

University of Southampton Research Repository ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

School of Medicine

**Characterisation of *BCR-ABL* and *FIP1L1-PDGFR* genomic
rearrangements in haematological malignancies**

by

Joannah Claire Score

Thesis for the degree of Doctor of Philosophy

September 2008

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES
SCHOOL OF MEDICINE

Doctor of Philosophy

CHARACTERISATION OF *BCR-ABL* AND *FIP1L1-PDGFR* GENOMIC REARRANGEMENTS IN HAEMATOLOGICAL MALIGNANCIES.

By Joannah Claire Score

Chromosomal translocations that produce oncogenic fusion genes are common in haematological malignancies but the mechanism by which they are formed is incompletely understood. Broadly there are three factors that are believed to be relevant to this process: (i) biological selection for specific gene fusions, (ii) susceptibility of particular chromosomal regions to breakage and (iii) the opportunity to recombine due to physical proximity. Analysis to date has indicated that translocations in leukaemia result from non-homologous end joining following two double stranded DNA breaks. Homologous recombination does not play a significant role. In lymphoid disorders DNA breaks may be a consequence of normal or aberrant RAG recombinase activity, whereas in myeloid disorders the reasons for breakage are not known.

The paradigm for fusion genes in leukaemia is *BCR-ABL*, produced as a consequence of the t(9;22). In chronic myeloid leukaemia, breaks within *BCR* are located in the major breakpoint cluster region (M-bcr), resulting in a p210 BCR-ABL protein. In Philadelphia chromosome positive acute lymphoblastic leukaemia (Ph+ ALL), however, the breaks within *BCR* are frequently located further upstream in the minor breakpoint cluster region (m-bcr), resulting in a smaller p190 protein. Since the p190 isoform is almost exclusively associated with Ph+ ALL, it is plausible that the breakpoints arise from a lymphoid progenitor cell specific mechanism, such as illegitimate RAG recombinase activity. To determine if RAG activity might be responsible for p190 breakpoints, fosmid fluorescence *in situ* hybridisation and multiplex long range PCR have been developed to identify the genomic breakpoints. This is technically challenging since the breaks are widely dispersed within both *BCR* and *ABL*, and thus there are only 2 cases have been fully characterised in the literature. Forward breakpoints in 25 p190 ALL cases, plus for comparison 25 breakpoints in patients with p210 ALL and 32 with p210 CML have been characterised.

Statistical analysis of the forward *BCR* breakpoints have revealed 2 distinct clusters in p210 ALL but none in CML or p190 ALL. Analysis of the forward *ABL* breakpoints showed evidence for 2 clusters in p210 ALL and one cluster at the 3' end of p190 ALL breakpoints, but no clustering in the CML breakpoints. When the forward breakpoints were compared to the proximity of sequence motifs the only significant association was a deficit of breakpoints falling in repeat regions for p210 CML *BCR* breakpoints and all p210 *BCR*

breakpoints. No significant differences were found in the distribution of forward breakpoints in *ABL* between the different subtypes of leukaemia. However, comparison of the patterns at reciprocal breakpoints (57% were characterised; 50% CML, 40% p210 ALL and 80% p190 ALL) showed differences between p190 and p210 (ALL and CML).

Functional analysis of a subset of cases with breakpoints found to be close to cryptic RSS sites revealed 1 of 5 p190 cases tested had a 'specific' RSS 4 bp from the forward breakpoint but at the exact location of the reciprocal breakpoint. Involvement of RAG was confirmed by the formation of both coding joints and signal joints, which is the defining feature of V(D)J recombination. In addition 1 of the 3 p210 ALL cases tested had a 'specific' RSS at the forward breakpoint, however only a hybrid joint and signal joints were formed therefore the significance of this is uncertain. Neither of the p210 CML cases tested gave 'specific' RSSs. This data therefore shows that aberrant RAG mediated recombination is involved in a subset of *BCR-ABL* translocations in ALL.

In parallel, a LR-PCR technique was optimised to identify genomic breakpoints in patients with chronic eosinophilic leukaemia (CEL) who have the *FIP1L1-PDGFR* fusion gene. Amplification of genomic breakpoints in these patients is difficult due to the large region in the *FIP1L1* gene in which the breakpoints can occur. To date 46 genomic breakpoints have been characterized using long range or bubble PCR, the LR-PCR technique is a more sensitive method of detecting the fusion at presentation with all 46/46 (100%) detected at first step LR-PCR compared to 22/44 (50%) detected by nested RT-PCR, the remaining 2 patients were nested negative by RT-PCR and therefore LR-PCR is now routinely used in screening for the fusion in patients at presentation. As gDNA LR-PCR was better at detecting the fusion at presentation, patient-specific gDNA RQ-PCR primers were designed to assess its sensitivity for monitoring for minimal residual disease (MRD) after imatinib treatment in selected patients. This was found to be a more sensitive test than both the routine reverse-transcriptase PCR (RT-PCR) and patient specific cDNA RQ-PCR MRD. Quantification of the amount of gDNA breakpoint in patients at presentation also revealed a large 40 fold variation, with many patients having low levels of the fusion and may explain why FISH can sometimes be unreliable for identification of this fusion, indeed 1/6 of the patients tested by both methods was gDNA RQ-PCR positive and FISH negative.

Analysis of genomic breakpoints has also enabled identification of a previously undescribed fusion of *KIF5B* to *PDGFR* in chronic eosinophilic leukaemia as a consequence of a complex translocation.

Contents

Contents.....	i
List of Figures.....	vii
List of Tables.....	x
Declaration of Authorship.....	xi
Acknowledgements.....	xii
List of Abbreviations.....	xiii

CHAPTER 1 : INTRODUCTION	1
1.1 Normal Haematopoiesis	1
1.2 Abnormal Haematopoiesis	3
1.3 Classification of Leukaemia	4
1.3.1 Chronic Leukaemia	5
1.3.2 Acute Leukaemia	5
1.4 The Philadelphia chromosome and the paradigm of the deregulation of the ABL kinase	6
1.4.1 The discovery of the <i>ABL</i> gene and its significance to leukaemia	6
1.4.2 The structure and function of the ABL protein	7
1.4.3 The role of BCR in activating the ABL proto-oncogene	10
1.4.4 BCR-ABL domains and their role in leukaemogenesis	10
1.4.5 The effects of BCR-ABL on signal transduction	11
1.4.6 Other oncogenic effects of the BCR-ABL fusion protein	14
1.4.7 BCR-ABL independent oncogenic effects and disease progression	15
1.4.8 The structure and diversity of the <i>BCR-ABL</i> fusion genes and their products	15
1.5 Targeted therapy	17
1.6 Other deregulated tyrosine kinases involved in MPNs	20
1.6.1 Platelet derived growth factor alpha (<i>PDGFRA</i>) gene fusions	21
1.6.2 Diagnosis of Idiopathic Hypereosinophilic Syndrome (IHES)	21
1.6.2.1 Discovery of <i>FIP1L1-PDGFRA</i> as a major cause of IHES/CEL	22
1.7 Mechanisms that give rise to translocations	23
1.7.1 Mechanism of the formation of the <i>BCR-ABL</i> fusion gene	26

1.8	Aims of this project	27
CHAPTER 2 : MATERIALS AND METHODS		28
2.1	Patient Samples	28
2.1.1	Preparation of blood and bone marrow samples	28
2.1.2	RNA extraction	29
2.1.2.1	cDNA synthesis	29
2.1.3	DNA extraction	29
2.2	Fluorescence in situ Hybridisation (FISH)	30
2.2.1	Culturing of clones for FISH	30
2.2.2	Extraction of BAC or Fosmid probes	30
2.2.3	Labelling	31
2.2.4	Probe precipitation	31
2.2.5	Slide preparation from fixed cell suspensions	31
2.2.6	Addition of probes to slides	31
2.2.7	Post hybridisation washes and detection	32
2.3	The Polymerase Chain Reaction (PCR)	32
2.3.1	High Fidelity PCR (Roche)	33
2.3.2	Expand Long Range PCR (LR-PCR)	33
2.3.2.1	Expand Long Range PCR Multiplex	33
2.3.3	Bubble PCR	34
2.4	Agarose Gel Electrophoresis	35
2.5	Cloning of PCR products	36
2.5.1	Ligations	37
2.5.2	Chemical transformation	37
2.5.3	Culturing and plasmid extraction	38
2.6	Restriction enzyme digests	38
2.7	Sequencing by dideoxynucleotide chain termination	39
2.7.1	ExoSAP	39
2.7.2	Sequencing reaction	39
2.7.3	Removal of unincorporated dye terminators	39
CHAPTER 3 : THE MECHANISM OF FORMATION OF <i>BCR-ABL</i> IN ALL		41
3.1	Introduction	41

3.1.1	Aims	44
3.2	Materials and Methods	44
3.2.1	Patient Samples	44
3.2.2	Whole Genome Amplification (WGA)	44
3.2.3	FISH to sub-localise the p190 breakpoint	45
3.2.4	PCR-based techniques	47
3.2.4.1	High Fidelity PCR	47
3.2.4.2	Expand Long Range PCR (LR-PCR)	47
3.2.4.3	Long Range Bubble PCR	47
3.2.4.4	Long distance inverse PCR (LDI-PCR)	48
3.2.5	Multiplex ligation-dependent probe amplification (MLPA)	49
3.2.6	Statistical and sequence motif analysis of breakpoints (performed by Ru-Fang Leh and Joe Wiemels)	52
3.2.6.1	Komolgorov-Smirnov test for distribution of breakpoints	52
3.2.6.2	Scan statistics to detect clustering of breakpoints	52
3.2.6.3	Silverman's smoothed bootstrap test for numbers of clusters	52
3.2.6.4	Proximity of breakpoints to sequence motifs	53
3.2.7	<i>In silico</i> analysis <i>BCR-ABL</i> breakpoints for cryptic RSSs	53
3.2.8	Constructs for use in RAG recombination assays	54
3.2.8.1	Preparing the vector	54
3.2.8.2	Preparing the insert	55
3.2.8.3	Ligating the insert into the cloning vector	56
3.2.8.4	Transforming <i>E.coli</i> cells with the construct	57
3.2.8.5	Endotoxin free plasmid DNA extraction	57
3.2.9	Transfecting NIH-3T3 cells	57
3.2.9.1	Growing NIH-3T3 cells	57
3.2.9.2	Preparing the NIH-3T3 cells for transfections	58
3.2.9.3	FuGene HD transfection	58
3.2.9.4	Plasmid DNA extraction from transfected cells	59
3.2.9.5	Digestion of parental plasmids	60
3.2.9.6	Ethanol precipitation of plasmid DNA	60
3.2.9.7	Transformation of <i>E.coli</i> with extracted, digested plasmid DNA	60
3.2.9.8	Screening of the colonies for RAG mis-recognition (see Figure 3-6)	61
3.3	Results	63
3.3.1	Screening for genomic breakpoints in p210 ALL	63
3.3.1.1	High Fidelity simplex PCR	63
3.3.1.2	Simplex LR-PCR	64
3.3.1.3	Multiplex LR-PCR	66
3.3.2	Control CML p210 patients using Multiplex LR-PCR	71

3.3.3	Screening for genomic breakpoints in p190 ALL	74
3.3.3.1	Fosmid FISH	74
3.3.3.2	LR-PCR	77
3.3.3.3	Multiplex LR-PCR on p190 samples	80
3.3.4	Long Range Bubble PCR	83
3.3.5	Long Distance Inverse (LDI) PCR	84
3.3.6	Multiplex LR-PCR with more primers	87
3.3.7	Reciprocal breakpoint detection and sequencing	93
3.3.7.1	Example of forward and reciprocal breakpoint determination for case Ital 10	93
3.3.7.2	Results of reciprocal breakpoint detection by LR-PCR	94
3.3.8	Detection of deleted reciprocal breakpoints by MLPA	96
3.3.9	Statistical analysis	98
3.3.9.1	Forward breakpoint locations	98
3.3.9.2	Clustering of breakpoints	98
3.3.9.3	Proximity of breakpoints to sequence motifs	102
3.3.10	Screening for a common sequence motif associated with recombination hot spots and genome instability in humans	104
3.3.11	Functional analysis of <i>BCR-ABL</i> breakpoints for cryptic RSSs	105
3.3.11.1	Measuring RAG recombination using the pGlow-TOPO expression vector.	106
3.3.11.1.1	Cloning the P _{trp} RSS oligo into pGlow	107
3.3.11.2	Measuring RAG recombination using pcDNA3.1/CT-GFP-TOPO vector	109
3.3.11.2.1	Amplification and cloning of the <i>Kan^R</i> gene into pcDNA	110
3.3.11.2.2	Amplification and cloning of the oop RSS fragment into pGlow TOPO	111
3.3.11.2.3	Sub-cloning oop RSS into pcDNA3.1 Kan ^R	112
3.3.11.3	Measuring RAG recombination using pGlow-TOPO vector	114
3.3.12	Assessment of cryptic RSSs using the extra chromosomal recombination assay (plasmids provided by Professor Nadel (207))	115
3.3.12.1	<i>In silico</i> identification of cryptic RSSs at <i>BCR-ABL</i> breakpoints	119
3.3.12.2	Additional constructs required for testing <i>BCR-ABL</i> breakpoint cryptic RSSs	122
3.3.12.3	Making constructs for testing patient breakpoints.	127
3.3.12.4	Optimising FuGene transfections	129
3.3.12.5	Cell culture assay with <i>BCR / ABL</i> test constructs.	132
3.3.12.5.1	Identification of ‘specific’ RSSs at the breakpoints in two patients	139
3.4	Discussion	144
3.4.1	Detection of forward <i>BCR-ABL</i> breakpoints	145
3.4.1.1	Detection of p210 ALL and CML genomic breakpoints	145
3.4.1.2	Detection of p190 ALL genomic breakpoints	147
3.4.2	Detection of reciprocal <i>ABL-BCR</i> breakpoints	151
3.4.3	Statistical analysis of forward breakpoints	154
3.4.4	Functional analysis of RSSs	155

CHAPTER 4 : UTILITY OF GENOMIC BREAKPOINT DETECTION IN PATIENTS WITH HYPEREOSINOPHILIA 160

4.1	Introduction	160
4.2	Materials and Methods	162
4.2.1	Patient samples	162
4.2.2	Preparation of blood and bone marrow samples	162
4.2.3	RT-PCR	162
4.2.4	Cloning, culturing and plasmid extraction	163
4.2.5	<i>PDGFRA</i> expression multiplex	163
4.2.6	Bubble PCR (see section 2.3.3)	164
4.2.7	Gel purification of bubble PCR products	164
4.2.8	Simplex LR-PCR to detect genomic <i>FIPILI-PDGFRA</i> breakpoints	165
4.2.9	Multiplex LR-PCR to detect genomic <i>FIPILI-PDGFRA</i> breakpoints	165
4.2.10	Qualitative gDNA MRD using nested PCR	165
4.2.11	TaqMan real time quantitative PCR (RQ-PCR)	166
4.2.11.1	Genomic DNA MRD detection using RQ-PCR	167
4.2.11.2	cDNA MRD using RQ-PCR	169
4.2.11.3	Absolute quantification of gDNA <i>FIPILI-PDGFRA</i> levels in presentation samples	169
4.2.12	FISH	170
4.3	Results	170
4.3.1	Detection of <i>FIPILI-PDGFRA</i> mRNA breakpoints	170
4.3.2	Detection of genomic <i>FIPILI-PDGFRA</i> breakpoints	171
4.3.2.1	Detection of <i>FIPILI-PDGFRA</i> genomic breakpoints by bubble PCR	171
4.3.2.2	Detection of <i>FIPILI-PDGFRA</i> gDNA breakpoints by simplex LR-PCR	174
4.3.2.3	Detection of genomic <i>FIPILI-PDGFRA</i> breakpoints by multiplex LR-PCR	176
4.3.2.4	Discrepancies between RT-PCR and multiplex LR-PCR	184
4.3.2.5	Overall comparison of RT-PCR and multiplex LR-PCR	186
4.3.3	Genomic DNA nested PCR for monitoring MRD	186
4.3.4	Genomic DNA real time quantitative PCR (RQ-PCR) for MRD detection	189
4.3.4.1	Comparison of gDNA and cDNA RQ-PCR and cDNA RT-PCR to assess the sensitivity of detecting MRD	196
4.3.4.2	gDNA RQ-PCR used for monitoring MRD	198
4.3.5	Variability of <i>FIPILI-PDGFRA</i> expression	199
4.3.6	Comparison of gDNA RQ-PCR and FISH	201
4.3.7	Structure of the <i>FIPILI-PDGFRA</i> breakpoints	202
4.3.8	Identification of a novel <i>PDGFRA</i> partner gene (292)	206
4.3.8.1	Results	207
4.3.8.1.1	Bubble PCR screening	207

4.3.8.1.2	Case report for E683	209
4.3.8.1.3	Confirmation of the presence of <i>KIF5B-PDGfra</i>	210
4.4	Discussion	213
4.4.1	<i>FIP1L1-PDGfra</i> breakpoint identification	213
4.4.2	Location of breakpoints	214
4.4.3	Comparison of methods to monitor MRD	215
4.4.4	Absolute gDNA RQ-PCR of presentation samples	217
4.4.5	Responses to imatinib	217
4.4.6	Identification of a novel <i>Pdgfra</i> fusion partner gene	218
CHAPTER 5	: CONCLUSIONS	220
	Appendix I	228
	Appendix II	230
	Appendix III	253
	Appendix IV	268
	Appendix V	270
	References	271

List of Figures

Figure 1-1: Haematopoiesis..	2
Figure 1-2: The clonal evolution of malignancy.	4
Figure 1-3: Mechanism of 1b ABL tyrosine kinase activation.....	9
Figure 1-4: Signal transduction by p210 BCR-ABL.....	12
Figure 1-5: Translocation breakpoint clusters in the <i>ABL</i> and <i>BCR</i> genes and the most common fusion transcripts.	16
Figure 1-6: RAG mediated recombination..	24
Figure 2-1: Bubble PCR.....	35
Figure 2-2: Invitrogen 1 kb plus DNA ladder sizes.....	36
Figure 3-1: Positions of the fosmids used in FISH and the primers used in screening for breakpoints.	46
Figure 3-2: Long Distance Inverse PCR.	49
Figure 3-3: A schematic representation of Multiplex Ligation-dependent Probe Amplification (MLPA).	50
Figure 3-4: Positions of the MLPA probes.	51
Figure 3-5: Coding and signal joint formation.	54
Figure 3-6: Experimental procedure to identify V(D)J recombination using Professor Nadel's constructs	62
Figure 3-7: High Fidelity breakpoint detection and sequencing for case 26231.	64
Figure 3-8: Normal control LR-PCRs.	65
Figure 3-9: Test multiplex LR-PCR for p210 ALL and CML breakpoint detection.....	67
Figure 3-10: Multiplex LR-PCR BREAKPOINT detection and sequencing for case 24375.....	68
Figure 3-11: Maps of <i>BCR</i> and <i>ABL</i> showing locations of the p210 ALL genomic breakpoints.....	70
Figure 3-12: Multiplex LR-PCR detection fo p210 CML cases.....	71
Figure 3-13: Maps of <i>BCR</i> and <i>ABL</i> showing the positions of the p210 CML control breakpoints	73
Figure 3-14: <i>ABL</i> fosmid FISH probe	75
Figure 3-15: Fosmid FISH on p190 cases.	76
Figure 3-16: Simplex LR-PCR breakpoint detection and sequencing of Ital 6.....	79
Figure 3-17: Simplex LR-PCR breakpoint amplification for Ital.....	80
Figure 3-18: Test multiplex LR-PCR to detect 25169 positive control.....	81
Figure 3-19: LR bubble PCR products.....	84
Figure 3-20: LDI-PCR products.....	86
Figure 3-21: Multiplex LR-PCR breakpoint detection for Ital 5.....	87
Figure 3-22: Multiplex LR-PCR breakpoint detection and sequencing for Ital 9	88
Figure 3-23: Multiplex LR-PCR breakpoint detection for P653	89
Figure 3-24: WGA LR-PCR products and multiplex LR-PCR breakpoint detection	90
Figure 3-25: Multiplex LR-PCR breakpoint detection and sequencing for 20763.....	91
Figure 3-26: Maps showing <i>BCR</i> and <i>ABL</i> forward breakpoints in p190 ALL.....	92
Figure 3-27: Reciprocal breakpoint detection using LR-PCR for Ital 10.....	94
Figure 3-28: Peak areas obtained from MLPA exported to Microsoft Excel	97

Figure 3-29: Analysis from clustering of p210 <i>BCR</i> breakpoints	99
Figure 3-30: Analysis from clustering of p190 <i>BCR</i> breakpoints.	100
Figure 3-31: Analysis from clustering of p210 CML, ALL and p190 ALL <i>ABL</i> breakpoints.	101
Figure 3-32: Positions of <i>ABL</i> versus <i>BCR</i> breakpoints.....	102
Figure 3-33: Proximity of breakpoints near sequence motifs.....	103
Figure 3-34: Modified from Raghavan <i>et al.</i> , (263), depicting the experimental procedure they used to detect V(D)J recombination.....	106
Figure 3-35: First strategy for determination of recombination using negative selection of the <i>GFP</i> gene.....	107
Figure 3-36: Cloning the P _{trp} RSS oligo into pGlow	108
Figure 3-37: Second experimental strategy for determining illegitimate RAG recombination.....	110
Figure 3-38: PCR and cloning of the RAG oop oligo	112
Figure 3-39: KpnI digest of the RAG oop oligo – pcDNA3.1 construct.....	113
Figure 3-40: Third experimental strategy for determining illegitimate RAG recombination.....	114
Figure 3-41: Recombination construct from Professor Nadel.....	116
Figure 3-42: The four possible ways that cryptic RSSs may be located at breakpoints.....	119
Figure 3-43: Agarose gels showing the process of making construct RS2	125
Figure 3-44: Sequencing of RS2.....	126
Figure 3-45: Constructs required to test RSS activity in 10 cases.....	128
Figure 3-46: Construction and verification of construct P851 RS8	129
Figure 3-47: Analysis of the RS39 positive control.....	131
Figure 3-48: BR events in construct 25654 RS24.....	135
Figure 3-49: Examples of colonies with ‘non-specific’ V(D)J recombination.....	137
Figure 3-50: Ital 6 RS8 ‘non-specific’ V(D)J recombination.....	138
Figure 3-51: P851 RS8 ‘non-specific’ V(D)J recombination.....	139
Figure 3-52: Ital 6 RS7 ‘specific’ V(D)J recombination.....	140
Figure 3-53: Ital 6 RS8 ‘specific’ V(D)J recombination.....	141
Figure 3-54: P851 RS7 ‘specific’ V(D)J recombination.....	142
Figure 3-55: P851 RS2 ‘specific’ V(D)J recombination.....	143
Figure 3-56: Breakpoints in p210 CML and p210 ALL compared to other previously published breakpoints	147
Figure 4-1: The FIP1L1, PDGFR α and FIP1L1-PDGFR α proteins.....	161
Figure 4-2: <i>PDGFRA</i> over-expression by RT-PCR	164
Figure 4-3: RQ-PCR using TaqMan probes.....	166
Figure 4-4: Flow diagram showing calculations for relative quantification.....	168
Figure 4-5: bubble PCR for HaeIII and RsaI digested <i>FIP1L1-PDGFR</i> RT-PCR positive and normal patients.....	172
Figure 4-6: E1080 single and nested step simplex LR-PCRs.....	175
Figure 4-7: E905 genomic DNA simplex LR-Pcr.....	176
Figure 4-8: Test simplex LR-PCRs on <i>FIP1L1-PDGFR</i> positive control	178
Figure 4-9: Positions of <i>FIP1L1</i> screening primers.....	179

Figure 4-10: Agarose gels of multiplex LR-PCR Mix 2 on positive controls	180
Figure 4-11: Agarose gels of <i>FIPILI-PDGFR</i> A positive control patients.....	181
Figure 4-12: <i>FIPILI-PDGFR</i> A genomic breakpoint LR-PCR multiplex on all positive controls.....	182
Figure 4-13: Agarose gels of <i>FIPILI-PDGFR</i> A multiplex LR-PCR on normal controls.....	183
Figure 4-14: <i>FIPILI-PDGFR</i> A gDNA multiplex LR-PCR on E1556.....	184
Figure 4-15: <i>FIPILI-PDGFR</i> A gDNA multiplex LR-PCR using one multiplex mix	185
Figure 4-16: genomic DNA nested PCR for detection of MRD in E359	187
Figure 4-17: Comparison of cDNA nested RT-PCR and gDNA breakpoint specific PCR for MRD analysis of <i>FIPILI-PDGFR</i> A positive patients.	189
Figure 4-18: RQ-PCR using the PrimerDesign probe.	191
Figure 4-19: Positions of cDNA and gDNA RQ-PCR probes in <i>PDGFR</i> A	192
Figure 4-20: RQ-PCR using the <i>PDGFR</i> A LNA probe and primers on normal DNA serial dilutions	193
Figure 4-21: <i>FIPILI-PDGFR</i> A specific RQ-PCR using the <i>PDGFR</i> A LNA probe	195
Figure 4-22: Graphs showing log reduction in <i>FIPILI-PDGFR</i> A fusion against time of imatinib for 7 <i>FIPILI-PDGFR</i> A positive patients and their follow-up samples.....	197
Figure 4-23: Graph showing log reduction in the levels of gDNA <i>FIPILI-PDGFR</i> A fusion against time of imatinib initiation in months for 13 patients.....	198
Figure 4-24: Graphs comparing the levels of gDNA <i>FIPILI-PDGFR</i> A fusion to cDNA RQ-analysis performed by Jovanovic <i>et al.</i> , (195).....	199
Figure 4-25: Absolute quantification using a <i>FIPILI-PDGFR</i> A plasmid standard curve	200
Figure 4-26: Graphical representation of percentage of cells harbouring the gDNA <i>FIPILI-PDGFR</i> A fusion gene in cases at presentation.....	201
Figure 4-27: The three types of <i>FIPILI-PDGFR</i> A splicing.	203
Figure 4-28: Gel showing <i>PDGFR</i> A over-expression results.....	207
Figure 4-29: UCSC BLAT results on sequence obtained from bubble PCR from patient E683.....	208
Figure 4-30: Domain structure of PDGFR α , KIF5B and KIF5B-PDGFR α	209
Figure 4-31: Detection of the <i>KIF5B-PDGFR</i> A fusion.....	210
Figure 4-32: Flanking FISH analysis of E683.....	211
Figure 4-33: Split FISH analysis of E683.	212

List of Tables

Table 3-1: Cases screened for p210 genomic breakpoints with simplex High Fidelity or Long Range PCR with <i>ABL</i> primers A and B.....	63
Table 3-2: The first set of cases screened with the p210 multiplex LR-PCR.....	68
Table 3-3: The second set of p210 ALL cases screened with the multiplex LR-PCR.	69
Table 3-4: Details of fosmid clones.	74
Table 3-5: Fosmid FISH results.	76
Table 3-6: Simplex LR-PCR primer combinations used in screening for p190 cases.....	78
Table 3-7: p190 cases screened by simplex LR-PCR.....	78
Table 3-8: Cases screened using the p190 multiplex LR-PCR.....	82
Table 3-9: Location of EcoR1 sites in <i>BCR</i> intron 1 and the size of fragment for LDI-PCR amplification.	85
Table 3-10: Forward and reciprocal breakpoints of 7 p190 ALL cases	96
Table 3-11: MLPA results on p190 and p210 (ALL and CML) cases who were negative for reciprocal breakpoints by LR-PCR.....	97
Table 3-12: Analysis of the breakpoint locations in the subtypes of leukaemia by the Komolgorov-Smirnov test.	98
Table 3-13: Frequencies of recombination motifs in the <i>BCR</i> and <i>ABL</i> breakpoint regions.....	104
Table 3-14: The locations of the upstream and downstream cassettes for use in the cell recombination assay..	117
Table 3-15: Authentic RSSs used in plasmid constructs.....	118
Table 3-16: Results of the EMBOSS Fuzznuc screen.....	121
Table 3-17: Cryptic RSS sites at the vicinity of the <i>BCR</i> or <i>ABL</i> breakpoints.....	122
Table 3-18: The locations of the upstream and downstream cassettes for use in the cell recombination assay.	124
Table 3-19: Comparison of assay results with those published by Marculescu <i>et al.</i> , (207).....	130
Table 3-20: Results of transfections with positive control constructs and test constructs.....	134
Table 4-1: <i>FIP1L1-PDGFR</i> A patients in whom genomic DNA breakpoints were found.....	174
Table 4-2: Details of the primers used in the LR-PCR multiplex to detect genomic <i>FIP1L1-PDGFR</i> A breakpoints.	177
Table 4-3: Comparison of the numbers of <i>FIP1L1-PDGFR</i> A patients positive by single step gDNA or cDNA	186
Table 4-4: Comparison of the percentage of <i>FIP1L1-PDGFR</i> A positive cells estimated by gDNA RQ-PCR and FISH.....	202
Table 4-5: Genomic breakpoints of <i>FIP1L1-PDGFR</i> A patients.	206

Declaration of Authorship

I, Joannah Claire Score declare that the thesis entitled:

Characterisation of *BCR-ABL* and *FIP1L1-PDGFR*A genomic rearrangements in haematological malignancies.

And the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly while in candidature for a research degree at this University;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly by others, I have made clear exactly what was done by others and what I have contributed myself;
- Parts of this work have been published as:
 - Identification of a novel imatinib responsive *KIF5B-PDGFR*A fusion gene following screening for *PDGFR*A overexpression in patients with hypereosinophilia. *Leukemia* 2006; 20(5): 827-832. Reference (292).
 - Cough and hypereosinophilia due to *FIP1L1-PDGFR*A fusion gene with tyrosine kinase activity. *Eur. Respir.* 2006; 27(1): 230-232. Reference (288).
 - Low-dose imatinib mesylate leads to rapid induction of major molecular responses and achievement of complete molecular remission in *FIP1L1-PDGFR*A-positive chronic eosinophilic leukemia. *Blood* 2007; **109**(11): 4635-4640. Reference (195).

Acknowledgements

I would firstly like to thank my supervisors Nick Cross and Francis Grand for their invaluable help, advice and guidance, without which none of this would have been possible. I would also like to thank all my fellow lab colleagues at the Wessex Regional Genetics Laboratory, especially Amy Jones, Claire Hidalgo-Curtis, Gemma Watkins, Vicky Hall, Andy Chase, Kathy Waghorn, Sebastian Kreil and Thomas Ernst for their friendship and advice both in matters relating to this thesis and otherwise. I would like to thank my family (including Xanthe, who I count as a family member) for their never-ending support in all I undertake and finally Adam for cooking all my meals, for all his support and for remembering what FISH stands for!

List of Abbreviations

aCML Atypical chronic myeloid leukaemia

AID Activation induced cytidine deaminase

ALL Acute lymphoblastic leukaemia

AML Acute myeloid leukaemia

Amp Ampicillin

A-MuLV Abelson murine leukaemia virus

ATP Adenosine triphosphate

BAC Bacterial artificial chromosome

BAD BCL-2 associated death promoter

BCL-2 B-cell lymphoma 2

BCR Breakpoint cluster region

bp base pair

BR Break recombination

CCT Cytosine triphosphate

cDNA Copy deoxyribonucleic acid

CEL Chronic eosinophilic leukaemia

CIAP Calf intestinal alkaline phosphatase

CJ Coding joint

CML Chronic myeloid leukaemia

CSR Class switch recombination

CYB Cytarabine

DAPI 4',6-diamidino-2-phenylindole

D β 1 D beta 1

D δ 2 D delta 2

ddNTP dideoxynucleotide triphosphate

DMEM Dulbecco's modified eagle's medium

DNA Deoxyribonucleic acid

dNTP deoxynucleotide triphosphate

DSB Double stranded break

EAC European against cancer

E.coli Escherichia coli

EDTA Etylenediaminetetraacetic acid

EIF4E Eukaryotic initiation factor 4E

ER Endoplasmic reticulum

ERK Extracellular related kinase

ET Essential thrombocythaemia

FACS Florescent activated cell sorting

FIP1L1 Fip-1-Like-1

FISH Fluorescence *in situ* hybridisation

FITC Fluorescein isothiocyanate

FLT3 FMS-like tyrosine kinase 3

GAB2 GRB2 associated binding factor 2

gDNA Genomic deoxyribonucleic acid

GDP Guanine diphosphate

GFP Green fluorescent protein

GRB2 Growth factor receptor bound factor 2

GTC Guanidine thiocyanate

GTP Guanine triphosphate

HES Hypereosinophilic syndrome

HR Homologous recombination

HSC Haematopoietic stem cell

IFN α Interferon alpha

IGH Immunoglobulin heavy

IHES Idiopathic hypereosinophilic syndrome

IL3 Interleukin-3

ICSBP Interferon consensus binding protein

ITD Internal tandem duplication

JAK Janus kinase

Kan Kanamycin

kb kilo base pair

KIF5B Kinesin family member 5B

LB Lauria Bertoni

LDI-PCR Long distance inverse polymerase chain reaction

LNA Locked nucleic acid

LR-PCR Long range polymerase chain reaction

MAPK Mitogen activated protein kinase

M-bcr Major breakpoint cluster region

m-bcr Minor breakpoint cluster region

MDS Myelodysplastic syndrome

MLPA Multiplex ligation probe amplification

M-MuLV Moloney murine leukaemia virus

MNC Mononuclear cell

MPN Myeloproliferative neoplasms

MRD Minimal residual disease

mTOR Mammalian target of rapamycin

NHEJ Non-homologous end joining

NRPTK Non-receptor protein tyrosine kinase

PBS Phosphate buffered saline

PCR Polymerase chain reaction

PDGFR α Platelet derived growth factor alpha

PDGFR β Platelet derived growth factor beta

Ph Philadelphia chromosome

PKC δ Protein kinase C delta

PI3K Phospho-inositol-3 kinase

PMF Primary myelofibrosis

PV Polycythaemia vera

RAG Recombinase activating gene

RNA Ribonucleic acid

ROS Reactive oxygen species

RPM revolutions per minute

RPTK Receptor protein tyrosine kinase

RQ-PCR Real-time quantitative polymerase chain reaction

RSB Re-suspension saline buffer

RSS Recombination signal sequence

RT-PCR Reverse transcriptase polymerase chain reaction

S/MAR Scaffold and matrix attachment region

SALSA Selective adaptor ligation, selective amplification

SCL Stem cell leukaemia

SDF1 Stromal derived factor 1

SDS Sodium dodecyl sulphate

SFK Src family kinase

SH1 Src-homology 1

SH2 Src-homology 2

SH3 Src-homology 3

SHC Src homology-2 containing

SHP SH2 containing phosphatase

SIL SCL-interrupting locus

siRNA Silent interfering ribonucleic acid

SJ Signal joint

SNAP Simple nucleic acid purification

SNP Single nucleotide polymorphism

SOCS Suppressors of cytokine signalling

SOS Son of sevenless

Src Sarcoma

SSA Single stranded annealing

SSC Sodium saline citrate

STAT Signal transducers and activators of transcription

T-ALL T-cell acute lymphoblastic leukaemia

TdT Terminal deoxynucleotidyltransferase

TTP Thymidine tri-phosphate

UCSC University California Santa Cruz

UV Ultra violet

V(D)J Variable (Diverse) Joining

WGA Whole genome amplification

WHO World health organisation

X-Gal 5, bromo-4-chloro-indolyl-beta-thio-galactoside

Chapter 1: Introduction

1.1 Normal Haematopoiesis

Haematopoiesis is the formation and development of blood cells by proliferation and differentiation from haematopoietic stem cells (HSCs). Maintenance of the haematopoietic system involves a tightly controlled steady-state of cell proliferation and cell death (apoptosis). As most mature blood cells have a limited life-span, haematopoietic stem cells continually produce new cells that become committed to granulocytic, monocytic, megakaryocytic, erythroid and lymphocytic lineages. In addition, the haematopoietic system must be able to allow quick cell production in response to biological stresses such as bleeding or infection and then be able to return to normal levels upon resolution of these stresses. Haematopoiesis is therefore a highly orchestrated system and as a result is regulated by complex genetic and environmental mechanisms that determine whether cells remain quiescent, proliferate, differentiate, self-renew or undergo apoptosis (1).

In adult humans the major site for haematopoiesis is the bone marrow, which contains all the different types of blood cells at various stages of development. Cell differentiation occurs from the morphologically indistinct stem cell down the progressively morphologically distinct multiple haematopoietic lineages via a successive series of committed progenitors (2) (see Figure 1-1).

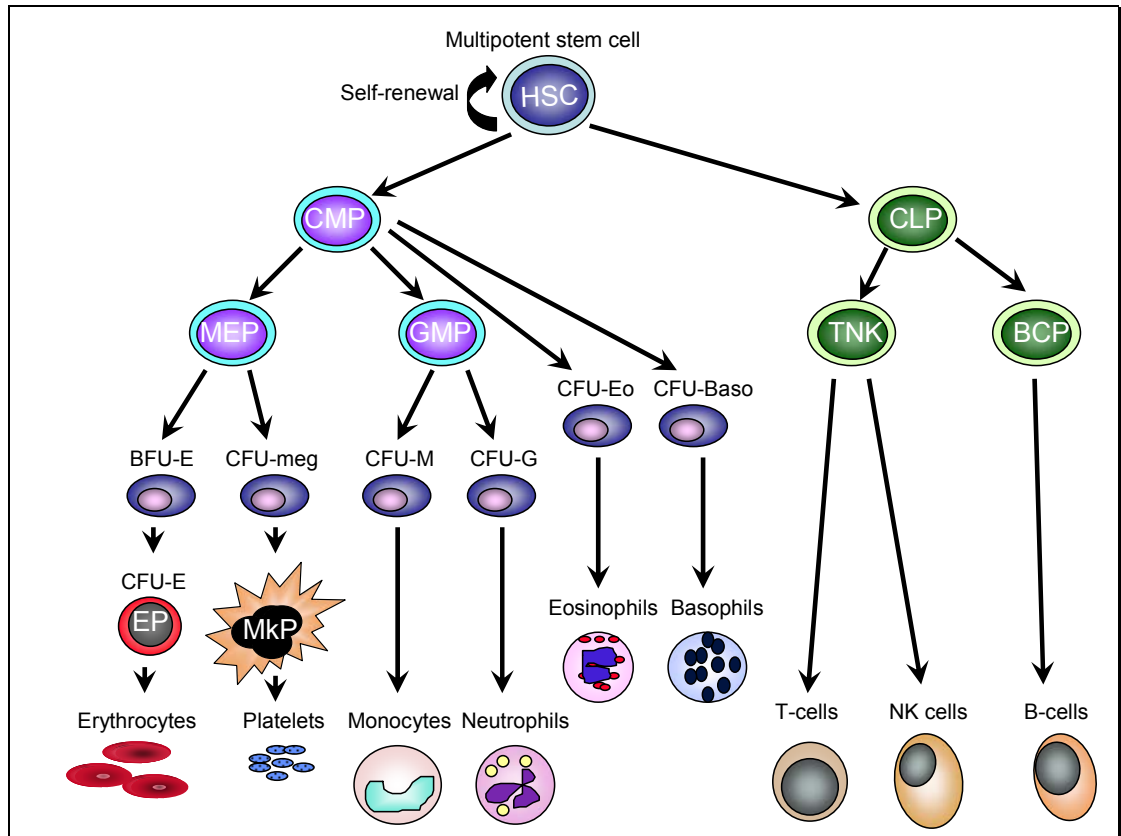


Figure 1-1: Haematopoiesis. Every mature blood cell is derived from a multipotent stem cell. HSC haematopoietic stem cell, CMP common myeloid progenitor, MEP megakaryocyte erythroid progenitor, GMP granulocyte-macrophage progenitor, EP erythroid progenitor, MkP megakaryocyte progenitor, CLP common lymphoid progenitor, TNK T-cell natural killer cell progenitor, BCP B-cell progenitor.

Haematopoiesis is mediated largely by cytokines which are soluble extracellular glycoproteins (3-5). In haematological cells, once the cytokines are bound to their cell surface receptors, they stimulate biological responses such as proliferation, survival and differentiation, though a variety of signal transduction pathways that impact on gene expression. Prominent amongst these is the JAK (Janus kinase)/STAT (signal transducers and activators of transcription) pathway, which is employed by a number of different receptors (6, 7).

Excessive signalling must be prevented and is therefore tightly controlled by several mechanisms. For example in the JAK/STAT pathway, JAKs may be inhibited by constitutive expression of antagonists such as SH2-containing phosphatases (SHPs) that

attenuate signal transduction (8, 9), or inducible inhibitors such as the suppressors of cytokine signalling molecules (SOCS) (10) which can directly inhibit signalling by JAK-associated cytokine receptors or target the receptor complex for ubiquitin mediated degradation (11, 12). Downstream signals through STATs may be blocked by protein inhibitors of activated STATs (PIAs) (13).

1.2 Abnormal Haematopoiesis

There are a number of haematological malignancies caused by the deregulation of the otherwise tightly controlled haematopoietic system, including leukaemia. Leukaemia, which is largely the result of acquired genetic abnormalities that confer a selective advantage to the host cell, leads to disruption of normal haematopoiesis. The progeny of the cell can acquire further mutations leading to further clonal expansion and disease progression by a variety of mechanisms, such as the generation of reactive oxygen species (ROS) (14) or inhibition of mismatch repair which gives rise to point mutations (15), both of which result in genetic instability and increased probability for further mutations (see Figure 1-2).

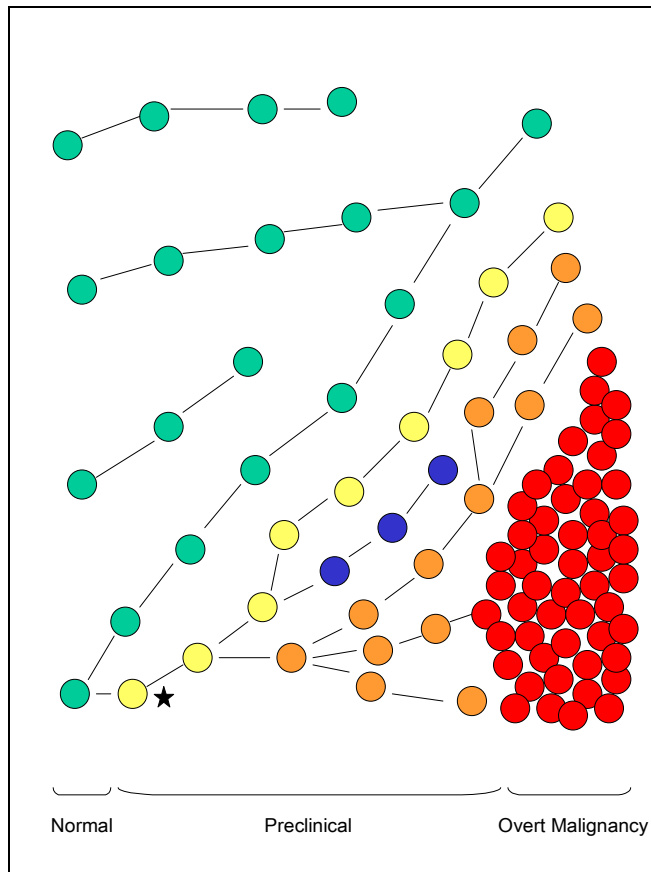


Figure 1-2: The clonal evolution of malignancy. The single yellow cell denoted with a star acquires a mutation providing a slight selective advantage over the normal green cells. The progeny of this cell acquire further mutations (blue, orange and red cells), some of which are deleterious causing the clone to die out (blue cells), however some result in uncontrolled expansion resulting in overt malignancy. In this example the end state contains a heterogeneous cell population (orange and red cells) with a dominant sub-population of cells (red cells).

1.3 Classification of Leukaemia

The World Health Organisation (WHO) classifies leukaemia into many distinct groups depending on the clinical course of the disease, cell morphology, clonal origins and the underlying genetic abnormalities. Broadly, leukaemia is categorised depending on whether the malignant cell type is lymphoid or myeloid. Chronic leukaemia results from proliferation and accumulation of differentiated clonally derived cells, which follow a near normal maturation (see Figure 1-1). Conversely acute leukaemia is characterised by a predominance of clonally derived cells that have a differentiation block causing accumulation of immature blast cells.

1.3.1 Chronic Leukaemia

Chronic leukaemia that is characterised by the over proliferation of cells from one or more of the myeloid lineages are termed by the WHO as chronic myeloproliferative neoplasms (CMPNs), and are: chronic myeloid leukaemia (CML), chronic eosinophilic leukaemia/hypereosinophilic syndrome (CEL/HES), polycythaemia vera (PV), essential thrombocythaemia (ET), primary myelofibrosis, chronic neutrophilic leukaemia and unclassifiable CMPNs. Chronic leukaemias that are characterised by both dysplastic and proliferative features of the myeloid lineages at presentation are termed the myelodysplastic/myeloproliferative neoplasms and include chronic myelomonocytic leukaemia (CMML), juvenile myelomonocytic leukaemia and atypical chronic myeloid leukaemia (aCML). Chronic leukaemia arising from the over-proliferation of cells from the lymphoid lineages includes chronic lymphocytic leukaemia and prolymphocytic leukaemia.

1.3.2 Acute Leukaemia

The WHO criteria for the diagnosis of acute leukaemia is denoted by the presence of at least 20% blast cells in the blood or bone marrow at presentation, although this can be lower if specific acute leukaemia-associated cytogenetic or molecular genetic abnormalities are present. Further classification is again based upon lineage, therefore acute myeloid leukaemia (AML) has blast cells that are derived from myeloid lineages with sub-classifications either based upon the genetic abnormalities involved such as *PML-RARA* (16) or *AML1-ETO* (17) gene fusions, or on morphology and immunophenotype. Acute lymphoblastic leukaemia (ALL) is divided into two types of disease, T-ALL and B-ALL again reflective of the lineage involved i.e. either T-lymphocytes or B-lymphocytes, respectively, and further sub-classifications based on the immunophenotype displayed by the abnormal cells. Again there are sub-classifications based on the type of genetic abnormalities, some of which overlap the myeloid leukaemias (such as the *BCR-ABL* gene fusion found in CML, rare cases of AML and in some B-ALL (18) and some that are more closely related to the biology of the malignant cell type, such as those involving fusions of antigen receptor genes.

1.4 The Philadelphia chromosome and the paradigm of the deregulation of the ABL kinase

The first chromosomal abnormality to be associated with a haematological malignancy was the Philadelphia chromosome (Ph). It was discovered in 1960, when Nowell and Hungerford (19) first described a correlation between CML and the presence of a 'minute' chromosome, which was later named the Philadelphia chromosome after its place of discovery. In 1973, Rowley (20) identified that the Ph chromosome was in fact a reciprocal translocation between the long arms of chromosomes 9 and 22, i.e. t(9;22)(q34;q11). The translocation was characterised molecularly in the 1980s and was the result of the first example of a fusion gene in malignancy. Specifically, the t(9;22) was shown to fuse the breakpoint cluster region (*BCR*) gene at 22q11 to the Abelson (*ABL*) gene at 9q34, resulting in the production of a BCR-ABL fusion protein with its increased tyrosine kinase activity (21) that is able to transform cells both *in vitro* and *in vivo* (22).

BCR-ABL has now become the defining marker of CML and by definition is present in 100% of cases (23). *BCR-ABL* is also present in 30% of adult ALL (24), 3 to 5% of childhood ALL (25) and in 2% of AML (26).

1.4.1 The discovery of the *ABL* gene and its significance to leukaemia

The *ABL* gene was discovered as a result of work carried out in the 1970s on the Abelson murine leukaemia virus (A-MuLV), although the full significance would not be realised until later. Abelson and Rabstein (27) discovered A-MuLV by inoculating the thymoma causing Moloney leukaemia virus (M-MuLV) into thymectomised mice, one of which developed a lymphosarcoma. The extracted A-MuLV virus reproducibly gave rise to lymphomas *in vivo* and could also transform lymphoid cells *in vitro* whereas the M-MuLV could not (28). One hypothesis for the difference in transforming ability of the A-MuLV was that a recombination event had taken place between the M-MuLV and a mouse host cellular oncogene (29, 30) and that this oncogene was likely to be the transforming component (31).

The acquired viral p120 protein was isolated from transformed cells and the resulting A-MuLV protein was found to be composed of the viral *gag* gene fused to a non-M-MuLV component derived from a normal cellular gene in mice, designated *C-abl* (32). It was subsequently shown that the human cellular counterpart *ABL* was located on chromosome 9 (33) and that *ABL* was not only translocated onto the Ph chromosome in Ph positive CML patients (34, 35) but was in fact localised next to the translocation breakpoint (36, 37) suggesting a role for *ABL* in the genesis of CML.

1.4.2 The structure and function of the ABL protein

It is now known that the human proto-oncogene *ABL* encodes a non-receptor tyrosine kinase. Tyrosine kinases are enzymes that catalyse the transfer of the terminal phosphate of adenosine triphosphate (ATP) to tyrosine residues of substrate proteins, however these enzymes may also autophosphorylate on specific tyrosine residues. It is believed that tyrosine kinases, of which there are 90 in the human genome, evolved to mediate aspects of multicellular communication and development. They are crucial for the response of haematopoietic progenitor cells to external growth stimuli and are involved in the regulation of cell proliferation, migration, differentiation, survival and cell adhesion. The activity of tyrosine kinases is normally very tightly controlled (38), and is deregulated in many cases of leukaemia (31).

ABL is highly conserved across evolution, expressed in many tissues throughout development, involved in many cellular processes and localised in the cytoplasm, nucleus and endoplasmic reticulum (ER) (39-42). The location, level of phosphorylation and expression of *ABL* dictates whether it has a positive or negative effect on cell growth (42-45). In normal cells, 14-3-3 proteins keep *ABL* in the cytosol, however, upon genotoxic stress, such as ionising radiation, *ABL* is released and translocates to the nucleus (42). Nuclear *ABL* exerts a negative effect on cell growth when it is activated in association with ataxia telangiectasia-mutated kinase to promote apoptosis (38, 46). *ABL* that is located in the ER may also induce apoptosis by communicating ER stress to the mitochondria when it is bound to protein kinase C delta ($PKC\delta$) (45) and therefore *ABL* ironically may also serve as a tumour suppressor gene.

ABL is also involved in signalling pathways regulating growth factor-induced proliferation when located in the cytoplasm (23, 42, 44). To facilitate understanding of the ABL tyrosine kinase and its role in the genesis of leukaemia, it is important to be able to relate its cellular functions to its protein structure. The ABL protein contains many different domains reflecting the various functions it has within the cell (see Figure 1-3). The first exon can be alternatively spliced and therefore the protein can exist in two different isoforms 1b or 1a (47). Figure 1-3 shows how the 1b form of ABL is maintained in its inactive form by tight intramolecular interaction of the N-terminal myristoyl group with the C lobe of the tyrosine kinase domain (48). It has been less clear how 1a ABL is regulated, although this isoform retains the serine 69 which, when phosphorylated, binds the ABL Src homology-2 (SH2) domain, where it may contribute to inhibition (49). A recent study has shown, however, that 1a ABL can be inhibited through binding of an ABL substrate called Abi1, that binds to phosphorylated tyrosine residue 312 and PXXP in the ABL SH2 and Src homology-3 (SH3) domains, respectively (50).

The ABL SH2 and SH3 domains provide docking sites for specific ligands (51), which activate the enzyme by changing its conformation and allowing oligomerisation and subsequent tyrosine phosphorylation of receptor subunits (52). This leads to repositioning of the activation loop away from the Src homology-1 (SH1) tyrosine kinase domain, allowing binding of ATP and substrate activation. Autophosphorylation of other tyrosine residues also creates binding sites for other SH2 containing signalling molecules, which are present in many signalling molecules (53). Both the ABL SH2 and SH1 domains are necessary for transformation (46).

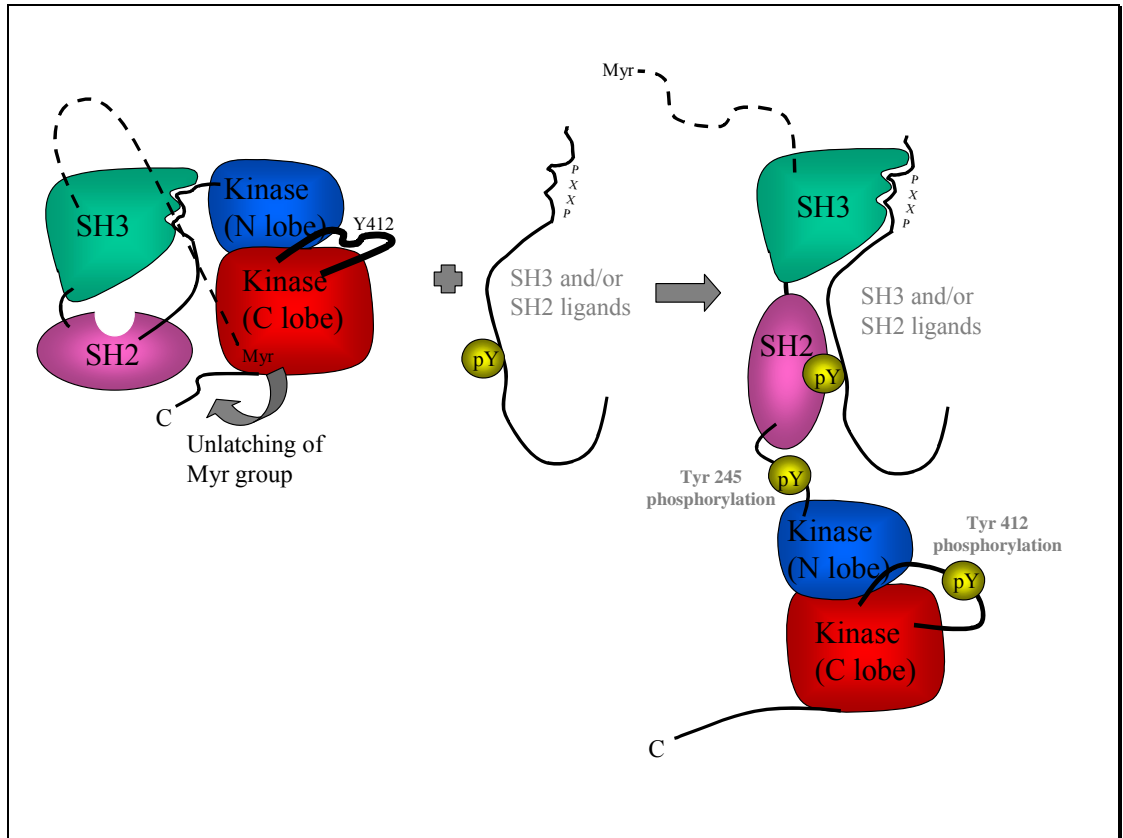


Figure 1-3: Mechanism of 1b ABL tyrosine kinase activation. ABL is maintained in an inactive conformation by anchorage of the myristoyl group and through clamping of the negative regulatory SH3 domain to the SH2 domain. ABL tyrosine kinase domain activation requires attachment of phosphotyrosine and proline rich proteins to the ABL SH2 and/or SH3 domain as well as dimerisation. The kinase domain then opens up, exposing the activation loop and enabling recruitment of ATP.

At the C-terminus of ABL there are nuclear localisation and export sequences, which allow it to move in and out of the nucleus (54), a unique DNA binding domain and an F-actin binding domain that allows ABL to bind to the cytoskeleton (55) and which may be involved in cell-cell interactions (39). The N-terminal SH3 domain also has an important role in regulating the tyrosine kinase activity of ABL by forming a clamp with the SH2 domain, keeping it in its inactive form, thus preventing binding of ATP (56). The SH3 domain is not present in the v-abl component of the A-MuLV and it is the lack of this regulatory domain that causes constitutive activation of the v-abl kinase (57). However in the BCR-ABL fusion protein, the SH3 protein is intact suggesting a different mechanism of activation (39).

1.4.3 The role of BCR in activating the ABL proto-oncogene

Once the structure and origin of the BCR-ABL fusion protein was elucidated, the mechanism by which BCR caused constitutive activation of ABL and hence transformation could be studied. It was already known that the tyrosine kinase activity of the fusion protein was increased and was likely to have an important role in transformation (31). A coiled-coil domain encoded by *BCR* exon 1 was found to be essential for constitutively activated tyrosine kinase activity in the fusion protein by causing dimerisation of the BCR-ABL proteins, thereby allowing the 2 adjacent BCR-ABL molecules to phosphorylate each other on tyrosine residues (58) resulting in increased tyrosine kinase activity. BCR-ABL was also found to have serine kinase activity that could not be inhibited by site directed ABL specific antisera, suggesting that this activity was not from the ABL component but BCR (59), a postulate that was subsequently confirmed (21).

1.4.4 BCR-ABL domains and their role in leukaemogenesis

Although the ABL kinase domain (SH1) is critical for BCR-ABL transforming ability, as demonstrated by the fact that ABL kinase inactivating point mutations in the ATP binding site inactivate the transforming ability of BCR-ABL (60) as well as by therapies that target kinase activity, (61) there are other important domains such as those that connect to downstream signalling pathways.

As mentioned previously, the BCR coiled-coil domain is necessary for transformation by constitutively activating ABL (58, 62), however, there is a BCR SH2 binding domain (encoded for by amino acids 176-242 of *BCR* exon 1) that binds to the SH2 domain of ABL making it into a more active and accessible conformation allowing interaction with tyrosine phosphorylated proteins (58). In addition there is a tyrosine residue on BCR designated Y177 that is phosphorylated and binds the growth factor receptor bound protein 2 (GRB2) adaptor molecule. GRB2 then recruits GRB2-associated binding protein 2 (GAB2) which leads to the activation of downstream cell proliferation signals and is required for the induction of MPN by BCR-ABL in murine models (63-65). GRB2 can also be recruited by the SH2 domain of ABL through the binding and phosphorylation of the adaptor molecule SHC (Src homology 2 domain containing) to the SH2 domain (66) and as such, mutations in SH2 reduce the ability of BCR-ABL to induce a CML-like

disorder in mice (67, 68). Therefore both BCR and ABL domains play a role in the transforming ability of BCR-ABL.

1.4.5 The effects of BCR-ABL on signal transduction

Although many signalling proteins interact with BCR-ABL through its various functional domains, the RAS pathway is critical for the transformation of cells (69), as proved by the use of dominant negative RAS mutants, which block transformation by BCR-ABL (70, 71). Indeed, deregulated RAS signalling has been implicated in a wide variety of human leukaemias (72, 73). RAS is directly activated through the binding of the SH2 domain of GRB2 to Y177 of BCR-ABL (65), BCR-ABL-GRB2 then recruits SOS (Son of sevenless), which stimulates conversion of the inactive GDP-bound form of RAS to its active GTP-bound state (69) (see Figure 1-4). However, BCR-ABL can also activate RAS indirectly by the activation of SHP2 by recruitment of the scaffold protein GAB2 to the BCR-ABL-GRB2 complex (66, 74) or through other SH2 containing adaptor molecules that can connect the tyrosine kinase signals to the RAS signalling pathway, such as activation by the SHC adaptor molecule (66). Once activated, RAS exerts its proliferative effects by activating the transcription factors FOS and JUN through the mitogen-activated protein kinase (MAPK) cascades ERK and JNK respectively (75). However, RAS, in addition to increasing the proliferation rates of cells, also has inhibitory effects on apoptosis therefore increasing cell survival (76) by inducing the expression of the anti-apoptotic molecules BCL-2 (B-cell lymphoma 2) and BCL-XL (77). JUNB is an antagonist of RAS and is therefore a negative regulator of cell proliferation and survival. The findings that JUNB is down regulated in CML cells isolated from patients (78) and that murine models with inactivation of JUNB leads to a CML-like disease (79) are consistent with the importance of RAS signalling in promoting the growth of BCR-ABL positive cells.

Phospho-inositol-3-phosphate kinase (PI3-K) is another pathway through which BCR-ABL transduces its oncogenic signals (80). PI3-K, like RAS has been linked to many human cancers (81) and its normal function is to stimulate DNA synthesis after ligand activation by some receptor tyrosine kinases. PI3-K is directly activated by the BCR-ABL-GRB2-GAB2 complex (82), by RAS and by the SH2 domain of BCR-ABL, and signals through the Akt/mTOR (mammalian target of rapamycin) pathway (83-85). mTOR both increases

translation of mRNAs necessary for cell cycle progression by up-regulation of the translation initiation complex EIF4E (Eukaryotic initiation factor 4E) (86) which can, in turn, up-regulate MYC (87). In addition, up-regulation of PI3K also inhibits pro-apoptotic proteins (88) and BCL-2 associated death promoter (BAD) (89) as well as inducing the expression of the anti-apoptotic molecule BCL-2 (90).

BCR-ABL also signals through the JAK/STAT pathway (91) particularly JAK2 (92-94). JAK2 interacts directly with the SH2 domain of BCR-ABL and is necessary for its full oncogenic effects (95), at least in some contexts. JAK2 exerts its effects partly through the up-regulation and suppression of degradation of MYC (95). Indeed, enforced expression of the IL3 receptor in NIH-3T3 cells results in susceptibility to transformation by BCR-ABL via a mechanism that involves JAK2 and STAT5 (96).

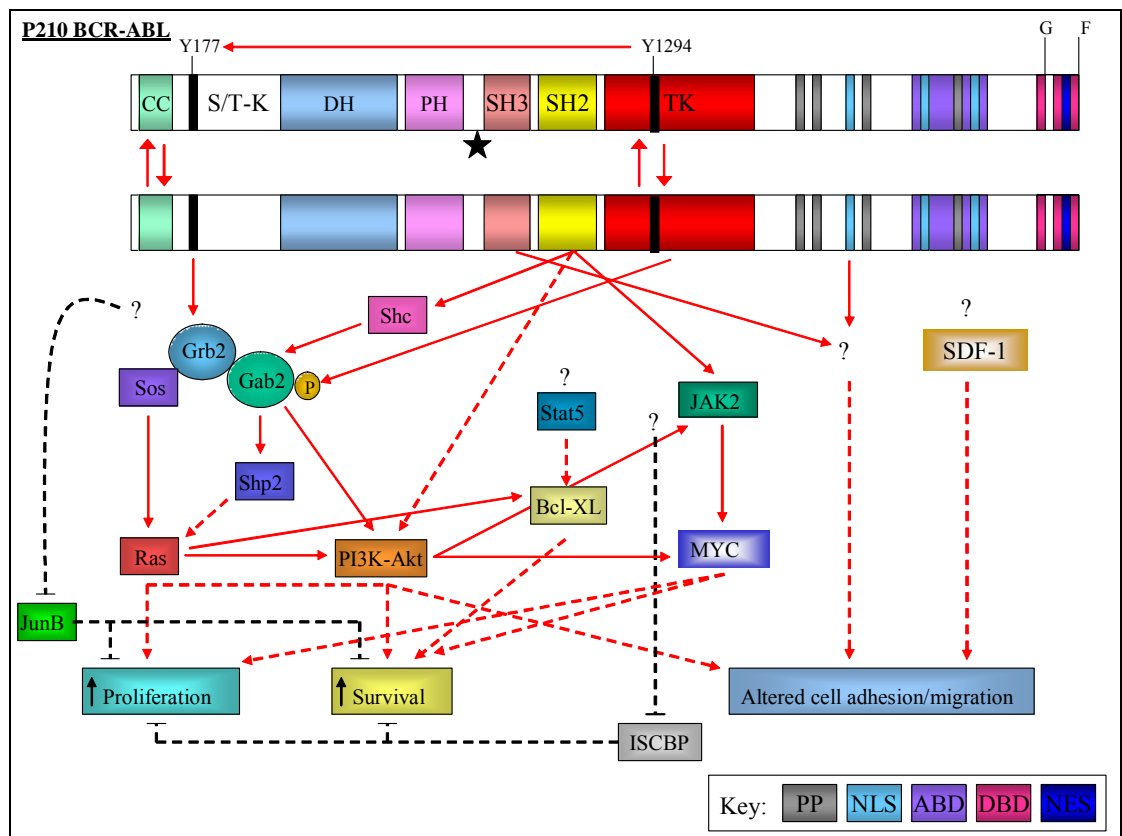


Figure 1-4: Signal transduction by p210 BCR-ABL. The breakpoint between BCR and ABL is denoted by a star (after the PH domain of BCR). Up regulation of signals is indicated by the red lines and down regulation by the black lines.

As already eluded to, the proto-oncogene MYC has also been implicated in the transforming activity of BCR-ABL. MYC is over-expressed in CML cell lines (66, 74, 97, 98) and introduction of dominant negative isoforms reduces the transforming ability of BCR-ABL (97). MYC transcription can be induced by mTOR, (87) JAK2 (95, 99) and the WNT signalling pathway (100). WNT regulates cell proliferation, differentiation and stem cell renewal, it has been shown that BCR is a negative regulator of WNT and that this putative tumour suppressor function of BCR is abrogated when BCR is fused to ABL (100).

Due to the complexities of signal cascades, experimental systems for testing the effects of signalling molecules in the genesis of leukaemia can lead to conflicting results. For example, STAT5 is constitutively activated in BCR-ABL positive leukaemias (93), and has been shown to activate the anti-apoptotic BCL-XL gene in cell cultures (101), however the expression of BCR-ABL in bone marrow cells of STAT5a/b knockout mice still led to a CML-like MPD, suggesting that neither STAT5a nor STAT5b are required for BCR-ABL mediated leukaemogenesis (102). However, it is now widely accepted that the strain of mouse that was used was not a true knockout (103) and further studies with both a complete STAT5a/b knockout mouse model and siRNA against STAT5 have confirmed a requirement for this molecule in the transformation of BCR-ABL positive leukaemia (104, 105).

There are also differences in signalling between BCR-ABL positive CML and B-ALL. For example, the use of a RAS inhibitor that blocks all 3 isoforms (NRAS, KRAS and Ha-RAS) showed that RAS is essential for BCR-ABL transformation in CML but not B-ALL (106). Conversely, the three Src-family kinases (SFK) that are expressed in haematopoietic cells (LYN, HCK and FGR) are over-expressed in BCR-ABL positive diseases (107), however a CML-like MPN can still be reproduced by retroviral transfer of BCR-ABL into Lyn/Hck/Fgr triple knockout mice, indicating that these kinases are not necessary for the pathogenesis of chronic phase CML (108). In B-ALL however, absence of any two of these three SFKs leads to drastically impaired disease (108). It is possible that these differences are linked to rapid disease progression seen in ALL.

Along with the increase in proliferative and survival abilities of cells through RAS and other signalling molecules, BCR-ABL may also suppress expression of negative regulators of haematopoiesis. Interferon consensus sequence binding protein (ICSBP) has the ability to down-regulate proliferation and survival of BCR-ABL cells. ICSBP is down regulated in cells from CML patients (109) and knockout murine models of ICSBP develop a CML-like disease (110). This suggests that ICSBP acts as tumour suppressor, the abrogation of which is important for the development of CML.

1.4.6 Other oncogenic effects of the BCR-ABL fusion protein

As described above, the BCR-ABL fusion protein deregulates tyrosine kinase activity through altering intracellular signalling cascades that have detrimental effects on the cells ability to control the intricate balance between cell proliferation and apoptosis (43). However, it is not just changes in signalling that have been implicated in BCR-ABLs transforming ability. The fusion of *BCR* to *ABL* also alters the sub cellular localisation of ABL, by blocking nuclear translocation signals and increasing association with the cytoskeleton and cytoplasm by activating the F-actin binding function (58). This exclusive localisation of BCR-ABL in the cytoplasm means that the pro-apoptotic effects of nuclear ABL no longer exist and there is inhibition of apoptosis.

For maintenance of normal haematopoietic cell regulation and development, the interaction between the bone marrow stromal cells and haematopoietic progenitor cells is critical. The bone marrow stromal cells, to which normal progenitor cells adhere, provide a highly controlled microenvironment where cell proliferation is tightly regulated which is crucial for the cells to develop correctly (43, 44). Normal ABL kinases are required for correct formation, maintenance and maturation of intercellular adhesion (111), however BCR-ABL positive cells have altered cellular interactions, most notably their decreased adhesion to the bone marrow stromal cells and their ability to proliferate in the extramedullary space (112). Stromal-derived factor 1 (SDF-1) and its receptor CXCR4 are critical for normal cell movement and adherence in the bone marrow microenvironment (113, 114). Both have been found to be deregulated by BCR-ABL (115) and may be required for the altered adhesive properties of CML cells (116), resulting in them leaving the inhibitory environment of the bone marrow and entering the circulation.

1.4.7 BCR-ABL independent oncogenic effects and disease progression

In addition to direct effects of BCR-ABL on intracellular signalling, location and adhesion, there are also seemingly unconnected mechanisms by which the oncogenic cells can proliferate or escape apoptosis. For example, micro RNAs have been recently linked to the pathogenesis of CML and ALL (117). Micro RNAs are small non-coding RNAs that silence specific target genes and are therefore able to regulate many cellular processes (118). ABL is thought to be a putative target of miR-203, which is thought to have tumour suppressive properties. In CML and ALL, miR-203 is hypermethylated but when re-expressed can reduce BCR-ABL fusion protein levels and inhibit cell proliferation (117).

BCR-ABL independent or co-operating mechanisms of transformation are believed to underlie the progression of CML to blast crisis and also the development of BCR-ABL positive ALL. One example is the B-lymphoid transcription factor Ikaros, which has been found to be deleted in the majority of BCR-ABL positive ALL and lymphoid blast crisis CML cases but is intact in chronic phase CML (119). Deletion of p16 has also been implicated in lymphoid blast crisis and BCR-ABL positive ALL, but the mechanisms responsible for myeloid blast crisis remain poorly understood.

1.4.8 The structure and diversity of the *BCR-ABL* fusion genes and their products

The *ABL* gene breakpoints can occur anywhere in the first 150 kb of the gene, more specifically upstream of exon 1b, between exon 1b and 1a or between exons 1a and a2 (120). The breakpoints in the *BCR* gene are more variable. In CML, the breakpoints in the majority of cases (approximately 99%) fall within the major breakpoint cluster region (M-bcr) which spans 5 exons now known to be exons 12 to 16, but historically called exons b1-b5. This usually gives rise to a hybrid messenger RNA (mRNA) with either a e13a2 (b2a2) or e14a2 (b3a2) junction encoding a p210 fusion protein (see Figure 1-5). The Ph chromosome is also found in 30% of adult ALL (24) and 3-5% of childhood ALL (25). In contrast to CML, only about 30-50% of patients with Ph positive ALL harbour the p210 fusion protein, with the remaining 50-70% harbouring the p190 fusion protein. This

smaller protein results from breakpoints in the *BCR* minor breakpoint cluster region (m-bcr) within the large intron 1, giving rise to a hybrid e1a2 mRNA (see Figure 1-5). Because of this transcript variability it has been postulated that some fusions may be loosely associated with various clinical phenotypes (18).

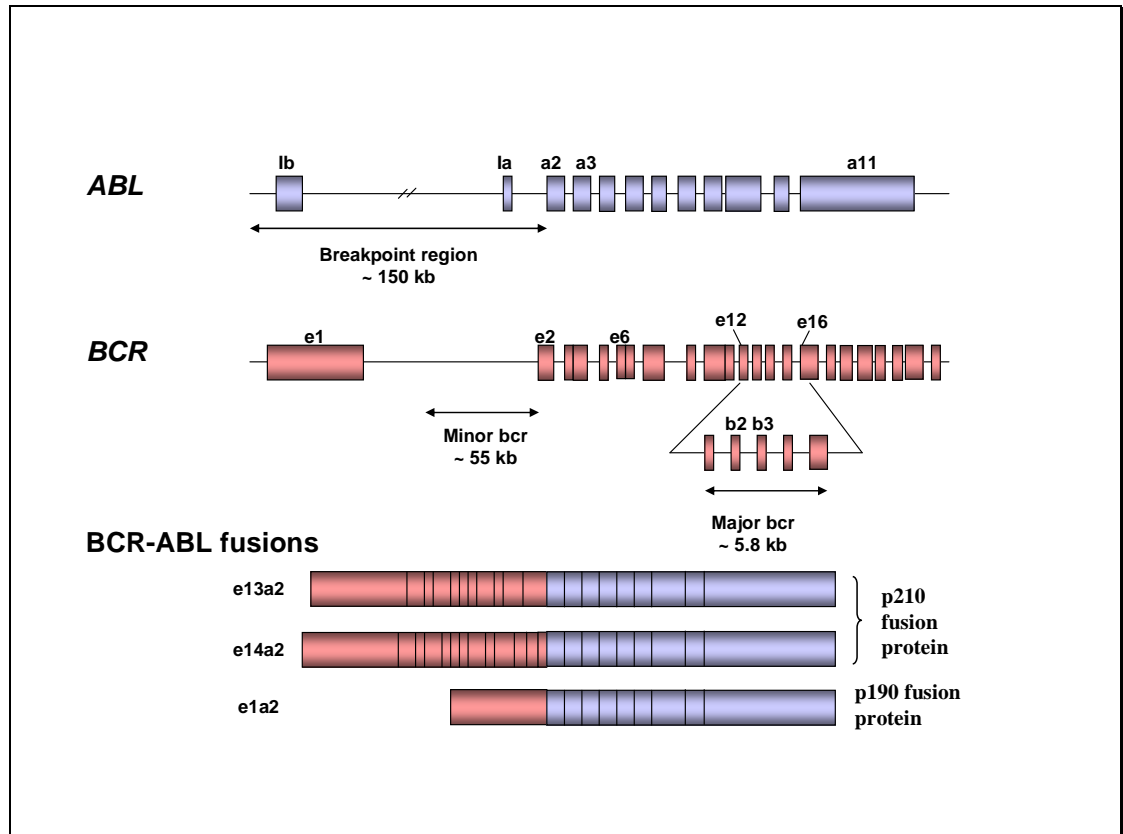


Figure 1-5: Translocation breakpoint clusters in the *ABL* and *BCR* genes and the most common fusion transcripts.

There are also rare cases of atypical *BCR-ABL* fusions in CML and ALL patients, for example e6a2, e8a2, e19a2 and fusions that lack *ABL* exon a2 (121-123).

The role of the reciprocal *ABL-BCR* fusion protein is much less clear, even though it appears to be transcriptionally active in 65% of CML patients (124). About 10-15% of CML patients, often those with atypical translocations, have large deletions on the derivative 9 chromosome that are believed to occur at the time of formation of the Ph chromosome (125, 126). Patients with deletions have poorer median survival compared to patients with no detectable deletion (125, 127, 128). There are two principal hypotheses as

to why this may be the case, (i) the deletion of neighbouring genes on chromosomes 9 and 22 that have tumour suppressive properties, or (ii) the deletions may reflect a higher genetic instability in the cell at the time of Ph chromosome acquisition and therefore may be more prone to acquiring further genetic alterations (125).

1.5 Targeted therapy

The cumulative effect of the intensely studied structure, function and effects of BCR-ABL has led to the first rationally designed targeted therapy for cancer (61). The transforming domain of the BCR-ABL protein is the SH1 tyrosine kinase domain. It is here that ATP binds and the transfer of phosphate from the bound ATP to the tyrosine residues of substrates takes place (129). It was postulated that an ATP mimic could bind to this site thus preventing phosphorylation of its substrates, thereby blocking activation of signalling transduction pathways and thus proliferation of the malignant clone (61). Several inhibitors were purified and tested but the best inhibitor was a 2-phenylaminopyrimidine derivative, subsequently called imatinib mesylate (130). Indeed, imatinib has been shown to inhibit CML cell growth and induce apoptosis both *in vitro* and *in vivo* (131).

Imatinib was initially trialled on CML chronic phase patients who had no response to the then current best chemotherapy, interferon- α (IFN α). The results were very encouraging (132, 133), and a large randomised trial of chronic phase CML patients was performed, tracking the progress of over 1000 patients who were given either imatinib or IFN α plus cytarabine (CYB). Imatinib was found to be superior to IFN α plus CYB therapy (achieving a complete cytogenetic response in over 75% of patients compared to less than 35% in the IFN α group). In addition the side effects and intolerance levels to imatinib were markedly lower (134). In a recent follow-up analysis of patients who received imatinib from diagnosis, the survival rate was 89% by 60 months with 7% of patients progressing to accelerated phase or blast crisis (135).

Despite the successful use of imatinib for CML, it has some limitations. First, residual leukaemia usually remains detectable by reverse transcriptase polymerase chain reaction (RT-PCR) and therefore imatinib treatment is not a cure (136, 137). CML patients on

imatinib therapy may become resistant through point mutations in the ATP binding region of *BCR-ABL* (138, 139) or other mechanisms. Extensive *in vitro* studies found that some mutations both inside and outside the tyrosine kinase domain can destabilise the inactive form of *BCR-ABL* to which imatinib binds, to its active conformation, which precludes imatinib binding (140). Alternatively, mutations may result in steric hindrance or eliminate critical H-bonds, either of which can reduce or eliminate imatinib binding (141). Other mechanisms of resistance are less well defined, but in some cases appear to be related to impaired ability of imatinib to enter or remain inside target cells.

The beneficial effects of imatinib are reduced for Ph positive ALL as well as CML in accelerated or blastic phases (142). This may be partly due to the higher rate of mutations compared to chronic phase CML (143). In addition, constitutively activated LYN kinase (144) has been observed in lymphoid transformation of CML (145), suggesting a possible beneficial role for SFK inhibitors (146).

The second generation of kinase inhibitors includes dasatinib, which is the only dual SFK/*BCR-ABL* inhibitor approved for use on patients with imatinib resistant CML and Ph positive ALL. Dasatinib is structurally unrelated to imatinib and was originally designed as a SFK inhibitor but because the structure of *ABL* is closely related to SFKs, it was also found to have inhibitory effects on *BCR-ABL*, *KIT*, *PDGFR* and ephrin receptor kinases (147). Dasatinib can bind to the *ABL* kinase domain in many conformations including the active conformation to which imatinib cannot bind (148, 149). As a result, dasatinib is active against all imatinib resistant mutations except T315I and has a 325-fold increased specific activity compared to imatinib *in vitro* (148). Results from Phase I, II and III trials have shown that patients have good responses in both patients with imatinib resistant mutations (excluding T315I) in all phases of CML and in Ph positive ALL (150-152). It also has good efficacy for many patients who are imatinib resistant without detectable mutations (153). The reason for this is unknown, but could possibly relate to inhibitory activity of dasatinib against the SFKs. However, it is not known how long responses will last and indeed some resistant patients do not respond to any treatment modality. This is likely to be due to the fast evolution of the malignant clone as the result of the acquisition

of further genetic abnormalities conferring complex mechanisms of resistance that are incompletely understood.

In addition to dasatinib there is a new imatinib derivative called nilotinib, that is in Phase II trials (154) and which has recently been licensed in Europe. Nilotinib has a 30-fold greater potency than imatinib due to its better fit to the ATP binding pocket and is effective on all kinase mutants except T315I, however unlike dasatinib it has no SFK activity (155).

Both nilotinib and dasatinib are ineffective against the kinase domain mutant T315I, mainly because this mutation eliminates the critical H-bonds required for binding to the ATP pocket (141). Although relapse with T315I is relatively uncommon, other approaches to targeting BCR-ABL are being developed, such as GNF-2 which binds to the myristoyl binding site of BCR-ABL leading to inactivation (156). Third generation inhibitors are also being developed to target T315I mutant, such as the aurora kinase inhibitor MK-0457 (157) which can also inhibit FLT3, ABL and JAK2 (158). Early clinical trials suggest that MK-0457 is well tolerated and gives encouraging responses in ALL and in some patients with T315I (159).

Despite these therapeutic advances, in order to achieve full and permanent remission by chemotherapy, quiescent Ph positive stem cells also need to be targeted. These non-proliferating cells have been found in Ph positive CML as well as ALL and, importantly, that their quiescence is reversible (160). The presence of quiescent Ph positive cells in patients with chronic phase CML on imatinib has been demonstrated and may be relevant for subsequent relapse (161, 162) as neither imatinib, dasatinib nor nilotinib are capable of targeting primitive quiescent CML cells (163-165). Targeting of these cells may be essential for a true cure of CML, which can only be achieved currently by allogeneic stem cell transplantation (164, 166).

It is clear that the BCR-ABL fusion protein is a constitutively active tyrosine kinase that can concomitantly deregulate cell growth and apoptosis by activation of multiple intracellular signalling pathways. However, it has become apparent that ABL is not the

only tyrosine kinase that, when constitutively activated, leads to transformation and leukaemia. Indeed since the discovery of *BCR-ABL*, many other activated tyrosine kinases have been identified in MPNs, lymphoma and some solid tumours such as papillary thyroid carcinoma, gastrointestinal stromal tumours and lung cancer. Some of these activated tyrosine kinases are inhibited by imatinib, e.g. patients with fusion genes that involve the platelet-derived growth factor receptors alpha or beta (PDGFR α and β) (167). Unfortunately, not all activated tyrosine kinases are imatinib responsive but increasing numbers of new tyrosine kinase inhibitors are being developed that are active against a wider range of targets (158).

1.6 Other deregulated tyrosine kinases involved in MPNs

More than 40 tyrosine kinase fusion genes have been described in MPNs, most commonly involving PDGFR α , PDGFR β or FGFR1 (168). Such fusions have several features in common; the N-terminal portion of the fusion partner gene replaces the extracellular ligand binding domain of receptor protein tyrosine kinases (RPTKs) such as the PDGFRs and the N-terminal domain of non-receptor protein tyrosine kinases (NRPTKs) such as ABL. The chimaeric protein always retains the kinase domain but the loss of regulatory domains means that the kinase is no longer responsive to its normal ligands or normal negative regulators.

However, it is not only translocations that can cause the deregulation of tyrosine kinase genes in MPNs. Point mutations can also be responsible for constitutive activation of tyrosine kinases. For example, the KIT receptor is activated by point mutations in the majority of cases of systemic mastocytosis, a disease that is clearly myeloproliferative (169, 170) but classified separately by the WHO. More recently, a single acquired point mutation in the *JAK2* gene has been identified in approximately 95% of patients with PV, and 40 to 50% of patients with ET and PMF (171-173). The *JAK2* mutation occurs at nucleotide 1849 substituting a guanine with a thymine, changing the amino acid from a valine to a phenylalanine at codon 617 (V617F). This mutation occurs within the highly conserved auto-inhibitory JH2 domain. Although the JH2 domain is homologous to the true tyrosine kinase domain (JH1) the JH2 domain lacks the important catalytic domains and is actually thought to negatively regulate the tyrosine activity (174). Thus the V617F

mutation is thought disrupt this negative regulatory role of the JH2 domain and lead to constitutive activation of JAK2 and its downstream signalling molecules (175). Indeed, expression of JAK2 V617F in murine models gives rise to a disease similar to PV (171). The *JAK2* mutation is also found in a small number of patients with HES, AML, MDS and CMML (176-178) but is not apparent in lymphoid malignancies (177). The *JAK2* mutation is heterozygous in most cases of ET (179) but is homozygous in a substantial proportion of PV and PMF cases as a consequence of acquired uniparental disomy at chromosome 9p (180). It is thought that ET, PV and PMF represent a continuum in disease progression, much like the different stages of CML, with acquisition of further copies of *JAK2* V617F and other mutations leading to transformation (181). However, it still remains controversial as to whether *JAK2* V617F is the only mutation that is required to give rise to this spectrum of diseases or whether there are other, as yet unidentified, co-operating mutations that are also necessary (182).

Internal tandem duplications (ITD) of *FLT3* in AML are another example of leukaemia-associated mutations that increase tyrosine kinase activity. In general, tyrosine kinases that are activated by point mutations, ITDs or gene fusions all result in activation of various signalling pathways that provide the transformed cells with survival and proliferative advantages.

1.6.1 Platelet derived growth factor alpha (*PDGFRA*) gene fusions

One example of an alternative activated tyrosine kinase involved in MPNs is the type III RPTK *PDGFRα*, which has been linked to atypical CML when fused to the *BCR* gene (183) and it is the most common recurrent fusion in patients with IHES/CEL when fused to the *FIP1-Like 1* (*FIP1L1*) gene (184).

1.6.2 Diagnosis of Idiopathic Hypereosinophilic Syndrome (IHES)

Hypereosinophilic syndromes (HES) can arise from clonal disorders of the bone marrow or from a number of reactive conditions such as parasitic infections and allergies. The diagnosis of idiopathic HES (IHES) is one of exclusion and characterised clinically as a persistent state of unexplained eosinophilia for longer than 6 months, with evidence of

organ dysfunction due to eosinophilic tissue infiltration and the consequent release of their granules and toxic contents (170). IHES is predominantly a disease of men, with a male to female ratio of 9:1 and is usually diagnosed between the ages of 20 and 50 years (185). It is only once clonal proliferation of the eosinophils and its cause is ascertained, that cases of IHES may be reclassified as the MPN chronic eosinophilic leukaemia (CEL) (186, 187) and treated appropriately. Until relatively recently, as the majority IHES patients are cytogenetically normal, establishing the clonality of IHES was difficult and had only been proved in a small number of patients with cytogenetically visible lesions (188) or by X-inactivation assays in female patients (189).

1.6.2.1 Discovery of *FIP1L1-PDGFR α* as a major cause of IHES/CEL

The elucidation of the molecular lesion that gives rise to many IHES/CEL cases was a result of the observation that some of the treatments for CML were also useful for IHES and they may therefore share a common pathogenetic mechanism. Specifically, imatinib was found to be very effective in some IHES cases, implicating a constitutively activated, imatinib sensitive, tyrosine kinase gene, as the cause of the disease (190), for example ABL, KIT, PDGFRs (191) or an unidentified target.

This bedside-to-bench approach eventually led to the identification of a cytogenetically invisible interstitial deletion on chromosome 4q12 that fuses the *FIP1L1* gene to the tyrosine kinase *PDGFR α* gene 800 kb downstream. In the initial description, *FIP1L1-PDGFR α* was found in approximately half of all patients with IHES (184). Further evidence that this fusion gene was the cause of IHES/CEL in these patients arose through the identification of a resistance mutation (T674I) in FIP1L1-PDGFR α in a patient who relapsed during imatinib therapy (187). Tyrosine kinase domains are highly conserved and the T315I mutation in BCR-ABL that confers resistance to imatinib (192) is homologous to the T674I mutation in FIP1L1-PDGFR α . In addition, it was shown that the fusion has deregulated, constitutive tyrosine kinase activity and is capable of transforming haematopoietic cells (184, 186, 193), in a manner analogous to BCR-ABL. *In vitro* experiments have confirmed the constitutive activation of PDGFR α deregulates signalling pathways such as those involving STAT5 (184).

IHES has a very ambiguous clinical presentation and is often associated with considerable morbidity and mortality. Most *FIP1L1-PDGFR* positive patients show dramatic responses to imatinib therapy with rapid normalisation of peripheral eosinophil counts and achievement of nested or real time quantitative reverse transcription polymerase chain reaction (RQ-PCR) negativity (194-196). Identification of patients who are positive for *FIP1L1-PDGFR* is therefore critical for appropriate clinical management (197). Exclusion of *FIP1L1-PDGFR* is also important as some patients may benefit from alternative therapies such as the IL-5 antibody mepolizumab (198).

1.7 Mechanisms that give rise to translocations

It is widely accepted that acquired translocations are initiated after two double stranded breaks (DSBs) have occurred in pairs of non-homologous chromosomes (199), that are in close proximity to one another. Furthermore, it has been shown that the distances between genes common in translocations are a lot shorter than control genes during cell division in both myeloid (200) and lymphoid (201) malignancies. DSBs may be the result of a number of exogenous factors such as ROS, ionising radiation, genotoxic chemicals and DNA replication stress (202) or by the cell itself, as ROS are by-products of normal respiration. The aberrant recombination that follows represents a failed attempt by repair enzymes to re-ligate the severed DNA (203). The mechanisms behind many chromosomal translocations in lymphoid malignancies are believed to be the result of illegitimate recombination mechanisms, such as V(D)J recombination (204) or class switch recombination (CSR) (205, 206), that normally give rise to the generation of diversity of antigen receptors found in lymphoid cells required to mount an effective immune response.

In V(D)J recombination, the recombinase-activating gene (RAG) enzyme complex recognises specific recombination signal sequences (RSSs) at the variable (V), diversity (D) and joining (J) segments and cuts these segments which are then joined together by non-homologous end joining (NHEJ) early in B-cell differentiation (see Figure 1-6). The RAG enzyme complex can, however, mis-target recombination and juxtapose proto-oncogenes such as *BCL2* to loci that encode antigen receptors (207). These receptors contain regulatory regions that activate the proto-oncogene, giving rise to lymphoid malignancies. It is also possible for two non-antigen receptor loci to be joined erroneously

by illegitimate V(D)J recombination, such as fusions between SCL-interrupting locus (SIL) and stem cell leukaemia gene (SCL) in some patients with T-ALL, where the SCL promoter region is replaced with the SIL promoter resulting in the inappropriate expression of SCL (208). Indeed, at such translocation breakpoints, signal sequences suggestive of V(D)J recombination such as cryptic RSSs and non-templated ('N' region) nucleotide additions have been found (204).

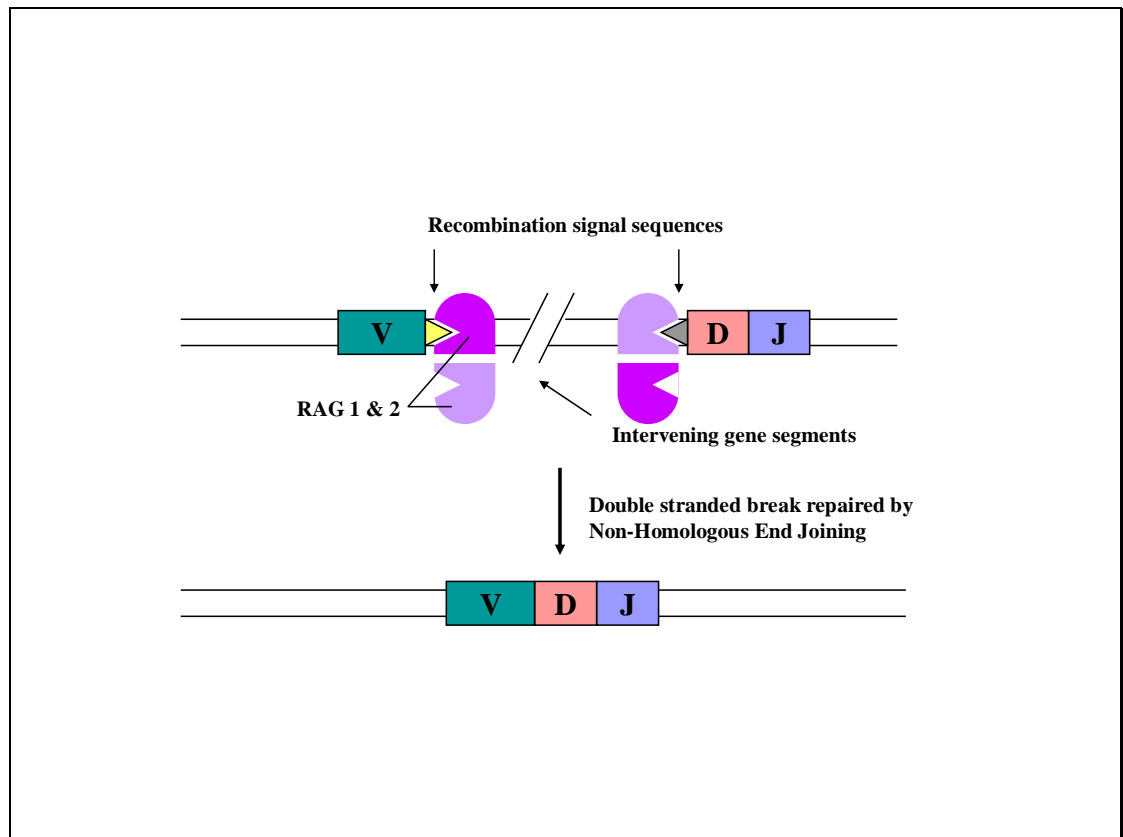


Figure 1-6: RAG mediated recombination. RAG 1 and 2 recognise specific recombination signal sequences (heptamers and nonamers) flanking the Variable (V), Diversity (D) and Joining (J) segments of antigen receptor gene loci. The RAG recombinase complex then makes double stranded breaks which are then repaired by non-homologous end joining.

CSR occurs later in B-cell differentiation and the first translocation of this type was identified over 20 years ago when the proto-oncogene MYC was found to be translocated to the immunoglobulin heavy chain (IgH) switch region in Burkitt lymphoma (209) and are thought to be the result of illegitimate activation induced cytidine deaminase (AID) activity (210). Many different translocations involving immunoglobulin switch regions have since been identified, particularly in multiple myeloma (205).

The mechanisms behind chromosomal translocations in myeloid and in the remainder of lymphoid malignancies, however, remain much more elusive. Broadly there are two main opposing hypotheses as to why some fusions, such as the t(9;22) are seen so frequently compared to other fusions such as those, for example, involving *FGFR1* or indeed, why recurrent fusions involve relatively few proto-oncogenes compared to the total number in the genome. The first is the ‘selection hypothesis’ which poses that such breaking and re-ligation events occur relatively frequently between chromosomes and that the gene fusions seen in leukaemia are a consequence of biological selection for oncogenic products. Any others which do not fulfil these criteria will never be seen and rapidly die out. Indeed the finding of *BCR-ABL* transcripts in healthy individuals suggests that *BCR-ABL* and other fusion genes are generated regularly but perhaps in a more committed progenitor without the necessary cell machinery for a selective growth advantage (211). Therefore only in-frame fusions in leukaemic stem cells may give rise to malignancy (160).

The second is the ‘hot-spot hypothesis’ which poses that the breaks are non-random and instead tend to occur at particular genomic regions as a result of mechanisms that use either sequence specific targets or sequence dependent secondary structures, which are therefore termed recombinogenic sequences (212). Sequence specific mechanisms include homologous recombination (HR) whereby translocations are mediated by large (at least 200 bp) repetitive sequences such as *Alu* sequences. These elements have been implicated in the generation of translocations in both myeloid (213) and lymphoid (214-216) Ph positive leukaemia, however as a relatively large proportion of the human genome is comprised of *Alu* repeat sequences the evidence is not hugely convincing.

Recombinogenic sequences include purine/pyrimidine repeat regions, which have been found in association with some ALL breakpoints (217, 218). Chromatin structural elements may also lead to chromosome rearrangements (219) and as such, scaffold and matrix attachment regions (S/MARs), have been implicated in mediating translocations such as those involving *MLL*, *AML1* and *AF9*. Breakpoint clusters within these genes have also been shown to be preferred cleavage sites during early apoptosis (220, 221) and therefore translocations could occur as a result of an aborted apoptosis event (222) in which cells

aberrantly repair cleaved DNA. In concordance with this, NHEJ as opposed to HR is thought to be the most likely mechanism for repair following DSBs, and indeed genomic junctions suggestive of NHEJ have been found in patients with balanced translocations with both lymphoid (223) and myeloid (224, 225) leukaemia.

As many as 50 DNA DSBs occur during each mammalian cell division (226) and it is postulated that impaired NHEJ or increased frequency of DNA DSBs could also lead to increased frequency of translocations. There is some evidence for this, including an increased risk of CML in patients exposed to radiation (227, 228) and also inherited conditions such as Nijmegen's breakage syndrome, which results from defects in the DNA DSB repair machinery, and is associated with an increased incidence of chromosomal rearrangement and cancer (229). It has also been found that mice deficient in components of the NHEJ machinery are prone to developing lymphoma (230). Features of NHEJ include short (1-6 bp) stretches of microhomology between the translocated chromosomes, short deletions and insertions and direct and indirect repeats at the breakpoints (224).

There is evidence to support both the 'selection' and 'hot-spot' hypotheses (221, 231), but it is likely that neither of these mechanisms is mutually exclusive and that some translocations may therefore be the result of more than one mechanism. It is, however, important to study the mechanisms behind translocations as it may allow the identification of individuals at risk and ultimately lead to the design of preventative measures.

1.7.1 Mechanism of the formation of the *BCR-ABL* fusion gene

The precise mechanisms responsible for the formation of the t(9;22) are unknown, partly because relatively few genomic breakpoints have been fully characterised. The reason for this is because of the very large regions within which breakpoints may occur, particularly within *ABL* but also within *BCR* for e1a2 fusions (see Figure 1-5). There are several possible candidate mechanisms for the generation of the t(9;22). The observation that the distances between the *BCR* and *ABL* genes are a lot shorter than control genes during cell division in haematopoietic stem cells (200) is probably a significant factor and implies that if breakages do occur in these regions then aberrant recombination might be significantly

increased due to the proximity to one another. This is further substantiated by the implication of *Alu* repeat sequences which may facilitate aberrant DSB repair and therefore be involved in the genesis of both Ph positive CML (213) and p190 ALL (214-216) which might mediate aberrant recombination if DSBs occurred during mitosis. Clearly, biological selection is important, for example it has been shown that ionising radiation can produce a variety of BCR-ABL fusion proteins in HL60 cells not naturally present in either CML or ALL (228). However, these factors do not account for all features of the BCR-ABL fusion protein. Specifically, why is the p190 variant common in ALL but very rare in CML? Why are breakpoints in CML and many cases of ALL generally limited to the M-bcr where there are many possible in-frame combinations of *BCR* and *ABL* that are known to be biologically active? It is possible that these anomalies are a consequence of the mechanism of recombination between *BCR* and *ABL*, and differences in that mechanism between CML and ALL.

1.8 Aims of this project

This project aims to explore and develop genomic translocation breakpoint analysis in order to pursue the hypothesis that the mechanisms of translocation may explain some of the unusual features of *BCR-ABL* described above. In addition, rapid breakpoint analysis may also be useful diagnostically. Thus the specific aims of this study are to:

1. Identify mechanisms that give rise to the Ph chromosome in ALL by characterising genomic DNA breakpoints in a cohort of patients.
2. Develop genomic breakpoint analysis to assist in the diagnosis and monitoring of IHES patients with *PDGFRA* fusions.

Chapter 2: Materials and methods

2.1 Patient Samples

Patient samples were either referred from hospitals throughout the UK for diagnostic analysis or provided by collaborators from abroad. Research on samples was approved by the Salisbury and South Wiltshire Research Ethics Committee in the study entitled ‘Mechanisms and consequences of tyrosine kinase activation in chronic myeloproliferative disorders and related conditions’ (LREC study number 05Q2008/6), either following informed consent or anonymisation.

2.1.1 Preparation of blood and bone marrow samples

Patient samples for processing were received as blood or bone marrow. Mononuclear cells (MNCs) were separated by density gradient centrifugation over LymphoprepTM (Axis Shield, Norway). Cell counts were then determined using a Neubauer type haemocytometer (BDH, Poole, UK). 2×10^7 cells were diluted in Hanks’ balanced salt solution (HBSS, Invitrogen, Paisley, UK) to a maximum volume of 25 ml. The cell suspension was layered of an equal volume of LymphoprepTM prior to centrifugation for 30 minutes at 1800 rpm (revolutions per minute). After centrifugation, the MNCs were removed from the interface using a sterile Pasteur pipette (BDH) and transferred to a clean tube. The recovered cells were washed twice in RPMI 1640 (Invitrogen, Paisley, UK) supplemented with L-glutamine, penicillin, and 10% fetal calf serum, before counting.

Granulocytes were obtained by discarding the supernatant before adding ice-cold red cell lysis buffer (see Appendix I for recipe) to the cell pellet. The cells were placed on ice for 10 minutes then spun for 5 minutes at 1600 rpm. The resultant cell pellet was washed twice in supplemented RPMI 1640 then counted. For DNA extraction at a later date, cell pellets of $1\text{--}2.5 \times 10^7$ cells were frozen at -70°C (see section 2.1.3). For RNA extraction, 1×10^7 cells were lysed with guanidine thiocyanate (GTC) mixed with $7 \mu\text{l}$ β -mercaptoethanol (Sigma, Poole, UK) per 1 ml GTC (see Appendix I).

2.1.2 RNA extraction

350 µl of GTC lysed cells for RNA extraction were mixed with an equal volume of 70% ethanol to provide the appropriate conditions for the RNA to bind to the silica-gel membrane, mixed, and applied to an RNeasy kit column (Qiagen, West Sussex, UK). The column was spun for 15 seconds at 10 000 rpm. The column was washed in accordance to the manufacturer's instructions and the RNA eluted with 20 µl deionised, RNase free water. RNA was then reverse transcribed into complementary DNA (cDNA).

2.1.2.1 cDNA synthesis

Prior to cDNA synthesis, the RNA was heated for 5 minutes at 65°C to destroy any tertiary structures and placed on ice. For cDNA synthesis, the RNA was mixed with 21 µl cDNA mix (50 mM Tris pH 8.3, 75 mM KCl, 3 mM MgCl₂, 1 mM DTT, 1 mM dATP, dTTP, dCTP, dGTP), 100 ng/µl random pd(N)₆ hexamers (Amersham Pharmacia, Amersham, UK), 15 units RNasin (Promega, Southampton, UK) broad spectrum ribonuclease inhibitor and 150 units of Moloney Murine Leukaemia Virus (MMLV) reverse transcriptase (Invitrogen, Paisley, UK) which synthesises cDNA from RNA. This reaction was incubated at 37°C for 2 hours before being terminated by heating the samples to 65°C for 10 minutes. All cDNA samples were stored at -20°C.

2.1.3 DNA extraction

DNA was salt extracted from stored patient granulocyte cell pellets. Cells were thawed prior to addition of 1 ml RSB (see Appendix I), 30 µl SDS (see Appendix I) and 200 µg proteinase K (Roche, UK) and incubated at 37°C overnight. 300 µl of 6 M NaCl was added to the cells and the cell debris was pelleted by centrifugation at 13 000 rpm for 30 minutes. The supernatant containing the DNA was carefully transferred into a clean tube to which an equal volume of 100% ethanol was added and the tube gently inverted to precipitate the DNA. The DNA was hooked out with a hyperdermic needle, washed in 70% ethanol and re-suspended in an appropriate volume of 1 x Tris EDTA (pH 7.4) before quantification of DNA using an ND-1000 spectrophotometer (NanoDrop) and storage at 4°C.

2.2 Fluorescence in situ Hybridisation (FISH)

FISH is the process by which DNA probes spanning the area of interest are labelled, hybridised and subsequently used to detect the presence of complementary DNA in cultured patient cells.

2.2.1 Culturing of clones for FISH

Ensuring aseptic techniques at all times, *E.coli* harbouring Bacterial Artificial Chromosomes (BACs; which span regions of approximately 200 kb) or Fosmid clones (which span smaller regions of approximately 40 kb) were streaked out onto the appropriate antibiotic plates (20 µg/ml chloramphenicol for BACs) and grown up overnight at 37°C. One colony for each clone was then picked and re-suspended in 30 ml LB broth (supplemented with 20 µg/ml chloramphenicol) and cultured in a flask overnight at 320 rpm and 37°C.

2.2.2 Extraction of BAC or Fosmid probes

On day one of this procedure, cultured cells were spun down for 10 minutes at 3000 rpm in a Sorvall centrifuge. Each pellet was then re-suspended in 1.8 ml of P1 (see Appendix I) and transferred to a plastic ultra-centrifuge tube. Slowly 1.8 ml of P2 (see Appendix I) was added and left for 5 minutes, this solution increases the pH, which denatures the bacterial DNA and the FISH or Fosmid clone DNA. 1.8 ml of P3 (see Appendix I) was then added and left on ice for 5 minutes. This lowers the pH, precipitating the denatured bacterial DNA and proteins, however, the supercoiled FISH or Fosmid clone DNA re-anneals and stays in solution. The tubes were then spun for 10 minutes at 10 000 rpm at 4°C. Once complete the supernatant containing the FISH or Fosmid clone DNA was cleaned by transferring the supernatant to 1.5 ml tubes containing 700 µl of chloroform and spun for 5 minutes at 13 000 rpm. The top layer was removed by pipetting and the DNA precipitated by transferral into a tube containing 800 µl of ice cold isopropanol and placed at -20°C overnight. On day two, the tubes were removed from -20°C and centrifuged at 13 000 rpm for 15 minutes. The pellets were then washed twice with 70% ethanol before air-drying and re-suspension in 40 µl of deionised water, ready for labelling.

2.2.3 Labelling

On ice, 1 µg of probe DNA was labelled by nick-translation in the presence of 0.015 U DNase I (Invitrogen) and 10 U DNA polymerase (Invitrogen). For biotin labelling, 5 µl of Bio mix (see Appendix I) was added, for digoxigenin labelling, in addition to the 5 µl of Dig mix (see Appendix I), 1 µl of 1:4 BSA (New England Biolabs, NEB, USA) and 25 µg of Digoxigenin were also added. All labelling reactions were made up to 50 µl with deionised water and incubated at 16°C for 90 minutes, reactions were then terminated with a denaturation step of 65°C for 10 minutes. To check the efficiency, labelling reactions were run on a 1.5% agarose gel (see section 2.4).

2.2.4 Probe precipitation

To prevent co-hybridisation of repetitive sequences such as *Alu* repeats, 10 µl of labelled DNA was co-precipitated with 10 µl of Cot-1 competitor DNA (Eur-X, Gdansk, Poland) in 40 µl 100% ethanol and 2 µl of 3M sodium acetate (pH 5.2) for -20°C for one hour. The DNA was then pelleted by centrifugation at 13 000 rpm for 15 minutes before washing with 70% ethanol, air drying and re-suspension with 12 µl hybridisation buffer (see Appendix I).

2.2.5 Slide preparation from fixed cell suspensions

Fixed cells were centrifuged and the pellet re-suspended in freshly made fix (3:1 methanol:acetic acid) solution. Cells were centrifuged again and the pellet re-suspended in as much fix as required to make the suspension clear. Microscope slides were prepared by using a Pasteur pipette to apply one drop of the suspension from about 30 cm above the slide. The slides were then allowed to dry in a humidity-controlled room, checked under a phase contrast light microscope and left at room temperature to age overnight.

2.2.6 Addition of probes to slides

Slides were prepared by soaking in 2 x SSC (sodium saline citrate to give 0.3M sodium citrate, 3M sodium chloride with pH 7.0, diluted from 20 x SSC from Sigma) for 30 minutes to 2 hours prior to incubation with 25 µg pepsin in the presence of 0.01M HCl for

10 minutes at room temperature in a moist chamber in order to remove excess proteins from the fixed nuclei that may interfere with hybridisation. The slides were then washed 3 times 5 minutes in 2 x SSC before being fixed in 1% paraformaldehyde (Sigma) in phosphate buffer solution (PBS) (Oxoid, Basingstoke, UK) for 10 minutes. Slides were washed a further 3 times for 5 minutes in PBS before dehydration through an ethanol series of 3 minutes in each of 70%, 90% and 100% ethanol respectively. The slides were then allowed to air dry prior to the addition of 12 µl of the appropriate probes, left on a 72°C hotplate for 5 minutes then placed in a sealed moist chamber and allowed to hybridise for 60 hours at 37°C.

2.2.7 Post hybridisation washes and detection

Post hybridisation washes of 3 times 5 minutes in 50% formamide/2 x SSC then 3 times 5 minutes in 2 x SSC were carried out in a 42°C water bath. Slides were incubated with 50 µl SSCTB (see Appendix I) for 20 minutes in a moist chamber at room temperature then washed 3 times in 4 x SSC. Successive antibody layers were then applied, each antibody was diluted in SSTB and incubated for 30 minutes under a cover slip in a dark moist chamber in the following order: (i) 1/500 sheep anti-digoxigenin FITC conjugated (Roche) (ii) 1/400 Texas red Avidin (Vector Laboratories Inc) and 1/200 Rabbit anti-sheep FITC (Vector Laboratories, Peterborough, UK) (iii) 1/100 Goat anti-Avidin biotinylated (Vector Laboratories) and 1/100 swine anti-rabbit FITC (Dako, Denmark) iv. 1/400 Texas red avidin. Between each antibody application the slides were washed 3 times for 5 minutes in SSCT (see Appendix I). After the final application, slides were washed for 5 minutes in SSCT then twice for 5 minutes in PBS before mounting under a cover slip in Vectorshield (Vector Laboratories) containing 1 µg/ml 4-6-diamidino-2-phenylindole (DAPI). Slides were then ready for examination.

2.3 The Polymerase Chain Reaction (PCR)

The PCR is a technique by which DNA is amplified *in vitro* by the use of two synthetic oligonucleotide primers that are complementary to two regions flanking the target DNA. Thermal cycling is then carried out whereby the target DNA is repeatedly denatured, the primers annealed and nascent strands extended from the primers, in conditions of excess

deoxynucleotides and a thermostable DNA polymerase. This enables exponential amplification, as the new strands act as a template for each cycle. Details of all primers used in this project can be found in Appendix II.

2.3.1 High Fidelity PCR (Roche)

On ice, PCRs were set up with 10 µl of Mix 1 (containing 2 µl High Fidelity buffer with 15 mM MgCl₂, 1 U Enzyme mix and 7.7 µl water) 7 µl Mix 2 (final concentrations of 200 µM dNTPs), 500 nM of forward and reverse primers and 1 µl DNA to give 400 ng DNA. For cycling programme see Appendix II.

2.3.2 Expand Long Range PCR (LR-PCR)

For detection and amplification of genomic DNA breakpoints, ranging from 300 bp to 12 kb the Expand Long Template PCR kit (Roche) was used. All primers were designed on Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) to be 23 bases long with a melting temperature of 66°C. PCRs were set up on ice with final concentrations of 500 µM dATP, dCTP, dTTP and dGTP, 300 nM forward primer, 300 nM reverse primer, 2.5 µl Buffer 2 (with a final concentration of 2.75 µM MgCl₂), 1 U of Enzyme Mix and 20 ng of patient DNA, made up to a final volume of 25 µl with deionised water. In all screening PCRs, controls using normal DNA were included with each primer combination and two PCRs using primer pairs to amplify up normal sequences were included to ensure the PCRs worked with all DNAs used. For cycling conditions see Appendix II.

2.3.2.1 Expand Long Range PCR Multiplex

As for section 2.3.2 except an appropriate forward or reverse primer mix containing 300 nM of each of the different primers was used, in place of either the single 300 nM forward or reverse primer respectively.

2.3.3 Bubble PCR

For the detection of some genomic breakpoints, a modified protocol of the allele-specific bubble PCR was used (as first described by Zhang *et al.*, 1995 (232), see Figure 2-1). The bubble oligo was made by annealing 50 μ M of the BUB-T and BUB-B oligos in 136 μ M Tris (pH 8), 10 mM MgCl₂ and 50 mM NaCl for 16 hours at 4°C and kept at this temperature at all times. Approximately 2 μ g of patient DNA was digested at 37°C overnight with the appropriate restriction enzymes (NEB, Ipswich, USA) in a total volume of 10 μ l. 190 μ l of distilled water and 200 μ l of phenol:chloroform was added, which was then centrifuged for 5 minutes at 14 000 rpm at room temperature. The top layer was removed into a clean tube and precipitated in two times the volume of 100% ethanol, one tenth the volume of 3M sodium acetate and 40 μ g of glycogen (Invitrogen) then placed at –70°C for 20 minutes. The precipitate was pelleted by centrifugation for 30 minutes at 14 000 rpm and washed twice in 500 μ l of 70% ethanol. The pellet was then left to air dry before re-suspension in 10 μ l distilled water. On ice, 2 μ l of the digested DNA was ligated to 10 μ mol bubble oligo with 1 unit of T4 DNA Ligase (Invitrogen) then incubated at 15°C overnight. To denature the ligase, the samples were heated at 65°C for 5 minutes then excess unligated bubble oligos were removed by using the MinElute® PCR Purification kit (Qiagen). The PCRs were then performed with the appropriate reverse primers and the NV1 and NV2 forward primers for first and second step PCRs respectively. For PCR and cycling conditions, see section 3.2.4.3 for Long Range bubble PCR and section 4.2.6 for *PDGFRA* bubble PCR.

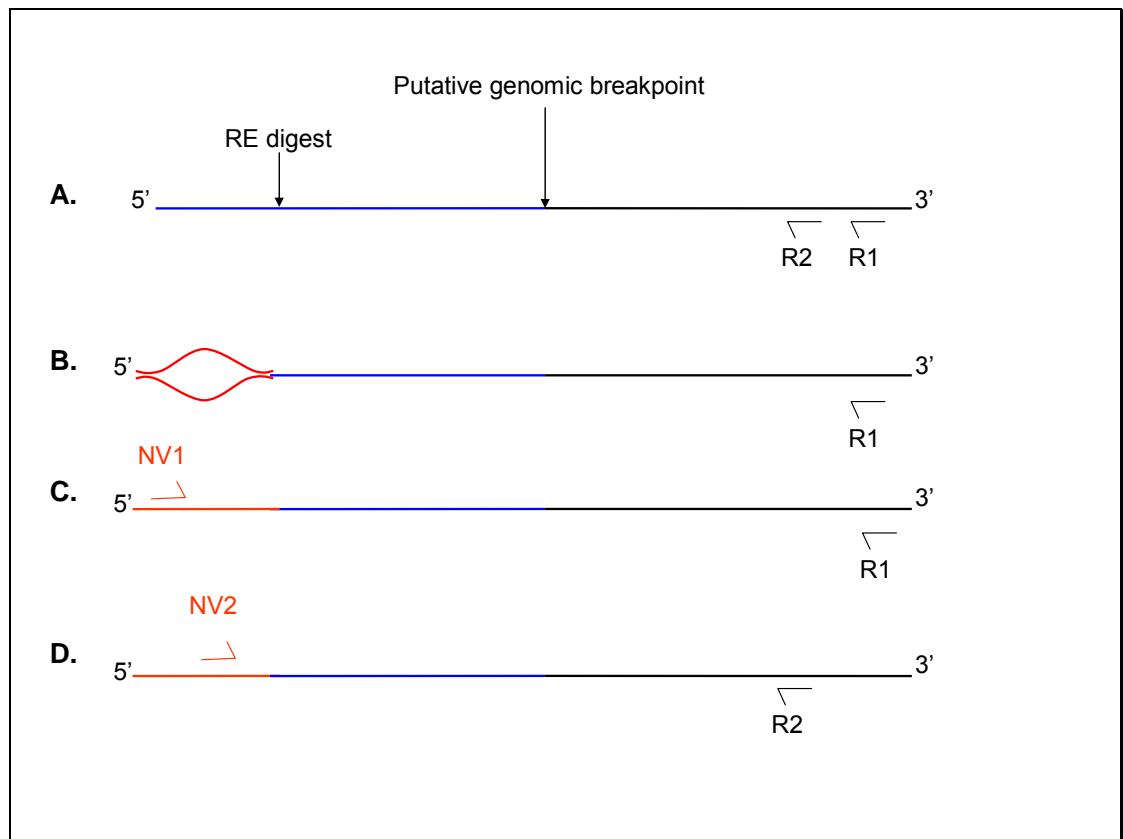


Figure 2-1: Bubble PCR.: (A). Digestion of putative genomic breakpoint with restriction enzymes. (B). Ligation of the bubble oligo and subsequent single stranded PCR with an allele-specific reverse primer R1 (C) to give rise to a unique site for the forward primer nvamp-1 (NV1) to anneal (D) to allow allele-specific amplification of the fusion gene by nested PCR.

2.4 Agarose Gel Electrophoresis

Gel electrophoresis allows separation of DNA by differential migration through the agarose according to size. DNA is negatively charged and when an electric current is applied, it will migrate towards the positive anode. The smaller molecules are able to move more quickly through the polymer than the larger molecules. Ethidium bromide is added to the gel to visualise the DNA, as the positively charged ethidium ion intercalates between the bases of the DNA and fluoresces under ultra violet (UV) light. As the putative breakpoint products of long range PCR can be any size from about 300 bp to 12 kb long, 1% gels were used. Gels were made by heating 0.5 g of agarose in 50 ml 1 x tris-borate (1 x TBE) buffer (Sigma) in a microwave, the gels were then cooled before addition of 0.2 mg/ml of ethidium bromide and then cast into the appropriate gel tank with the correctly sized combs. Before loading of PCR products the gel was immersed in TBE buffer and the

combs removed. 10 µl of PCR product was mixed with 2 µl of 6 x Orange-G Loading Buffer (see Appendix I) on parafilm then loaded. The orange-G enables monitoring of the approximate location of the bands and the high sucrose concentration sinks the DNA into the wells. To allow correct sizing of bands of all gels, a lane at either end of the comb was loaded with 1 kb plus ladder (Invitrogen) (see Figure 2-2). Gels were run at 80 volts for approximately 20 minutes and bands visualised using a UV transilluminator.

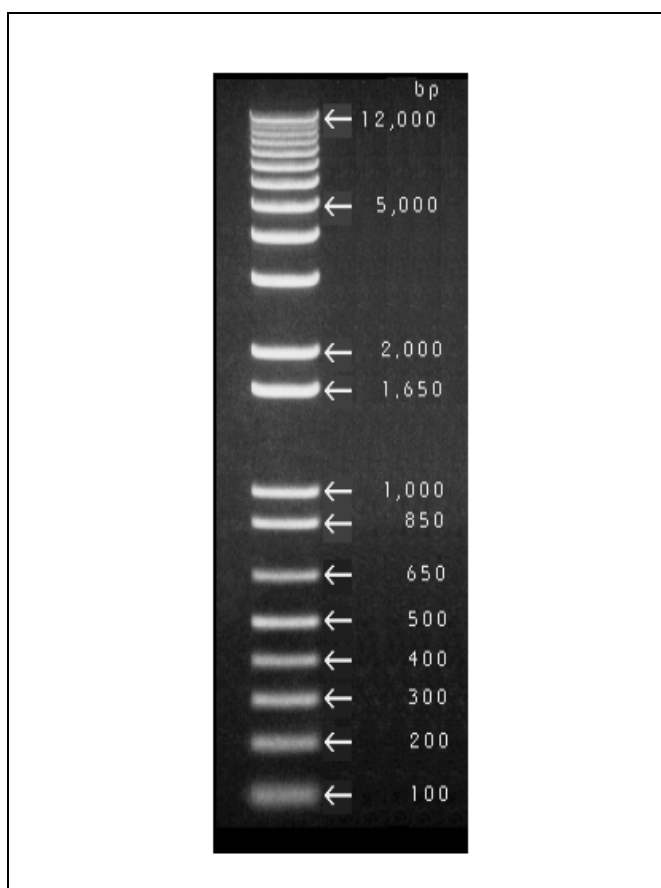


Figure 2-2: Invitrogen 1 kb plus DNA ladder sizes. All gel products were run alongside this ladder unless otherwise stated in the figure legend.

2.5 Cloning of PCR products

Fragments less than 4 kb are suitable for cloning with the TOPO TA Cloning kit for sequencing (Invitrogen).

Taq polymerase has template independent terminal transferase activity, which adds a deoxyadenosine (A) to the 3' ends of the PCR products. The linearised pCR®4-TOPO vector has single overhanging 3' deoxythymidine (T) residues which allows the PCR products to ligate efficiently.

Selection of positive colonies is based on the acquisition of the *lacZ* gene, derived from the *E.coli lac* operon. The *lacZ* gene is carried on the plasmid vector and codes for the enzyme β -galactosidase. X-gal (5, bromo-4-chloro-indolyl-beta-thio-galactoside) is an artificial, chromogenic substrate for β -galactosidase. Transformed bacteria containing the religated vector lacking the desired insert, when plated onto X-gal and agar form blue colonies as these cells will contain a functional *lacZ* gene, which is able to convert the colourless substrate X-gal into a blue product. If successful ligation of the PCR product takes place into the reading frame of the *lacZ* gene, it is disrupted and there is no production of active enzyme and the colonies remain white, in a process known as insertional inactivation.

2.5.1 Ligations

1 μ l salt solution (1.2 M NaCl, 0.06 M MgCl₂), 1 μ l fresh PCR product and sterile water to a final volume of 5 μ l were mixed before addition of 1 μ l of pCR4 vector. The ligation was carried out at room temperature for 5 minutes before being placed on ice to await transformation.

2.5.2 Chemical transformation

For each transformation, one tube of TOP10 *E.coli* chemically competent cells provided with the kit were removed from storage at -80°C and thawed on ice. 2 μ l of the cloning reaction was added to the cells, gently mixed and incubated on ice. After between 5 – 60 minutes the cells were heat shocked for exactly 30 seconds at 42°C and placed immediately back on ice for 2 minutes. 250 μ l of room temperature SOC medium (2% tryptone, 0.5% yeast extract, 10 mM MgSO₄, 20 mM glucose) was aliquoted into the tube and the cells were shaken at 37°C at 225 rpm to recover. After recovery for an hour, the transformed

cells were spread onto pre-warmed (to 37°C) Luria Bertani (LB) agar plates (see Appendix I) supplemented with 50 µg/ml ampicillin and pre-spread with 30 µl X-gal (50 mg/ml).

Once the liquid was absorbed the plates were inverted and placed at 37°C overnight. After incubation, plates were placed at 4°C to allow the colour to develop.

2.5.3 Culturing and plasmid extraction

Only white colonies were picked and each colony was grown in 2 ml LB broth (see Appendix I) 50 µg/ml ampicillin, overnight at 37°C and 320 rpm. Plasmid DNA was extracted the following morning from the cultured cells using the QIAprep spin Miniprep kit (Qiagen) according to the manufacturer's guidelines. Similar to the extraction of FISH probes described in section 2.2.2, the principle of the procedure is based on alkaline lysis of the cells. The high [OH]⁻ lyses the bacterial cells, denatures the bacterial DNA, but leaves the supercoiled vector DNA unaffected. The lysate is then neutralised and adjusted to high salt conditions, which precipitates the bacterial DNA and proteins leaving the plasmid DNA in the supernatant upon centrifugation. The supernatant is then applied to the silica column upon which the plasmid DNA is selectively absorbed where it can be washed with ethanol and eluted with water. Plasmid DNA was eluted with 30-50 µl of deionised water.

Plasmid DNA extracts were checked for the presence of the correct inserts by agarose gel electrophoresis of EcoR1 digested plasmid DNA (see section 2.6) and then directly sequenced using M13 forward and M13 reverse primers (see section 2.7).

2.6 Restriction enzyme digests

Restriction enzymes recognise specific nucleotide sequences that are usually 4 bp to 8 bp long, they then cleave the DNA at these sites. 1 U of enzyme was used to digest approximately 1 µg of DNA in the appropriate buffers and then incubated at 37°C overnight, prior to running on an agarose gel.

2.7 Sequencing by dideoxynucleotide chain termination

The dideoxynucleotide chain termination method is a PCR based technique for sequencing that utilises the addition of all four fluorescently labelled dideoxynucleotides (ddNTPs) to homogeneous, single stranded template DNA with the correct primer, DNA polymerase and deoxynucleotides (dNTPs). The ddNTPs lack the 3' hydroxyl group, necessary for the formation of 3' – 5' phosphodiester bond and therefore halts chain elongation at random when incorporated into the synthesised template DNA. The resultant mixture is a series of DNA molecules at different lengths which, when run through the capillary gel electrophoresis separate on the basis of size and are detected when the laser excites the fluorescent dye attached to the ddNTPs.

2.7.1 ExoSAP

If direct sequencing of PCR products was performed, removal of unused primers was performed by incubating 2 µl of the PCR product in 4 U of exonuclease (NEB) and 0.8 U of shrimp alkaline phosphatase (Roche) (made up in a ratio of 1:4) at 37°C for 15 minutes and then denatured by heating to 80°C for 15 minutes. The DNA from this reaction was then a suitable template for sequencing.

2.7.2 Sequencing reaction

All sequencing was performed using 50 ng of the appropriate forward or reverse primer, 1 µl of 5 x v3.1/v1.1 reaction buffer (Applied Biosystems, Foster City, California, USA) and 0.5 µl of v1.1 BigDye Terminator (Applied Biosystems, USA) with template DNA from either Exosap PCR product, extracted plasmid DNA or gel extracted DNA in a total volume of 10 µl (see Appendix II for cycling conditions).

2.7.3 Removal of unincorporated dye terminators

DyeEx 2.0 spin columns (Qiagen) or Montage sequencing reaction clean-up kits (Millipore, Watford, UK) were used for removal of un-incorporated dye terminators, depending on the numbers to be cleaned up. Deionised formamide or deionised water was

used to re-suspend the cleaned sequencing reactions, which were then loaded onto an Applied Biosystems 3100 or 3130 sequencing machine.

Chapter 3: The mechanism of formation of *BCR-ABL* in ALL

3.1 Introduction

Acute lymphoblastic leukaemia (ALL) is characterised by the over production and accumulation of immature lymphoid cells that are devoid of function. The Philadelphia (Ph) chromosome is present in 30% of adult ALL (24) and 3-5% of childhood ALL (25) and its presence indicates a worse prognosis than Ph negative ALL (24, 233, 234).

As detailed in section 1.4.8, in all Ph positive leukaemia, the breakpoints in *ABL* invariably occur anywhere in the ~150 kb region 5' of exon 3 (120) whereas the breakpoints in *BCR* are more variable (see Figure 1-5). Nearly all CMLs and 30-50% of Ph positive ALLs have breakpoints in the M-bcr (e14a2 or e13a2 fusions) that give rise to a p210 fusion protein. A very small number of CMLs and 50-70% of Ph positive ALLs have breaks in the m-bcr (intron 1 of *BCR*), giving rise to the e1a2 fusion and a p190 BCR-ABL fusion protein (24). These differences lead to the supposition that these two fusion proteins might be disease specific, or at least disease-related, and possibly transform different target cells (18, 235).

In order to try and elucidate the basis and mechanism for the different BCR-ABL isoforms, it is first necessary to understand how and at what stage in the development of the cell they arise. In CML, it is thought that the transforming event takes place in a multipotent stem cell, inferred from the range of cell lineages involved (236). In Ph positive ALL, it is still of some debate as to whether the acquisition of the Ph chromosome occurs in a multipotent stem cell or a committed lymphoid progenitor. Indeed, these two hypotheses need not be mutually exclusive as Ph positive ALL may in fact be two diseases, covert CML presenting at blast phase after a silent chronic phase transforming very early to lymphoid disease which would be of multilineage origin and lymphoid restricted *de novo* ALL (237). Therefore the p190 fusion protein, found rarely in CML, but the majority of ALL (238) would give rise to *de novo* ALL whereas p210 would be more likely to be covert CML and therefore of stem cell origin. This would account for the over representation of Ph positive adult ALL compared to Ph positive childhood ALL, as CML is known to be primarily an adult disorder and therefore those adult ALL patients with a silent chronic phase,

presenting at blast phase may be misconstrued as true *de novo* ALL patients, rather than CML patients.

However, there is evidence both for and against this hypothesis of disease specific fusion proteins. Some of the evidence for a multipotent stem cell origin of p190 ALL, and therefore against disease specific fusion proteins, includes (i) the observation that human p190 Ph positive cells are capable of inducing the same characteristics as leukaemic stem cells in non-obese diabetic mice with severe combined immunodeficiency (239) and (ii) the use of interphase FISH to show the presence of the p190 Ph chromosome in granulocytes of ALL patients (240). Conversely, in studies into the acquisition of further genetic abnormalities of over 500 cases of Ph positive CML, *de novo* ALL and CML in lymphoid transformation, the 3 types showed significant differences in karyotype at diagnosis, disease progression and relapse, suggesting different modes of disease initiation and progression (241). It has also been shown that some p190 ALL patients have no detectable myeloid involvement, indicating a more restricted lineage (242). Studies into the origin of Ph positive childhood ALL have shown that the Ph positive cells originate in primitive lymphoid progenitor cells without myeloerythroid developmental potential (243). Indeed, Castor *et al.*, (2005) (244) used fluorescent-activated cell sorting (FACS) to study the origins of ALL and found that p210 and p190 ALLs fell into 2 distinct biological entities. The p210 ALL fusions arose from multipotent haematopoietic stem cells while p190 ALL fusions with no myeloid involvement arose from B-cell committed progenitor cells.

In vitro and *in vivo* models have also indicated that the reason that the p190 fusion is seen in a much higher proportion of adult ALLs compared to CMLs is a result of when the translocation occurs during development. Indeed, murine models have demonstrated that both p190 and p210 can give rise to a similar polyclonal CML like disease when stem cells are targeted for transduction but when committed progenitor cells are targeted for transduction monoclonal, B-ALL results (245). The p190 BCR-ABL fusion protein has been demonstrated to have a higher intrinsic tyrosine kinase activity in lymphoid cell lines and can transform them to growth factor independence quicker than p210. The reason for this difference is thought to be p190s exclusive up regulatory effect on STAT6 (93).

However, both p210 and p190 can give rise to B-ALL (246) but in myeloid cells there is no apparent difference in their transforming potencies (245). It may therefore be argued that the rarity of p190 in human CML is not due to a lack of transforming potency but rather that *BCR* intron 1 breaks may be much more frequent in ALL because they are formed by a lymphoid-specific mechanism (245).

It is therefore plausible that a mechanism encompassing both physical and biological selection causes the genesis p210 and p190 BCR-ABL. One such mechanism that may explain the genesis of e1a2 breakpoints in ALL could be the action of RAG recombinase, which would only be present in lymphoid restricted cells. The RAG enzyme complex is responsible for a very specific mechanism of recombination. RAG recognises specific heptamer and nonamer RSSs at the V, D, and J segments and cuts these segments which are then joined together by NHEJ (see Figure 3-5). Indeed, sequence specific signatures suggestive of NHEJ following double stranded breaks have been found at the breakpoint junctions at a variety of recurrent translocations in both myeloid (224, 247) and lymphoid malignancies (248), indicating NHEJ as a common mechanism of aberrant fusion rather than HR.

Methylation is a known inhibitor of V(D)J recombination (249-251) and cell lines with high levels of RAG activity, such as the Reh cell line, retain stable karyotypes over long periods of time due to high levels of methylation (252). This maybe important as in a study of promoter methylation differences in *ABL*, all p190 ALL samples were found to have no methylation, whereas p210 CML and p210 ALL samples were found to have *ABL* promoter methylation (253). This further supports the idea that RAG could be pivotal in the genesis of p190 breaks.

It has even been shown that RAG can misrecognise sequences that have no heptamer or nonamer sequences. For example, in the *BCL-2* major breakpoint region in which the majority of breakpoints are found in follicular lymphomas, a common non-B DNA structure is found in 20-30% of human alleles and is more prone to be misrecognised by RAG even though no heptamer or nonamer signals are present. Thus, there is a higher rate

of recombination within the *BCL-2* M-bcr region relative to other *BCL-2* regions in a RAG expressing cell line but not in cell lines unable to express the RAG complex (254).

3.1.1 Aims

The aim of this part of the study was to try and understand mechanisms that give rise to the Ph chromosome by focusing on the characterisation of genomic breakpoints in Ph positive ALL in order to ascertain if the initial transforming event occurs in multipotential or lineage restricted haematopoietic cells. This has not been done previously because of the difficulty in characterising translocation breakpoints that fall in such large genomic regions. The hypothesis that breakpoints may be restricted to the lymphoid lineage as a result of RAG misrecognition was then tested using functional assays to determine if any of the breakpoints contain cryptic V(D)J RSSs, by cloning the sequences into an appropriate vector and subsequently co-transfecting them with RAG1, RAG2 and TdT into mouse NIH-3T3 cells. This assay gives quantitative results which were then compared with controls (p190 ALL versus p210 CML and p210 ALL). Additionally, differences in clustering of breakpoints in *ABL* were also looked at between the three types of breakpoints.

3.2 Materials and Methods

3.2.1 Patient Samples

All samples were confirmed to have the e13a2, e14a2 or e1a2 *BCR-ABL* fusion mRNA by multiplex RT-PCR (255) after a clinical diagnosis of adult ALL or CML and usually also the detection of the Ph chromosome by conventional cytogenetics or the *BCR-ABL* fusion by FISH.

3.2.2 Whole Genome Amplification (WGA)

WGA was performed on samples where only small quantities of DNA were available for screening, using the GE Healthcare (Amersham) GenomiPhi DNA Amplification kit. This method enables production of microgram quantities of genomic DNA (gDNA) up to 10 kb

in length, from as little as 1 ng of gDNA and combines Phi29 DNA polymerase, random hexamer primers, dNTPs and input DNA to amplify DNA representative of the original sample by utilising strand displacement amplification.

For one reaction, on ice, 1 µl (approximately 10 ng) of sample DNA was dispensed into 9 µl of GenomiPhi sample buffer in a thin walled PCR tube. This was denatured for 3 minutes at 95°C and placed immediately back on ice. In a separate tube, also on ice, 1 µl of enzyme mix was added to 9 µl of GenomiPhi reaction buffer. This was gently mixed and all 10 µl was dispensed into the tube containing sample DNA and sample buffer. The reagents were then incubated at 30°C for 16 hours after which the enzyme was denatured by heating the reaction to 65°C for 10 minutes. The amplified DNA was quantified on the NanoDrop ND-1000 spectrophotometer, run on an agarose gel and checked for quality by use in LR-PCR to amplify normal sequences of up to 10 kb (see section 2.3.2).

3.2.3 FISH to sub-localise the p190 breakpoint

For the purposes of this project, small fosmid DNA probes of approximately 40 kb were used to delineate the breakpoint region in the p190 patients for whom suitable cytogenetic suspensions were available. See section 2.2 for general materials and methods, except probe precipitation (see section 2.2.4) where optimisation experiments showed 20 µl of labelled DNA gave much stronger signals for these much smaller probes. Initially, 7 fosmid probes were ordered from the Sanger Institute (see Table 3-4) and tested on normal metaphases to optimise the technique and ensure hybridisation to the correct regions. For breakpoint analysis, the best 5 probes were selected; 3 fosmid probes spanning the *ABL* breakpoint region and 2 probes spanning the *BCR* intron 1 (see Figure 3-1).

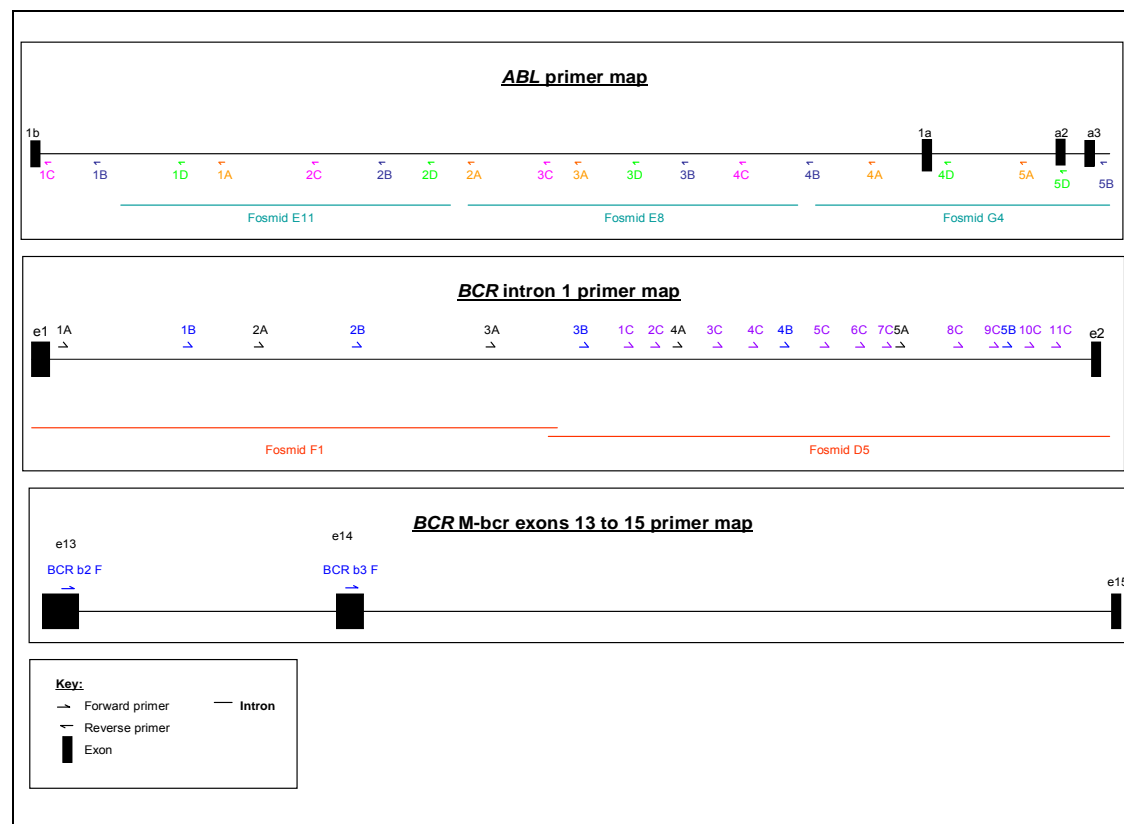


Figure 3-1: Positions of the fosmids used in FISH and the primers used in screening for breakpoints (to scale) in p190 ALL (intron 1 primer map) and p210 ALL and CML (M-bcr primer map). *ABL* primers used for both p190 and p210 samples are shown on the *ABL* primer map.

3.2.4 PCR-based techniques

3.2.4.1 High Fidelity PCR

See section 2.3.1. For ease of use, *ABL* was split into 5 sections designated 1-5 (using NM_007313 starting from the ATG the designated sections were as follows: ABL 1 = 129-30877; ABL 2 = 30878-63340; ABL 3 = 63341-94401; ABL 4 = 94402-124341; ABL 5 = 124342-141000) and *BCR* intron 1 was split into 3 sections designated Bt 1 to Bt 3 (using NM_004327 starting from the ATG, the designated sections were as follows: Bt 1 = 1280 (start of intron 1)-19454; Bt 2 = 19455-43454; Bt 3 = 43455-72904 (just passed the start of exon 2)). All primers designed were therefore labelled and numbered according to their section of origin, i.e. ABL 3 (19522) F would anneal to base pair number 19522 in *ABL* section 3. The same primers and numbering systems were used for LR-PCR.

3.2.4.2 Expand Long Range PCR (LR-PCR)

See section 2.3.2. For detection and amplification of genomic ALL and control CML breakpoint containing fragments ranging from 300 bp to 12 kb, the Expand Long Range PCR kit (Roche) was used. Primers were designed to span *ABL* and *BCR* regions within which the breakpoints in Ph positive ALL and CML are known to occur (see section 3.2.4.1). Before screening, however, the PCR was extensively optimised using normal genomic DNA and each primer was tested with its appropriate normal partner (see Appendix II for all primer lists and sequences) to yield a product of approximately 6-10 kb to ensure that all primers worked at the same standard conditions.

3.2.4.3 Long Range Bubble PCR

A modified protocol for the allele-specific bubble PCR was assessed for use in genomic breakpoint detection for p190 patients that were negative by multiplex LR-PCR. Long Range bubble PCR was performed as described in section 2.3.3 with nested PCR performed on the appropriately digested, bubble-ligated and cleaned DNA, using the bubble specific forward primers NV1 and NV2 with *BCR* reverse primers R1 and R2 that were designed at the 3' end of each specific *BCR* digested fragment using the LR-PCR protocol (see section 2.3.2).

3.2.4.4 Long distance inverse PCR (LDI-PCR)

LDI-PCR is a technique that has been used to amplify long breakpoint fragments in B-lymphoid cell malignancies (256) and which was adapted to amplify genomic breakpoints in multiplex LR-PCR negative p190 patients. Genomic DNA was digested as in section 2.6, which was then purified according to the bubble protocol (see section 2.3.3) except for the ligation step where the DNA was re-suspended in 179 µl of distilled water with 3 Units of T4 Ligase in 1 x Ligase buffer and incubated at 15°C overnight. This high dilution factor is necessary to allow the digested fragments to self-ligate. Nested LR-PCR was then performed (see section 2.3.2) using the correct *BCR* primers according to the digest performed (see Figure 3-2). PCR products were run on an agarose gel, the bands extracted and sequenced.

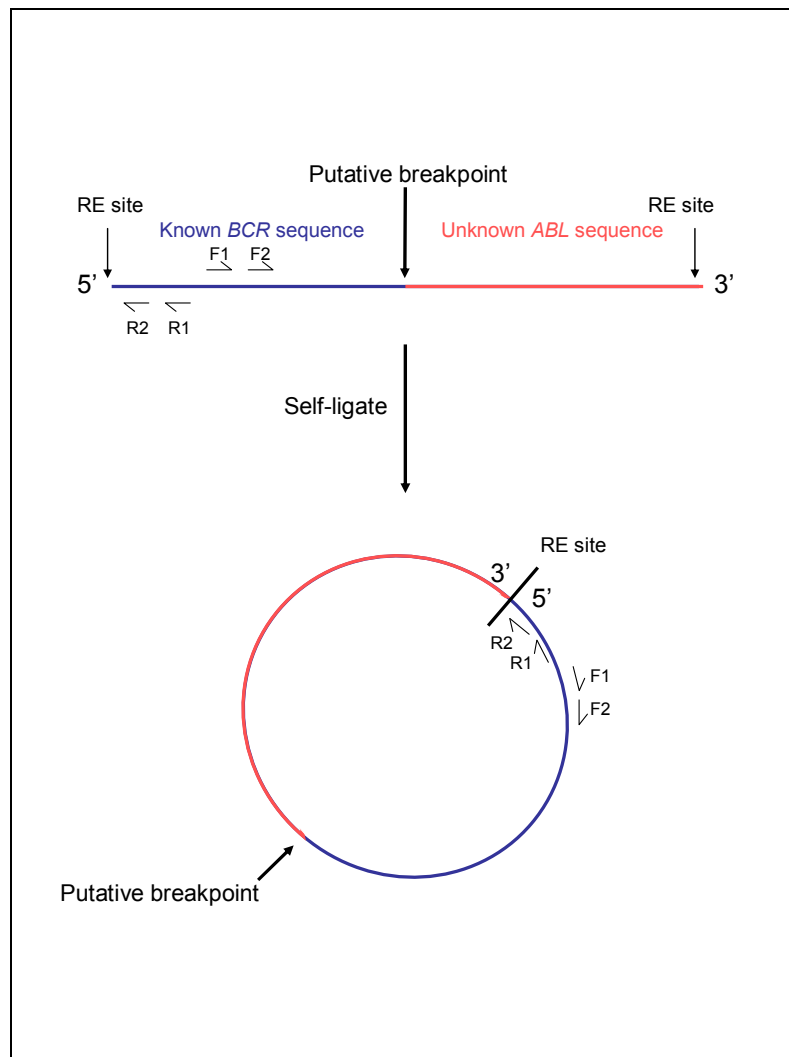


Figure 3-2: Long Distance Inverse PCR. Digestion of putative BCR-ABL genomic breakpoint followed by self-ligation and nested LR-PCR with primers F1 + R1 and F2 + R2, respectively. Primers were designed at the 5' end of the known BCR sequence to amplify into the unknown ABL sequence and across the putative BCR-ABL breakpoint.

3.2.5 Multiplex ligation-dependent probe amplification (MLPA)

MLPA quantifies target sequences by hybridising 2 adjacent complementary oligonucleotide probes which are then ligated and subsequently amplified in a multiplex PCR reaction using unique primer sequences that are tagged onto the probes. Only probes that hybridise and ligate will be amplified in the PCR. The primers are fluorescently labelled (FAM labelled), allowing the differentially sized products to be analysed by capillary gel electrophoresis. The area under the peak of each product relative to other peaks and normal controls reflects the copy number of the target sequence, allowing changes such as deletions and/or duplications to be identified (see Figure 3-3).

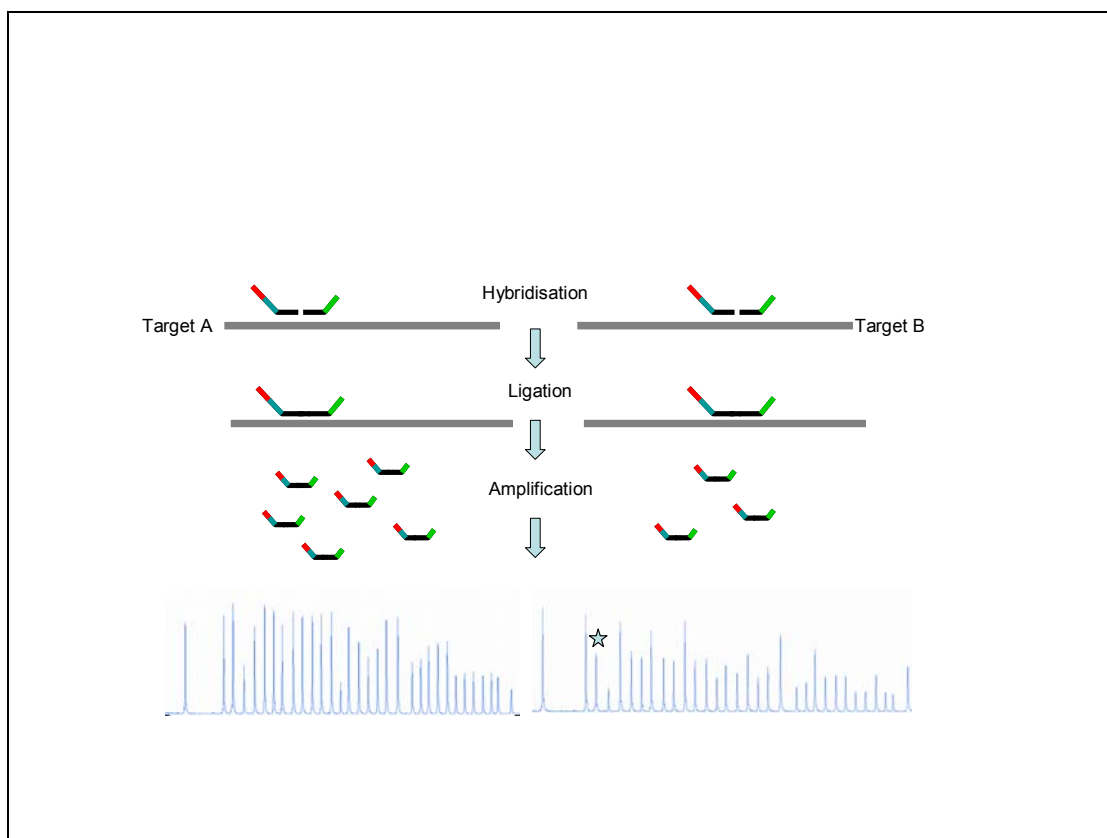


Figure 3-3: A schematic representation of Multiplex Ligation-dependent Probe Amplification (MLPA). Target A represents a normal MLPA trace with the probe represented in black, the 'stuffer' sequence in grey (variation in the size of the stuffer is used to produce final products of defined sizes), and the primers are shown in red and green. The 5' primer (red) is fluorescently labeled. The probes are hybridised to the target sequence, ligated together and finally amplified. The amplified products are analysed on an ABI 3100 analyser and a representative trace is shown at the bottom. Target B represents a control where some of the peaks are deleted, the star indicating an example of a reduction in peak height compared to the normal.

MLPA was performed to detect deletions on the reciprocal 9q+ chromosome in patients for whom the forward *BCR-ABL* breakpoints were characterised, but the reciprocal breakpoints could not be amplified by LR-PCR. This was performed as described by Kreil *et al.*, (128), who designed oligonucleotide probes to specific, loci flanking the 9q+ *ABL-BCR* breakpoint region (see Figure 3-4), each probe set generating products of a unique size (as first described by Schouten *et al.*, (257)).

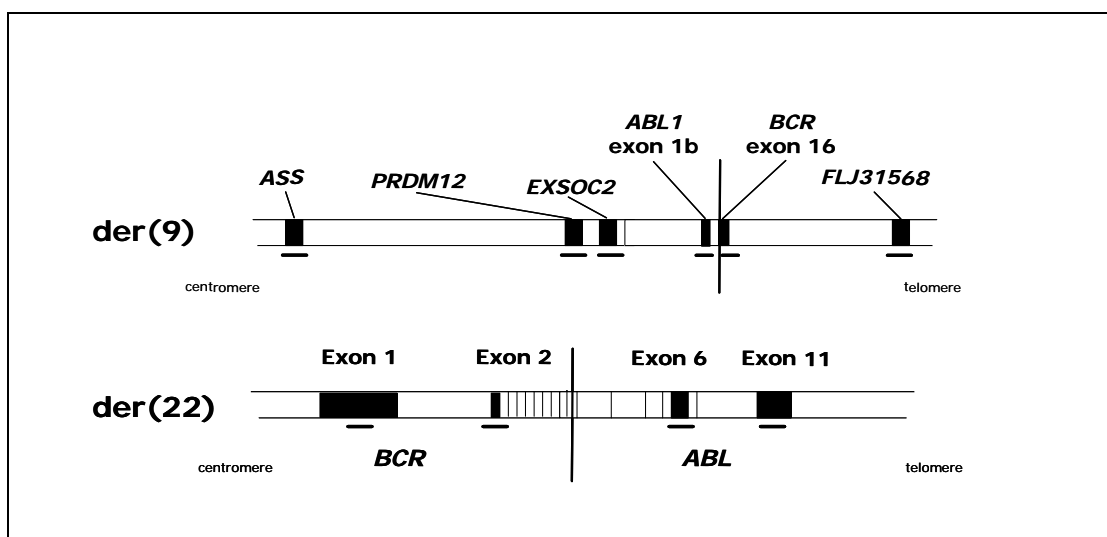


Figure 3-4: Positions of the MLPA probes (indicated by horizontal lines) used to detect reciprocal 9q+ ABL-BCR breakpoint deletions.

To perform MLPA, 2 µl (containing approximately 50 ng genomic DNA) of sample was added to 3 µl of 0.1 x Tris EDTA and heated at 98°C for 5 minutes and then cooled to 25°C, to denature. For hybridisation of the probes, 1.5 µl of probe mix (containing 1.33 fmol of each probe) was mixed with 1.5 µl of MLPA buffer and added to the denatured DNA at 25°C and incubated at 60°C for 16 hours. For the ligation reaction, the temperature was reduced to 54°C whilst the ligation mix (3 µl Ligase-65 buffer A, 3 µl Ligase-65 buffer B, 25 µl deionised water and 1 µl Ligase-65) was dispensed, and then incubated at 54°C for 15 minutes. The ligase was then inactivated by heating for 5 minutes at 98°C. Once completed, 5 µl of the MLPA ligation reaction was added to the PCR mix (2 µl Selective Adaptor Ligation, Selective Amplification SALSA PCR buffer, 1 µl of the SALSA PCR primers, 1 µl of SALSA Enzyme Dilution buffer, 0.25 µl SALSA Polymerase and 15.75 µl deionised water) and amplification was performed for 35 cycles (30 seconds at 95°C; 30 seconds at 60°C, 60 seconds at 72°C) followed by a final incubation for 20 minutes at 72°C.

For analysis, 1 µl of sample DNA was added to 8.9 µl Hi-Di Formamide (Applied Biosystems) and 0.1 µl GeneScan Rox 500 (Applied Biosystems) added as a size standard, in a new 96 well plate and analysed on the ABi 3100. Using Genotyper version 2.0, results were interpreted and the peak areas calculated. These results were exported into an Excel spreadsheet which calculates the ratios between all probe peak areas for each individual.

For normal sequences a ratio of 1.0 is expected, however if a deletion or duplication is present the ratio should be 0.5 or 1.5 respectively. Reactions were performed in duplicate with 2 normal controls and a positive control for an *ABL-BCR* deletion on each run to ensure that the assay was working correctly. In addition, probes were designed to sequences that were not expected to be deleted to provide an internal control to which the other probes could be compared.

3.2.6 Statistical and sequence motif analysis of breakpoints (performed by Ru-Fang Leh and Joe Wiemels)

See Supplementary information for details of all statistical analyses.

3.2.6.1 Komolgorov-Smirnov test for distribution of breakpoints

This test was used to assess if there were any differences between the forward breakpoint distributions by comparing the subtypes of leukaemia. A P-value was considered significant if it was less than 0.05.

3.2.6.2 Scan statistics to detect clustering of breakpoints

Clustering of breakpoints was assessed by scan statistics, which tests for evidence of clustering by sliding a window along the breakpoint region and comparing the observed number of breakpoints in that window by the number that would be expected by chance alone. The result of this test is influenced by the size of the window used for analysis and so several window sizes were used. If any real clustering is present, then there would be significant values in more than one window size.

3.2.6.3 Silverman's smoothed bootstrap test for numbers of clusters

The Silverman's test was used to provide evidence for the number of clusters in the breakpoint regions. The null hypothesis states that there are one or less clusters, so if this hypothesis is rejected then there is evidence for more than one breakpoint cluster.

3.2.6.4 Proximity of breakpoints to sequence motifs

Two methods were used to assess if there were any significant associations between the position of the breakpoints and sequence motifs within 50 bp either side of the breaks. The first method was to map known sequence motifs (such as repeat regions, s/MARs, V(D)J sites, etc) and to assess if these were enriched around the breakpoints. This was done by comparing the sequence around each breakpoint with 200 simulations of uniformly generated breakpoints within each sequence region. The percentage of breakpoints that overlapped given motifs was then calculated. The second method was to search for enrichment of unspecified motifs in association with particular types of breakpoints (p210 CML, p210 ALL or p190 ALL), by taking the sequences 50 bp either side of the breakpoints and running through the Bioprospector software, web address: <http://ai.stanford.edu/~xslie/cgi-bin/BPsearch.cgi>.

3.2.7 *In silico* analysis *BCR-ABL* breakpoints for cryptic RSSs

RAG recognises specific heptamer and nonamer RSSs, makes double stranded breaks which are then repaired by NHEJ (see Figure 3-5). As non-consensus RSS sites can be mis-targeted by the RAG complex and these can be highly variable and thus difficult to look for by eye, the EMBOSS Fuzznuc tool (<http://bioweb.pasteur.fr/seqanal/interfaces/fuzznuc.html>) was used to select the most likely candidate breakpoints for RAG recombination (p190 ALL, p210 ALL and p210 CML) and was done using the definition CACAGTG (heptamer), followed by 12 or 23 nucleotides and then ACAAAAACC (nonamer). The CAC in the heptamer had to be present and up to nine mismatches were allowed in the remaining heptamer or the nonamer sequences (258, 259). Sequences 70 bp either side of the breakpoints in both *BCR* and *ABL* were inputted into Fuzznuc, and the results of any putative forward or reverse RSSs recorded. Based on these results, constructs were made as in section 3.2.8.

The action of RAG recombinase results in two products: the coding joint which results in the assembly of the V(D)J gene segments and the signal joint, which is the excised by-product of the recombination containing the RSSs which is ligated into a circular piece of DNA and subsequently degraded (see Figure 3-5).

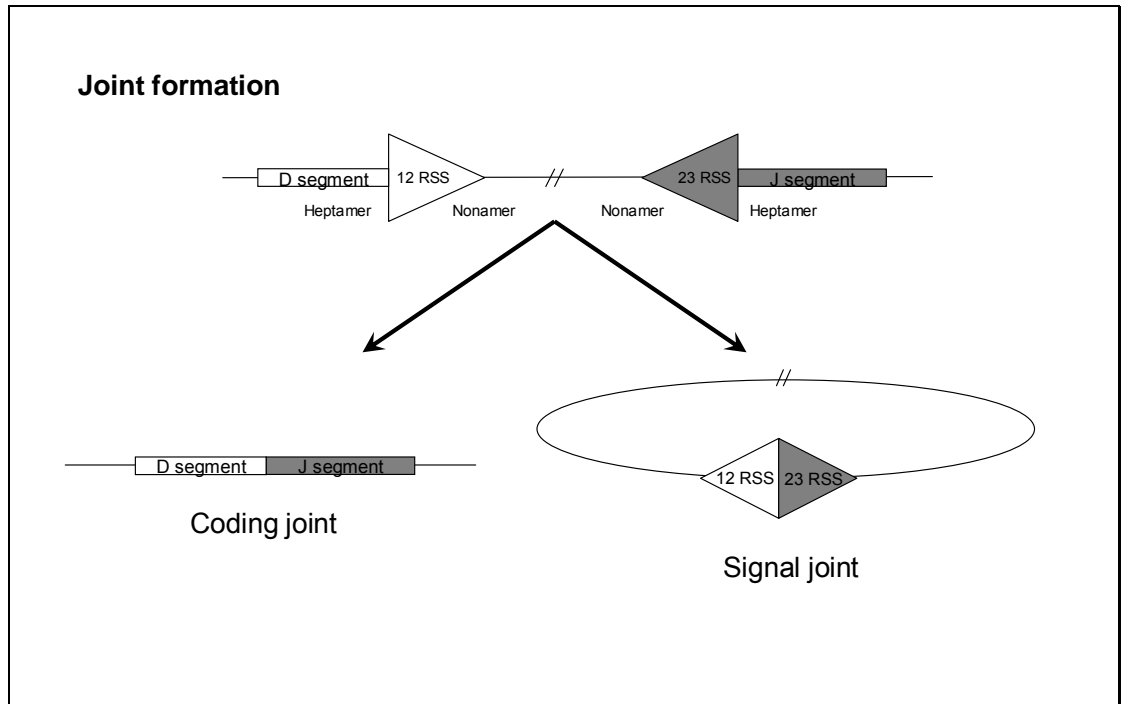


Figure 3-5: Coding and signal joint formation following DSBs by RAG recombinase and subsequent ligation.

3.2.8 Constructs for use in RAG recombination assays

In order to formally test if breakpoint sequences were acting as cryptic RAG sites, it was necessary to assess their ability to act as RAG substrates in a cell culture model. Constructs were made by inserting sequences flanking the breakpoints into a vector that would act as a model for RAG activity.

3.2.8.1 Preparing the vector

To clone fragments of DNA into the vector of choice, it was necessary to prepare the vector so that it was in the correct state to clone in the desired insert. This varied depending mainly on the restriction sites that both the vector and the insert contained. Vector DNA was either linearised or a fragment cut out by digesting between 5-10 µg with 1 U of the appropriate enzyme(s) per 1 µg of vector DNA in a total volume of 50 µl. This reaction was then incubated at 37°C for between 3 to 15 hours. The vector was then purified by crystal violet agarose gel extraction

Preparation of a crystal violet agarose gel is similar to the method described in section 2.4, however, there are some important differences. First, to make constructs it is important to limit DNA nuclease (DNase) contamination and therefore all apparatus was washed with distilled water prior to running the gels and sterile tips were used throughout. Also sterile 1 x Tris-acetate EDTA (1 x TAE) was used to both make the agarose gel and cover it whilst the gel was running, and 40 µl of 2 mg/ml crystal violet was added to the melted agarose in order to visualise the bands under white light and thus minimise DNA damage. In addition, when loading the gel, 8 µl of 6X crystal violet loading dye was used. Appropriate bands were then excised with a sterile scalpel blade and placed in a sterile 1.5 ml eppendorf tube. Using the Simple Nucleic Acid Purification (SNAP) UV-free purification kit (Invitrogen), the DNA was extracted from the gel according to the manufacturers instructions and eluted in 30 µl of deionised water. The concentration of DNA remaining after extraction was then quantified on the NanoDrop.

To prevent vector recircularisation of single digested vectors, DNA alkaline phosphatase was used to remove the 5' phosphate groups. For dephosphorylation of 5' overhangs with Calf Intestinal Alkaline Phosphatase (CIAP, New England Biolabs), 2 U were added to approximately 2 µg of linearised, crystal violet gel extracted vector and incubated at 37°C for 30 minutes. A further 1 U of CIAP was added to the reaction and incubated for a further 37°C for 30 minutes and then the DNA was crystal violet gel extracted and the amount of DNA quantified using the NanoDrop prior to use in ligations.

3.2.8.2 Preparing the insert

High Fidelity PCR (see section 2.3.1) was used for amplifying inserts for sub-cloning into TOPO vectors (Invitrogen) as it contains an enzyme blend of proof-reading and Taq DNA polymerase and therefore generates PCR products both with and without A overhangs. The cycling conditions were as Appendix II, except an extension time of 1 minute was used for these smaller fragments. Taq polymerase adds deoxyadenine (A) to the 3' ends of PCR products and the linearised TOPO vectors have single overhanging 3' deoxythymidine (T) residues which allows the PCR products to ligate efficiently. Thus, any DNA inserts amplified using a Taq polymerase were able to be sub-cloned into TOPO vectors easily prior to cloning into the construct vectors. The inserts were cleaved from the TOPO vector

with appropriate restriction enzyme(s) and crystal violet gel extracted prior to ligation into the cloning vector. For some of the constructs it was necessary to add restriction enzyme sites to the PCR product using primers tagged with the appropriate restriction enzyme sites for cloning into the RAG vector. These products were cloned into TOPO cloning vectors and then cut out using the engineered restriction enzyme sites before crystal violet purification and cloning into the RAG vectors.

To generate blunt ended PCR products which were also error free, *PfuUltra* High Fidelity DNA Polymerase (Stratagene, California, USA) was used. PCRs were set up with 500 nM forward primer, 500 nM reverse primer, 200 μ M of each dATP, dTTP, dCTP, dGTP and 2.5 Units of polymerase in 1 x *Pfu* buffer with 1 μ l of template DNA made to a final volume of 50 μ l. For cycling conditions see Appendix II. The inserts were then crystal violet gel extracted prior to blunt end ligation into the blunt ended cloning vector.

In some instances, single-stranded extensions were removed from inserts by mung bean nuclease digestion. 1 U of mung bean nuclease was used per 1 μ g of crystal violet gel purified DNA in 1 x mung bean nuclease buffer or 1 x NEB buffers 1, 2 or 4. The reaction was then incubated at 30°C for 30 minutes, then immediately crystal violet gel extracted to purify the DNA from the enzyme reaction.

3.2.8.3 Ligating the insert into the cloning vector

For the ligation reaction, 50 ng of cloning vector DNA was added to the insert DNA in a ratio of 1:3 vector:insert with 3U of T4 DNA Ligase (Promega), 1 x Ligase buffer, made up to a final volume of 10 μ l with nuclease free water and incubated at 16°C overnight. The amounts of DNA used were calculated according to the equation: {[Amount vector DNA (ng) x Size of the insert (bp)] / Size of the vector (bp)} x Ratio (i.e. 3). Vector only controls were included to ensure that the digest and dephosphorylation steps had worked.

3.2.8.4 Transforming *E.coli* cells with the construct

See section 2.5.2 for chemical transformation with the ligated vector and insert. If there were lots of colonies on plates transformed with the ligated vector plus insert compared to the vector only control, then colonies were picked, grown up and the plasmid DNA extracted, digested and sequenced to check for the correct insert (see sections 2.5.3, 2.6 and 2.7).

3.2.8.5 Endotoxin free plasmid DNA extraction

Once the correct inserts had been ligated into the cloning vectors, they were ready to be transfected into the cell line. For efficient transfections it is important to have the plasmid DNA extracted using endotoxin free techniques. Therefore the *E.coli* containing the constructs for the RAG assay were cultured and the vector DNA extracted using the Maxi-prep endotoxin free kit (Qiagen) prior to transfections. As less DNA was required for the test constructs these were extracted using the HQ mini-prep kit (Invitrogen) which is designed for use in transfections.

3.2.9 Transfecting NIH-3T3 cells

In order to test if the sequences surrounding *BCR-ABL* genomic breakpoints have functionally active, cryptic RSS sites, these constructs were co-transfected with RAG1, RAG2 and TdT expression plasmids (kindly provided by Professor Nadel, (207)) into the NIH-3T3 cell line.

3.2.9.1 Growing NIH-3T3 cells

Ensuring aseptic techniques were used at all times when performing tissue culture work, a vial of cryopreserved NIH-3T3 cells were defrosted. The cells were then washed by slow re-suspension in 5 ml Dulbecco's Modified Eagle's Medium (DMEM) supplemented with L-Glutamine, Penicillin, Streptomycin and 10% Fetal Calf Serum (FCS) and then spun down for 5 minutes at 1200 rpm. The supernatant was removed and the cells re-suspended in the residual media by flicking the tube. The cells were washed twice more before being diluted in 30 ml DMEM. For counting, 10 µl of the cell suspension was removed and

mixed with 10 μ l of Trypan Blue in an Eppendorf and the live (unstained) cells were then counted under a light microscope using a Neubauer haemocytometer. The number of cells was estimated and each 75 cm^2 flask required to be seeded with between 1×10^4 cells per cm^2 , therefore the appropriate volume of the cell suspension was added to 25 ml of DMEM grown at 37°C in 5% carbon dioxide.

3.2.9.2 Preparing the NIH-3T3 cells for transfections

Approximately 18 hours prior to transfections, the adherent NIH-3T3 cells that were in log-phase and at a density of approximately 70-80%, were trypsinised. This was done by discarding the media, washing the monolayer of cells briefly with 4 ml of trypsin (0.05%, Invitrogen), adding a further 4 ml of trypsin and then incubating at 37°C for 5 minutes. The trypsin was then inactivated with 10 ml of complete DMEM and the resultant cell suspension transferred into a 50 ml centrifuge tube and centrifuged for 5 minutes at 1300 rpm. The supernatant was discarded and the cell pellet re-suspended in the residual media by flicking the tube. For counting, 5 ml of DMEM was added to the cell suspension and the cell number counted in the same manner as in section 3.2.9.1. For use in transfections, 1×10^6 cells were plated into each 100 mm culture plate in 10 ml of DMEM supplemented with only FCS and L-Glutamine (as the antibiotics interferes with the transfection). The plates were then incubated at 37°C and 5% carbon dioxide for approximately 18 hours to achieve a confluency of 40% prior to transfections.

3.2.9.3 FuGene HD transfection

To determine the approximate optimum volume of FuGene transfection reagent to use in the assay, 2 μ g of pEFLacZ was transfected with different volumes of FuGene from 3 μ l to 8 μ l in 1 μ l increments and mixed in 100 μ l of diluent. After 60 hours, cells were washed 3 times with 1xPBS, re suspended in 1 ml paraformaldehyde (4%) in 1xPBS, incubated at 37°C for 10 minutes, and then washed with 1xPBS 3 times. A staining mix was made by adding 1 ml BetaBlue (Invitrogen) buffer and 25 μ l BetaBlue X-Gal for each well to be stained. This was added to the cells and, following incubation at room temperature for 10 minutes, 200 cells (blue and non-blue cells) were counted. The optimum volume of FuGene HD reagent was then estimated and used in the RAG assay.

For the RAG assay: several hours before transfection, the Opti-MEM diluent and FuGene HD transfection reagent (Roche) were removed from storage at 4°C and allowed to equilibrate to room temperature. For each transfection reaction, 500 µl of Opti-MEM and 2.5 µg of each plasmid (RAG1, RAG2, TdT and the plasmid with the sequence to be tested) were added to a 5 ml polystyrene tube. To each transfection reaction, 40 µl of Fugene HD reagent was added directly into the solution and immediately mixed. After 10 minutes at room temperature, the medium of each of the dishes was replaced with fresh DMEM (again with no antibiotics). After a further 5 minutes at room temperature, the transfection reagent / DNA complex was added in a drop wise manner below the surface of the medium into the dish containing the NIH-3T3 cells. After swirling to ensure good dispersal of the transfection reagent / DNA complex over the cells, the dishes were incubated at 37°C and 5% carbon dioxide for approximately 40 hours.

3.2.9.4 Plasmid DNA extraction from transfected cells

For each dish of transfected cells, the medium was removed and the cells gently washed with 1 ml of trypsin, which was then discarded and a further 1 ml of trypsin added and incubated at 37°C. After 5 minutes the trypsin was inactivated with 4 ml of complete DMEM and the cells transferred to a 50 ml centrifuge tube. The dish was washed in sterile 1xPBS to ensure all cells were in the tube and then completed to 30 ml with 1xPBS. This was centrifuged for 7 minutes at 1000 rpm at 4°C. The supernatant was discarded and the pellet re suspended in the residual PBS and then completed to a volume of 40 ml with 1xPBS. The cell suspension was centrifuged as before and the supernatant discarded before being re suspended in 1.2 ml 1xPBS, transferred to a 1.5 ml tube and centrifuged as previously described. After the supernatant was discarded, 100 µl of cold P1 (see Appendix I) was added and mixed by pipetting up and down to lyse the cells. To denature the NIH-3T3 genomic DNA, 200 µl of freshly made P2 (see Appendix I) was then added, mixed by inversion and incubated on ice for 5 minutes. To neutralise the suspension and allow the plasmid DNA to re-anneal and stay in solution, whilst precipitating the NIH-3T3 genomic DNA and proteins, 150 µl of cold P3 (see Appendix I) was added and mixed by inversion. This mixture was incubated upside down on ice. After 1 hour the suspension was centrifuged for 15 minutes at 14 000 rpm at 4°C. To remove impurities, the

supernatant containing the DNA was transferred into a new 1.5 ml tube with an equal volume of phenol/chloroform and centrifuged at 14 000 rpm for 5 minutes. The DNA contained in the upper phase was transferred to a new 1.5 ml tube and precipitated with two times cold 100% ethanol, 10% 3M sodium acetate and 20 µg of mussel glycogen. After incubation at -80°C for 20 minutes, the DNA was pelleted by centrifugation at 14 000 rpm for 15 minutes at 4°C. The pellet was then washed twice with 750 µl of 70% ethanol and allowed to air dry for 10-15 minutes prior to re-suspension of the pellet in 17 µl of nuclease free water. To ensure complete re-suspension the DNA was incubated at 65°C for 10 minutes.

3.2.9.5 Digestion of parental plasmids

As only the plasmids that have been replicated in the cells are required to assess for RAG mis-recognition, the parental plasmids must be removed prior to transformation. The methylated parental plasmids are specifically digested with the methylation sensitive enzyme DpnI, leaving only replicated plasmids. Therefore prior to transformation, to the 17 µl of DNA from section 3.2.9.4, 2 µl of the appropriate 10x buffer was added with 10U of DpnI and incubated at 37°C for at least 2 hours.

3.2.9.6 Ethanol precipitation of plasmid DNA

For DNA extraction, 80 µl of distilled water was added to the 20 µl of the DpnI digested plasmid DNA (see section 3.2.9.5) and then precipitated in two times 100% ethanol, 10% 3M sodium acetate and then placed at -70°C for 20 minutes. The precipitate was pelleted by centrifugation for 15 minutes at 14 000 rpm and washed twice in 500 µl of 70% ethanol. The pellet was then left to air dry before re-suspension in 10 µl distilled water at 65°C for 10 minutes.

3.2.9.7 Transformation of *E.coli* with extracted, digested plasmid DNA

As section 2.5.2 except all 10 µl of the product from section 3.2.9.6 was transferred to the chemically competent TOP10 *E.coli* cells, they were incubated for 60 minutes on ice and

the recovered cells were plated on ampicillin (50 µg/ml) and chloramphenicol (5 µg/ml) plates.

3.2.9.8 Screening of the colonies for RAG mis-recognition (see Figure 3-6)

PCR and then sequencing was used to screen approximately 25 colonies from each transfection with a *BCR* or *ABL* breakpoint segment. Each colony was plucked into AmpliTaq Gold (Applied Biosystems) PCR mix (1.5 mM MgCl₂, 0.5 µM primers, 200 µM dNTPs) and a hotstart PCR program: 95°C 15 mins (96°C 20 sec; 60 °C 50 sec; 72 °C 50 sec) for 25 cycles, then 72°C for 10 mins, using the RAG forward and RAG reverse primers that flank the putative and consensus RAG recognition sequences (see Appendix II).

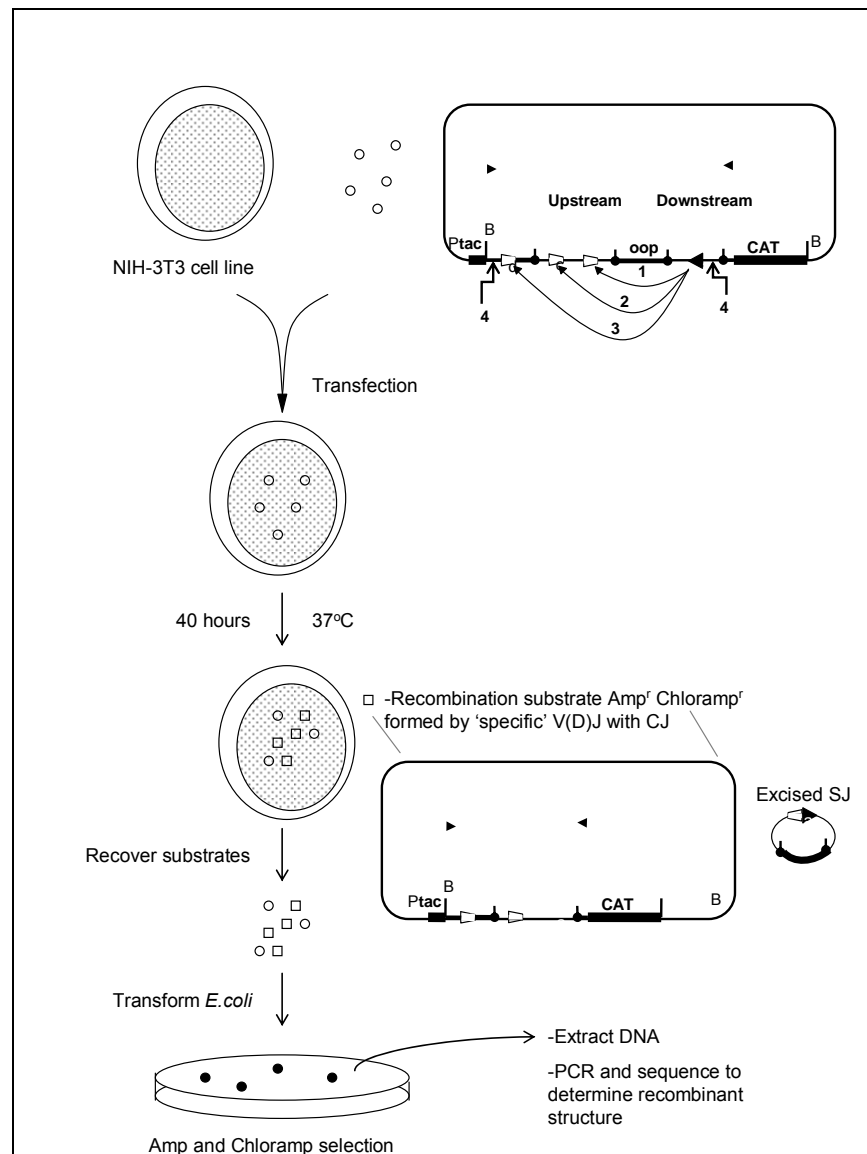


Figure 3-6: Experimental procedure to identify V(D)J recombination using Professor Nadel's constructs. The test constructs are introduced into the NIH-3T3 cell line in addition to the expression plasmids (RAG1, RAG2 and TdT, not shown). After 40 hours, the test constructs are recovered and transformed into *E. coli* for detection of recombinants on Amp plus Chloramp plates. The recombination is depicted between a consensus 23 (dark triangle) and 12 (open trapezoid) 'specific' cryptic RSS, leading to a coding joint formation (the excised signal joint is also shown for clarity. See Figure 3-41 for full description of recombination substrate.

3.3 Results

3.3.1 Screening for genomic breakpoints in p210 ALL

3.3.1.1 High Fidelity simplex PCR

The system for amplification of *BCR-ABL* breakpoints was developed in several stages. Initially, a total of 7 p210 cases (5 ALL and 2 CML) were screened using simplex High Fidelity PCR (see Table 3-1).

ID Nbr	Diagnosis	e13a2/ e14a2	Breakpoint detected	High Fidelity or LR- PCR	Size of PCR product	Primer mix	Breakpoint sequencing primer
NP241	CML	e13a2	Yes	High Fidelity	9 kb	4	BCR b2 1F
E656	CML	e14a2	Yes	LR-PCR	3 kb	20	BCR b3 1F
P679	ALL	e13a2	No	LR-PCR	NA	NA	NA
P682	ALL	e14a2	Yes	LR-PCR	1 kb	17	BCR bp(851) F
P851	ALL	e14a2	Yes	High Fidelity	2 kb	15	BCR bp(851) F
22251	ALL	e14a2	Yes	LR-PCR	4 kb	18	BCR b3(1850) F
26231	ALL	e14a2	Yes	High Fidelity	10 kb	19	BCR b3(2340) F

Table 3-1: Cases screened for p210 (e13a2 or e14a2) genomic breakpoints with simplex High Fidelity or Long Range PCR with *ABL* primers A and B.

The *ABL* breakpoint region is the same for both p210 and p190 but in p210 the breakpoint region in *BCR* is much smaller and therefore only 2 *BCR* forward primers were designed to anneal to *BCR* exon 13 and exon 14 to detect p210 breakpoints (see Figure 1-5). Five *ABL* reverse primers (designated ABL A) spanning the genomic sequence from *ABL* exon 1b to *ABL* exon a3 (see Figure 3-1) were also designed and tested prior to patient screening. All primers were tested with the appropriate *BCR* reverse or *ABL* forward complementary primer on normal DNA to ensure that they worked correctly. The resultant 10 primer combinations (each *BCR* forward primer with each *ABL* reverse primer) were then used to screen the 7 p210 cases. Breakpoint containing fragments were amplified from two of the ALL cases (cases 26231 see Figure 3-7A, and P851) plus one of the CML cases (case NP241). For the other cases, no amplification products were obtained.

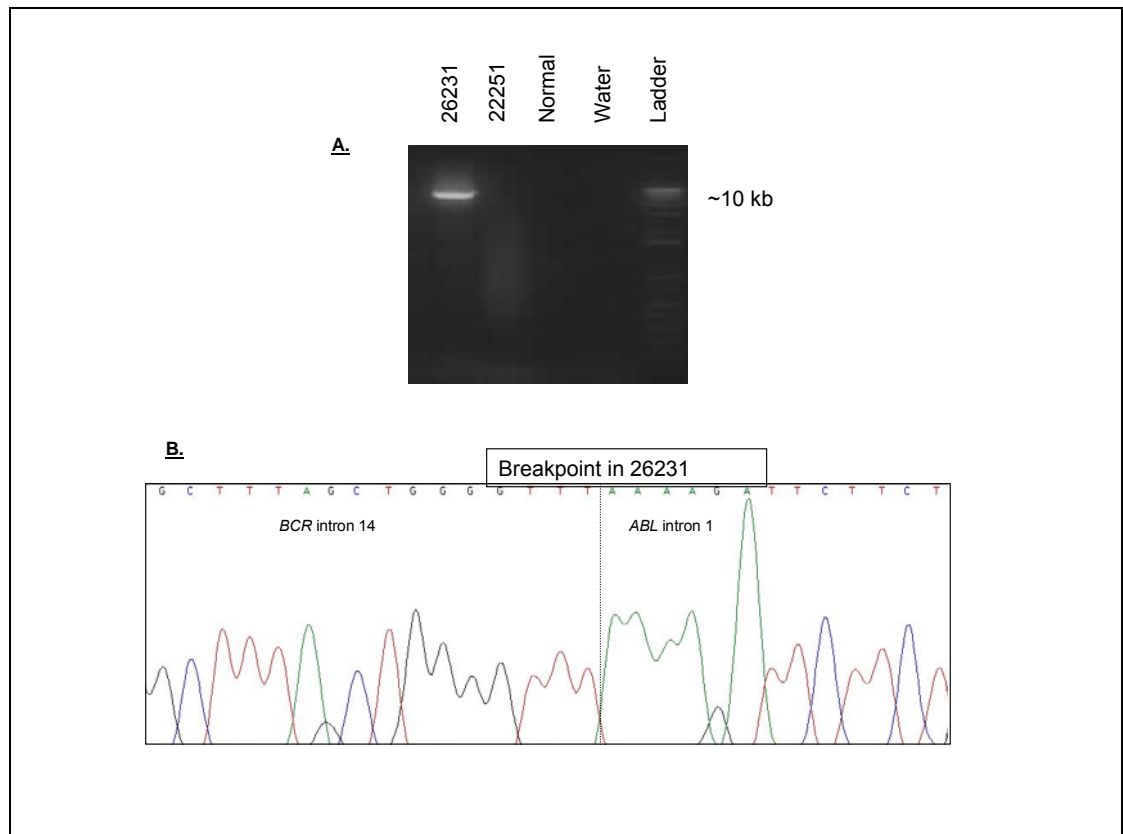


Figure 3-7: High Fidelity breakpoint detection and sequencing for case 26231. (A) Agarose gel of breakpoint containing fragment of approximately 10 kb amplified by High Fidelity PCR from case 26231 using primers *BCR* B3 F and *ABL* 1A R. (B) Electropherogram of 26231 breakpoint sequence.

The bands from the 3 positive cases were cloned and subsequently sequenced with plasmid DNA specific M13 forward and reverse primers to confirm amplification and cloning of the correct breakpoint sequence rather than a non-specific product. For NP241, sequencing with the M13 forward primer allowed sequencing across the breakpoint, however, for the other 2 cases, 3 successive sequencing primers were designed along the M-bcr region to enable sequencing across the breakpoints (see Figure 3-7 B and Table 3-1 for sequencing primers used).

3.3.1.2 Simplex LR-PCR

Since only 3 of the 7 p210 cases had amplified with the 10 primer combinations in the simplex amplification procedure, it was necessary to explore and optimise a more efficient

approach. The Expand LR-PCR kit was used in place of the High Fidelity PCR kit as it is designed to amplify larger sequences up to 20 kb and required less template DNA in the PCR reaction. Prior to use in screening, the LR-PCR was extensively optimised with changes made to the cycling conditions, buffer types, enzyme concentrations and patient DNA concentrations.

As it was a new kit all the primers were tested with a normal forward or reverse primer (see Figure 3-8) and re-designed if they did not work.

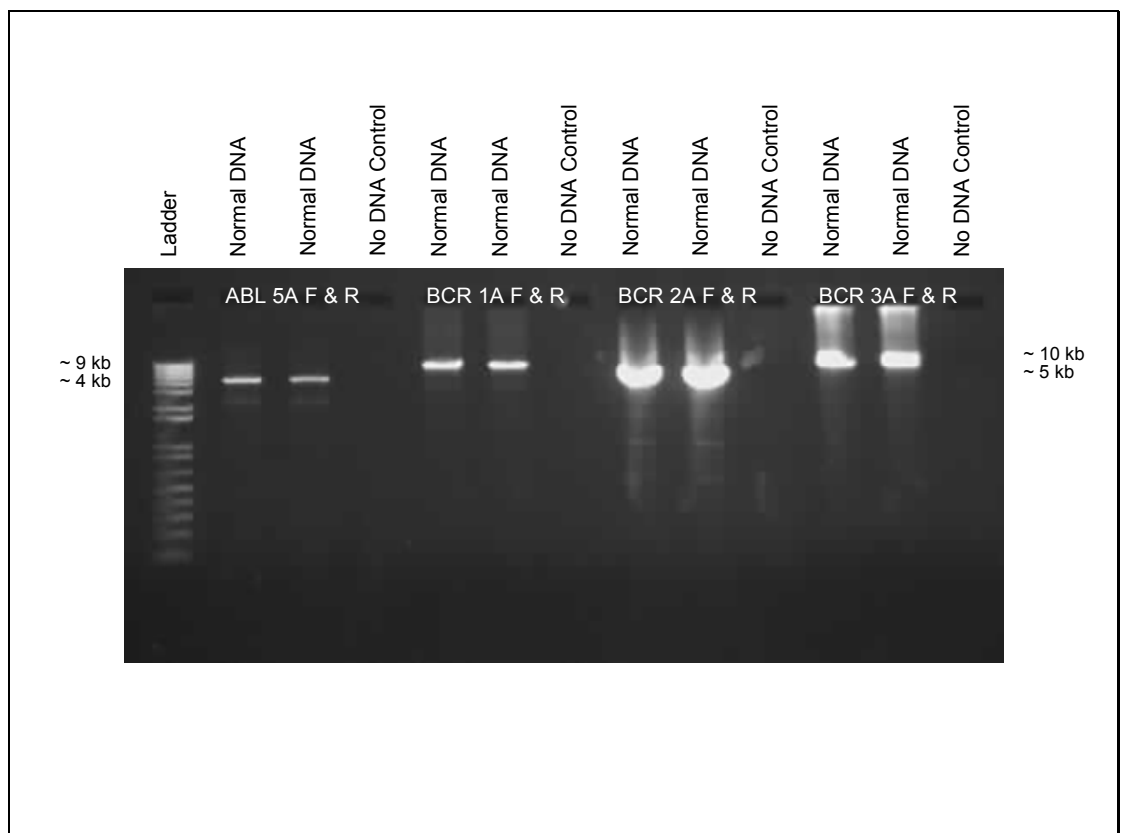


Figure 3-8: Normal control LR-PCRs. Bands amplified by different sets complementary primer pairs were extensively optimised until each primer pair worked well at the same PCR conditions.

In addition to the re-designed original set of 5 reverse *ABL* A primers, 5 more *ABL* reverse primers (designated *ABL* B) were designed to cover more comprehensively the large ~ 150 kb *ABL* region from exon 1b to exon a3 (see Figure 3-1). The remaining 4 samples in which the breakpoints had not been detected were then screened using the new *ABL* B reverse primers in conjunction with the original *BCR* forward primers. This led to the

identification of a further 2 ALL breakpoints (cases P682 and 22251) and the other CML breakpoint (case E656) after cloning and sequencing, therefore simplex PCR amplified 6/7 (85%) of the p210 breakpoints. This left only one p210 ALL breakpoint unidentified (case P679, see Table 3-1) after screening with *ABL* primers A and B.

3.3.1.3 Multiplex LR-PCR

As there was still one patient in whom a breakpoint had not been identified by LR-PCR with the *ABL* A and B primers, 10 more *ABL* forward and reverse primer pairs were designed and optimised (see Figure 3-1 for locations). At this stage, a multiplex LR-PCR method of detecting the genomic breakpoints was designed and optimised. This decreased the number of PCRs required to screen each patient from 40 to 8 and therefore the amount of patient DNA required, which for many patients was very limited. As all the primers used were designed and optimised for the same PCR conditions, the new *ABL* reverse primers were placed into groups of 5, and designated ABL C and D (see Figure 3-1 and Appendix II for the primers in each mix). Each *ABL* primer set (sets A to D) were made up so that each of the 5 *ABL* reverse primers in each mix had a final concentration of 300 nM, both of the *BCR* forward primers also had a final concentration of 300 nM. The PCR was then performed as in section 2.3.2.1. The multiplex was initially tested on 3 positive controls already detected by the simplex High Fidelity or LR-PCR, to ensure that the breakpoints would be detected, and on two p210 patients that should not yield a band from the primers used to ensure that there were no false positive bands produced by the multiplex (see Figure 3-9).

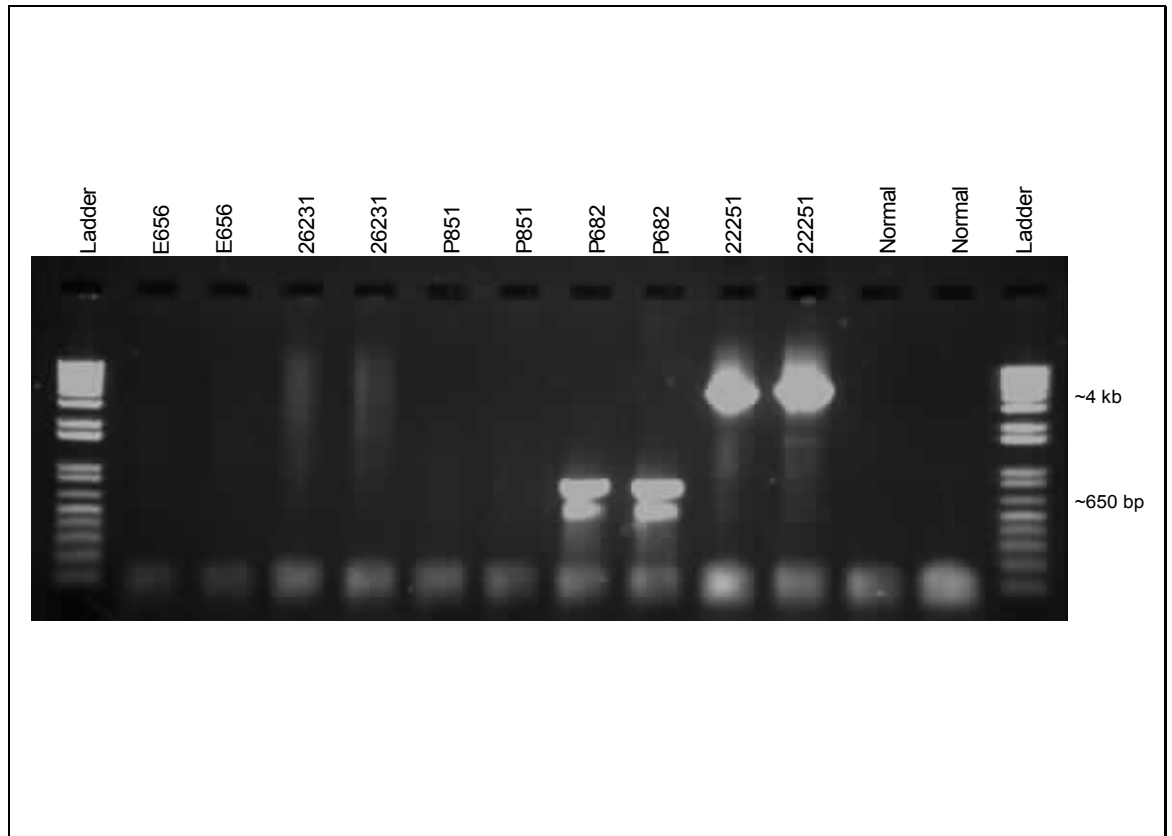


Figure 3-9: Test multiplex LR-PCR for p210 ALL and CML breakpoint detection. Agarose gel of p210 multiplex on negative (26231 and P851) and positive (E656, P682 and 22251) controls. The correct sized bands are seen in 2 out of the 3 positive controls (P682 and 22251). Each sample was amplified in duplicate.

Out of the 3 breakpoints, 2 amplified very well with no false positive bands and therefore the multiplex was suitable for use at least on a preliminary screen for breakpoints in p210 patients. Subsequent to this experiment, 4 more p210 ALL cases were tested in addition to the case in which a breakpoint could not be found (P679) with the ABL A and B primers. The multiplex was performed on these samples combining the appropriate *BCR* forward primer (depending on the breakpoint, either exon 13 primer - B2 F or exon 14 primer - B3 F) with each of the *ABL* reverse primer multiplex mixes A to D making a total of 4 PCR reactions for each case (if done in simplex the total would be 40 PCRs for each case). Breakpoint bands in 3 out of the 5 cases tested were observed (see Table 3-2) and two of these breakpoints were found using the newly designed and optimised ABL C or D reverse primer mixes.

ID Nbr	Diagnosis	e13a2/e14a2	Breakpoint detected	Size of band	Reverse Primer	Breakpoint seq primer
P679	ALL	e13a2	No	NA	NA	NA
19715	ALL	e13a2	Yes	1.5 kb	ABL 3C R	BCR B2 (368) F
24375	ALL	e14a2	Yes	5 kb	ABL 4C R	BCR bp (P851) F
26219	ALL	e14a2	Yes	4 kb	ABL 1A R	BCR B3 F
26498	ALL	e14a2	No	NA	NA	NA

Table 3-2: The first set of cases screened with the p210 multiplex LR-PCR.

All breakpoints identified by LR-PCR were confirmed by a second split-out PCR, whereby the *BCR* forward primer was combined with each *ABL* reverse primer (see Figure 3-10 B) in separate reactions and any products sequenced (see Figure 3-10 C).

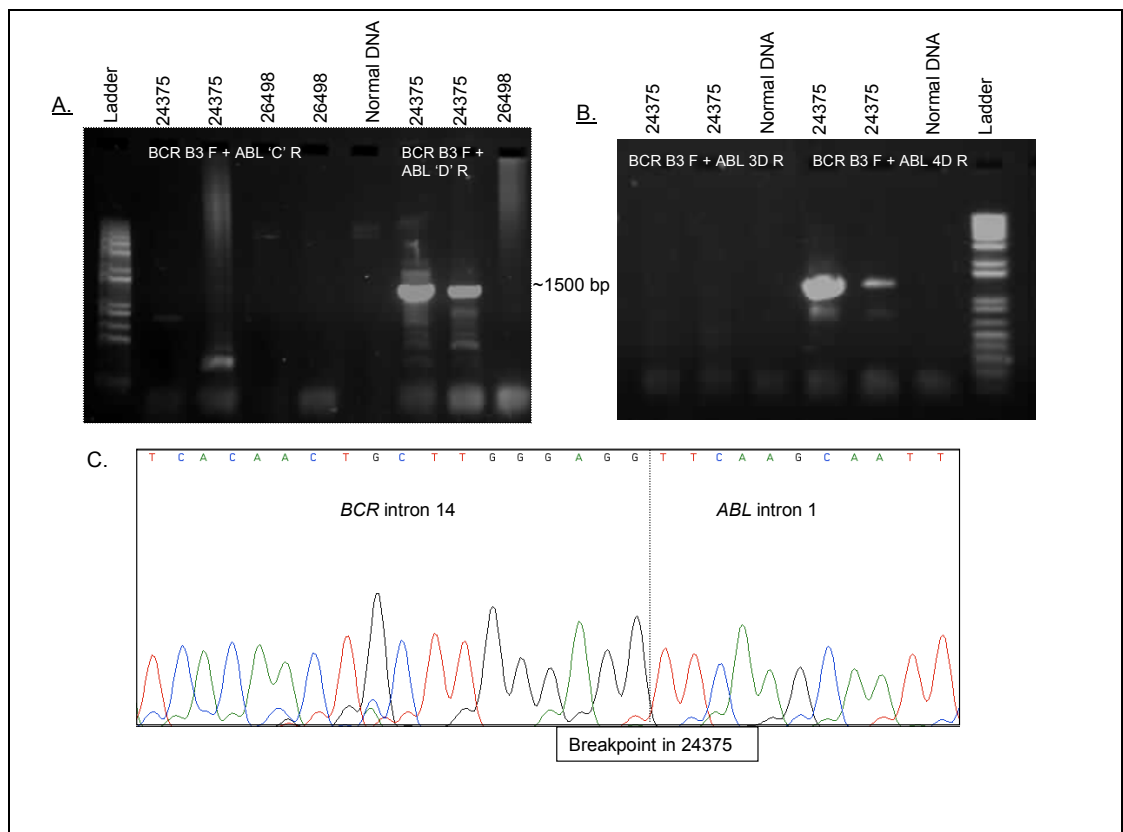


Figure 3-10: Multiplex LR-PCR breakpoint detection and sequencing for case 24375. (A) Agarose gel of multiplex LR-PCR on 24375 using *BCR* B3 F and *ABL* reverse primer mix D to amplify the breakpoint. (B) Individual simplex LR-PCR showing the breakpoint amplified with the *BCR* B3 F and *ABL* 4D R primer combination. (C) Electropherogram of 24375 sequence (see Table 3-2 for patient details).

A further 23 p210 ALL cases were subsequently received and screened with the LR-PCR multiplex. 12/23 (52%) of these resulted in amplified bands, which were then sequenced and confirmed (see Table 3-3). Simplex LR-PCR was performed on the remaining 11 multiplex LR-PCR negative cases, whereby the *ABL* primers in the multiplex were used separately with the BCR B3 or B2 forward primer to give 20 PCRs. 6 out of these 11 cases gave putative breakpoints which were sequenced and confirmed.

ID Nbr	Breakpoint detected	Multiplex or Simplex	Size of band (kb)	Reverse Primer	Breakpoint seq primer
6875	Yes	Multiplex	9.4	ABL 2B R	BCR B3 (1850) F
7982	No	NA	NA	NA	NA
10887	Yes	Simplex	3.3	ABL 5B R	BCR B2 (368) F
10258	Yes	Multiplex	2.1	ABL 1D R	BCR bp (P851) F
11129	Yes	Multiplex	3.1	ABL 3C R	BCR B3 (1850) F
11128	Yes	Multiplex	2.3	ABL 4A R	BCR B2 (368) F
8436	No	NA	NA	NA	NA
F23302	No	NA	NA	NA	NA
F22666	Yes	Multiplex	1.0	ABL 2C R	Cloned
F26154	No	NA	NA	NA	NA
F17685	Yes	Simplex	8.8	ABL 5B R	BCR B3 (1762) F
F17730	Yes	Simplex	5.3	ABL 4D R	BCR B2 F
F18544	Yes	Multiplex	2.9	ABL 4B R	BCR B2 F
F17237	Yes	Multiplex	8.3	ABL 2D R	BCR B3 (2340) F
F16604	Yes	Multiplex	3.6	ABL 1A R	BCR B2 (368) F
F15010	Yes	Multiplex	0.9	ABL 4C R	BCR B2 (368) F
F22757	Yes	Multiplex	6.1	ABL 3C R	BCR B3 (P851) F
F14472	Yes	Multiplex	3.4	ABL 5A R	BCR B2 (368) F
F11329	Yes	Multiplex	4.2	ABL 4B R	BCR B2 (368) F
F11751	Yes	Simplex	3.4	ABL 2C R	BCR B3 F
F12469	No	NA	NA	NA	NA
F24559	Yes	Simplex	1.2	ABL 2C R	BCR B2 (368) F
F20099	Yes	Simplex	6.8	ABL 2C R	BCR B3 (2340) F

Table 3-3: The second set of p210 ALL cases screened with the multiplex LR-PCR.

In total, therefore, 32 p210 ALL cases were screened (see Tables 3-1 to 3-3) and 25 breakpoints successfully amplified and sequenced, all of which were consistent with the RT-PCR results. See Figure 3-11 and Appendix III for positions of breakpoints along *BCR* and *ABL*.

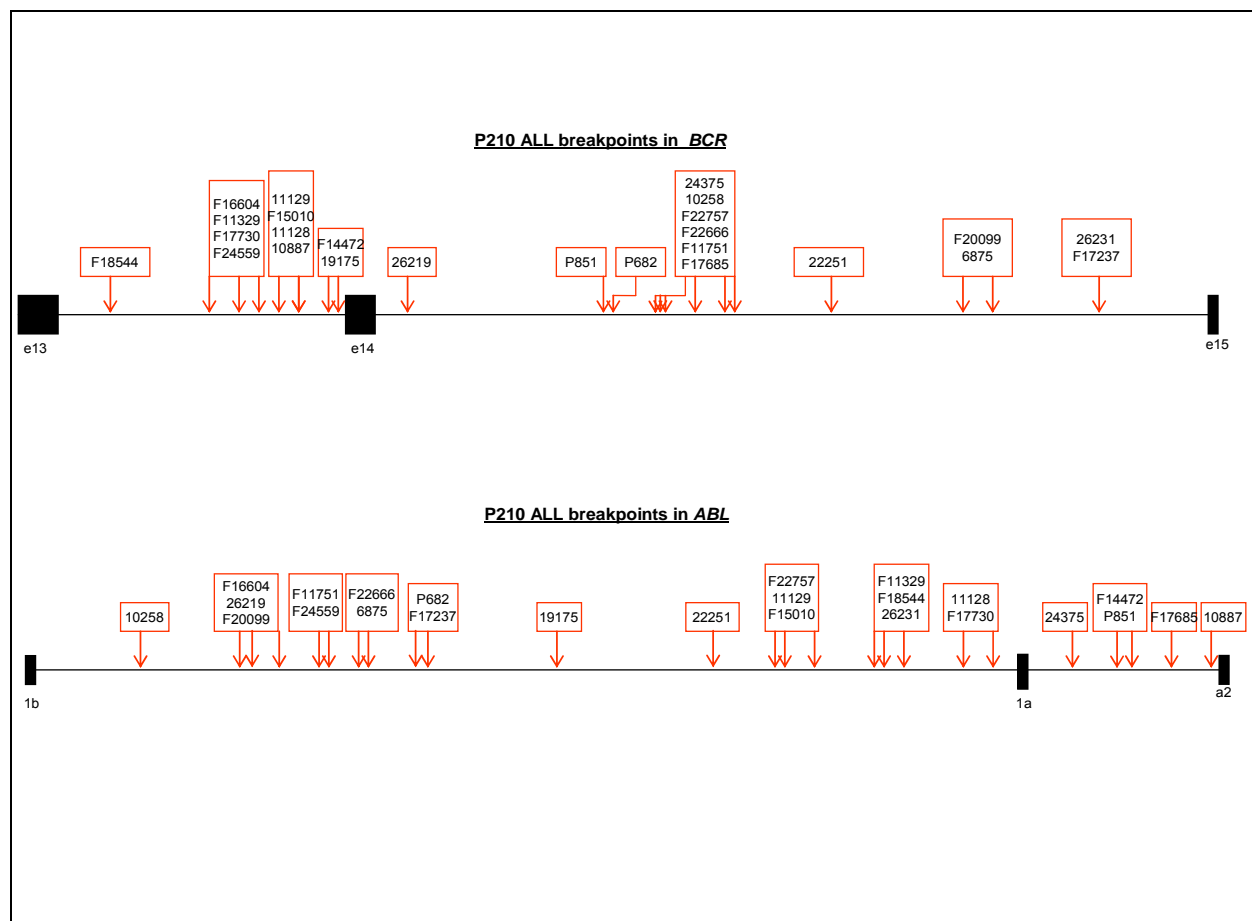


Figure 3-11: Maps of *BCR* and *ABL* showing locations of the p210 ALL genomic breakpoints.

3.3.2 Control CML p210 patients using Multiplex LR-PCR

As the multiplex was found to work well, and was quick and easy to perform on p210 patients, 33 more p210 CML controls were screened for breakpoints, in addition to the 2 CML breakpoints found previously (see sections 3.3.1.1 and 3.3.1.2). This screening provided a good panel of controls for comparison with the ALL results. 32 out of 35 of these cases resulted in the amplification of putative breakpoint bands (see Figure 3-12) which were confirmed on sequencing. Of these positives, 6 were negative on the multiplex but positive by simplex LR-PCR.

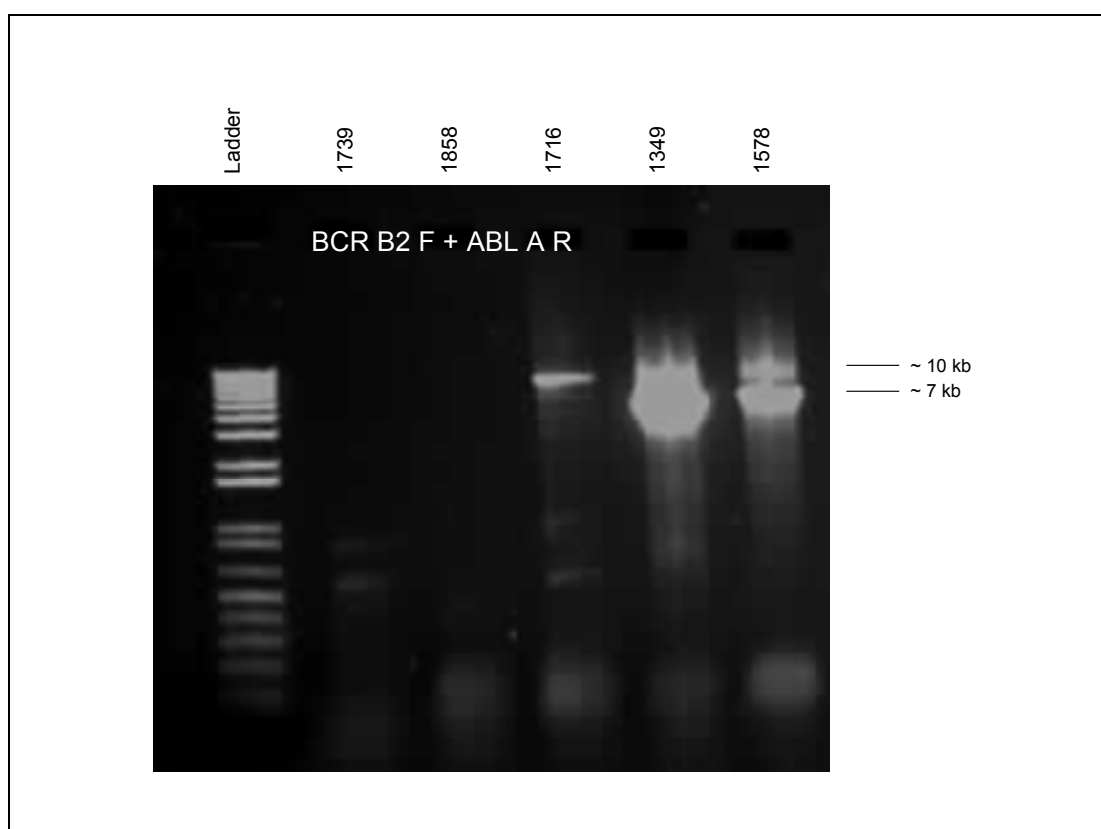


Figure 3-12: Multiplex LR-PCR detection for p210 CML cases using *BCR* B2 F and *ABL* A R primers resulted in amplification of 3 breakpoints (1716, 1349 and 1578).

The positions of the sequenced breakpoints are shown below (see Figure 3-13). Of the 5 that failed to produce a breakpoint band, 3 failed to produce normal control bands and thus these failed results are due to poor DNA quality rather than the failure of the LR-PCR. Overall, the LR-PCR successfully amplified breakpoints in 32/32 (100%) of patients known to express p210 BCR-ABL (see Figure 3-13 and Appendix III).

From Figure 3-13, it is also clear that there are 2 cases with breakpoints outside the normal regions. Case 286 had a breakpoint upstream of *ABL* intron 1b and 1103 had a breakpoint within *BCR* exon B4, which concurs with the RT-PCR results for this case.

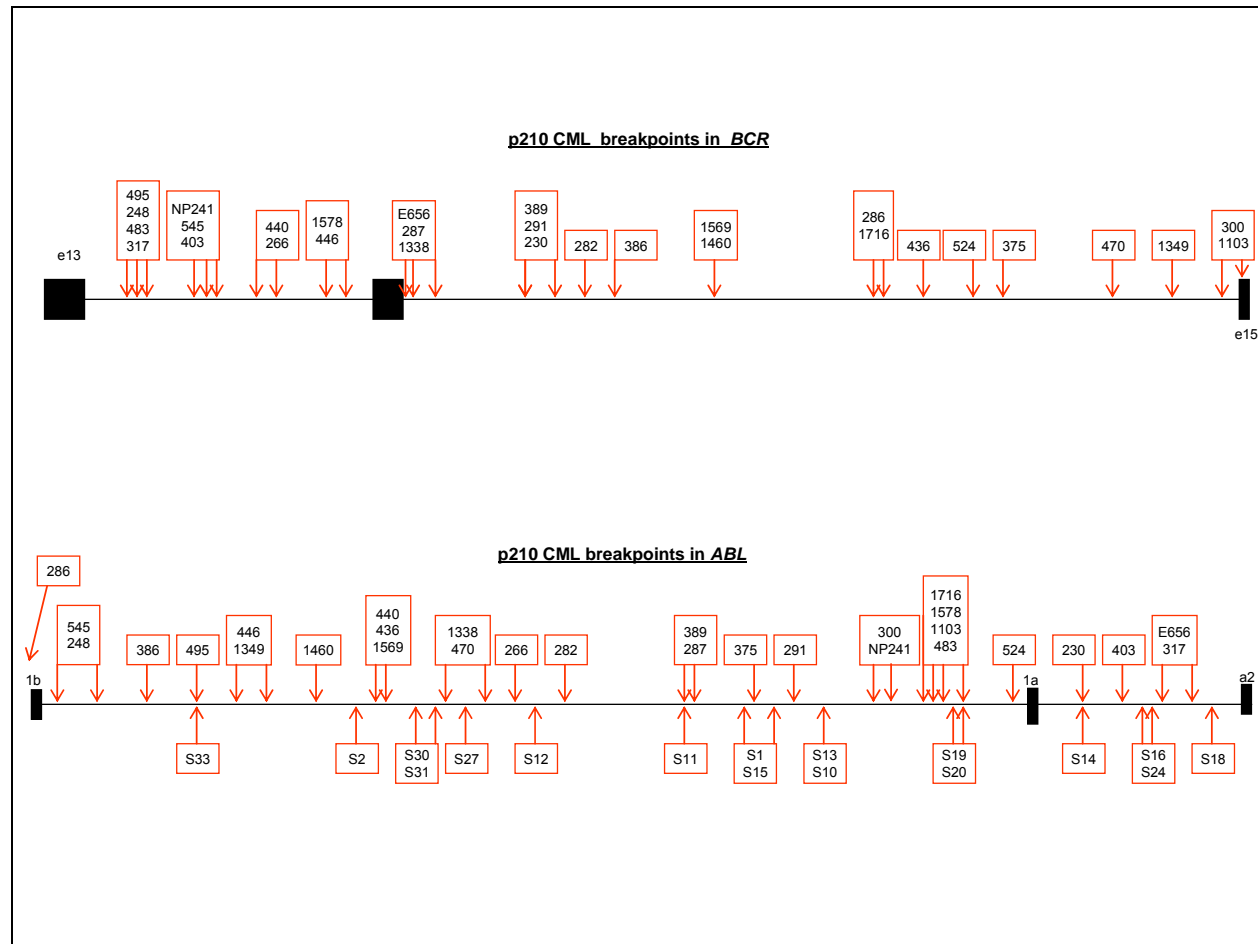


Figure 3-13: Maps of *BCR* and *ABL* showing the positions of the p210 CML control breakpoints. Breakpoints shown below the line are those from collaboration with Junia Melo which were also used for statistical analysis.

3.3.3 Screening for genomic breakpoints in p190 ALL

3.3.3.1 Fosmid FISH

Due to the large size of the regions in both *BCR* and *ABL* where breakpoints can occur in p190 ALL, fosmid FISH was used to sub-localise the breakpoint region and therefore decrease the area over which LR-PCR was required for cases with available cytogenetic material. Fosmids are approximately 40 kb in length and in total 7 fosmid clones (5 in *ABL* and 2 in *BCR*) were ordered and received from the Sanger Institute (see Table 3-4).

Full name	Abbreviated name	<i>BCR</i> or <i>ABL</i>	Worked correctly	Used in screening
G248 P81363 D5	D5	<i>BCR</i>	Yes	Yes
G248 P8761 F1	F1	<i>BCR</i>	Yes	Yes
G248 P84175 E8	E8	<i>ABL</i>	Yes	Yes
G248 P86688 G4	G4	<i>ABL</i>	Yes	Yes
G248 P85890 D11	D11	<i>ABL</i>	Yes	No
G248 P89107 E11	E11	<i>ABL</i>	Yes	Yes
G248 P800008 G3	G3	<i>ABL</i>	Failed	No

Table 3-4: Details of fosmid clones.

Before use in screening, all probes were tested on normal metaphases. It was found necessary to increase the amount of DNA in the probe precipitation step (see section 3.2.3) from 10 µl to 20 µl, before the probes could be visualised in interphase and metaphase cells. The fosmids were then tested to ensure hybridisation to the correct regions (see Figure 3-14). Only one fosmid (G3) failed to give signals, the remaining 6 hybridised correctly and gave strong signals. As a result of these initial experiments, 5 fosmids were chosen that gave the least background, the strongest signals and the best coverage of the 2 regions. Three were chosen to span the *ABL* breakpoint region up to exon a3 and two chosen to span the *BCR* intron 1 up to exon 2 (see Figure 3-1 and Table 3-4).

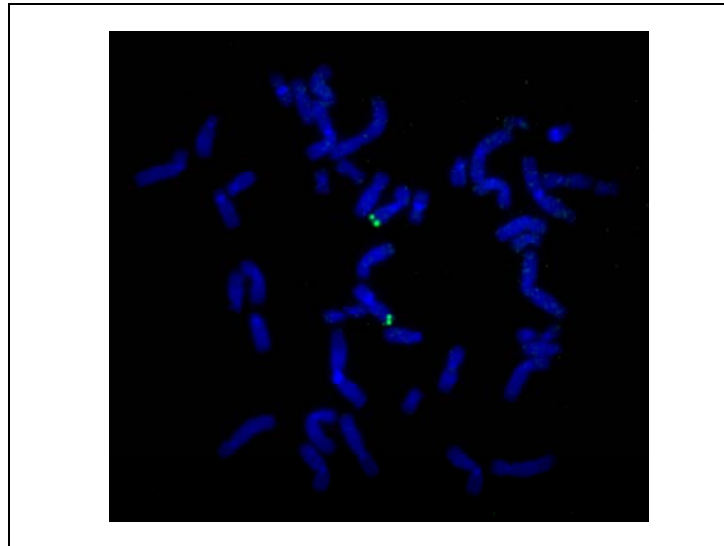


Figure 3-14: *ABL* fosmid FISH probe G4 (green) showing hybridisation to 9q34 on normal metaphases.

Six p190 ALL cases had cytogenetic suspensions available for FISH analysis. Systematic FISH analysis was then performed to narrow down the region within which the breakpoints were located, using different combinations of probes. Due to the low frequency of metaphases in ALL patient samples, interphases were also scored to ensure the most accurate interpretation of the FISH data. One patient with a very poor cytogenetic suspension was unscorable and was screened using all LR-PCR primer combinations. Three of the remaining 5 samples, 25373, 25169 and 25654 showed positive fused red and green signals with the D5 and G4 probes indicating breakpoints within the regions covered by them. 23827 showed fusion signals with F1 and E8 (see Figure 3-15 A) and on subsequent FISH showed fused signals with F1 and E11. 21296 showed a fusion between F1 and E11 (see Figure 3-15 B).

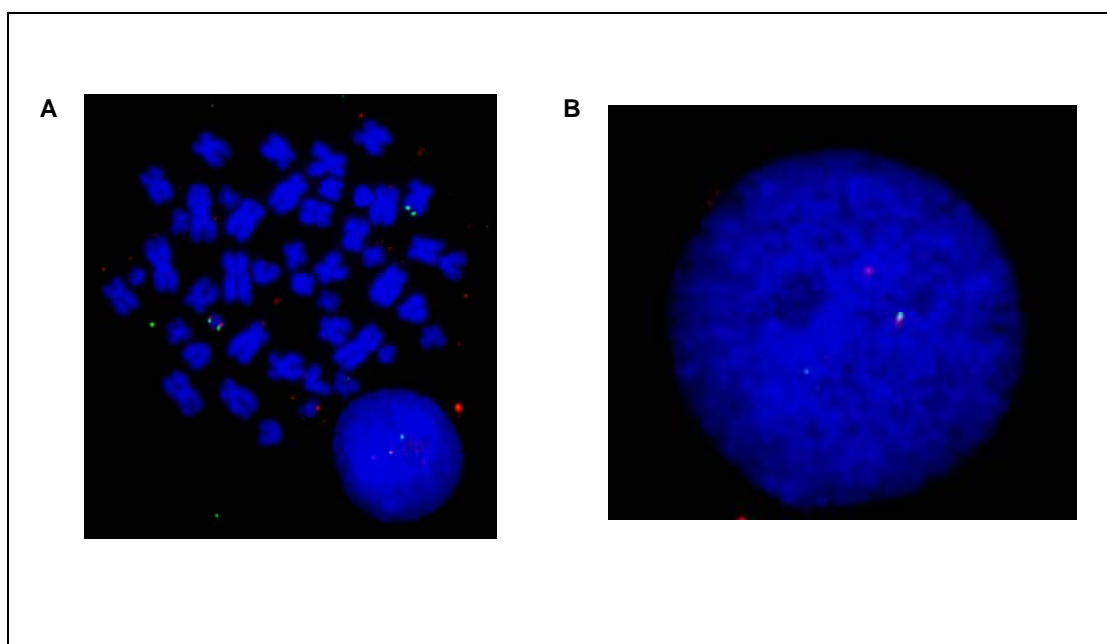


Figure 3-15: Fosmid FISH on p190 cases. (A) Fosmid FISH on 23827 using *BCR* (red) labeled probe F1 and *ABL* (green) probe E8, clearly showing fused signals in both the metaphase and the interphase cell. (B) Fosmid FISH on 21296 using *BCR* (red) labeled probe F1 and *ABL* (green) labeled probe E11 showing a fused signal and two normal signals.

LR-PCR was performed on all 5 cases with good cytogenetic suspensions, with primers spanning the FISH probes and 3 of the breakpoints were amplified and sequenced. Two of the breakpoints (21296 and 23827) were not amplified by LR-PCR using primers based on the predicted fosmid pattern of F1 E11, and it was not until LR-PCR screening, that the actual fused fosmids of D5 E11 were ascertained (see Table 3-5 and Figure 3-1).

ID Nbr	Predicted fosmids seen	LR-PCR BCR primer	LR-PCR ABL primer	Size of band	Actual fused fosmids	Breakpoint seq primer
21296	F1 E11	BCR 5A F	ABL 1A R	11.5 kb	D5 E11	BCR bt3 (17086) F
23827	F1 E11	BCR 4B F	ABL 2D R	9.5 kb	D5 E11	BCR bt3 (13916) F
25169	D5 G4	BCR 5A F	ABL 4A R	3.8 kb	D5 G4	BCR 5A F
25373	D5 G4	BCR 5B F	ABL 4D R	6.0 kb	D5 G4	BCR bt3 (25887) F
25654	D5 G4	BCR 5B F	ABL 4A R	5.4 kb	D5 G4	BCR bt3 (22807) F

Table 3-5: Fosmid FISH results.

3.3.3.2 LR-PCR

Although it might have been possible to obtain cytogenetic material from more cases, it was very unlikely that material would be available from all of them and therefore a LR-PCR strategy was developed, following on from the success of the LR-PCR for p210 breakpoints. Although the region in which breakpoints can occur in *ABL* remains the same in p190 ALL, the breakpoints in *BCR* occur in the m-bcr with the possibility of them occurring in the rest of the large intron 1 of *BCR*. This meant that more *BCR* forward primers were required to adequately cover the region for breakpoint screening and consequently increased the number of primer combinations than were necessary for p210 breakpoint determination. Initially, for screening of cases for which no cytogenetic material was available, 50 primer combinations were used for screening (5 *BCR* forward and 10 *ABL* reverse primers) however in the first 4 cases screened no breakpoints were amplified. A further 5 *BCR* primer pairs were designed and optimised making 100 primer combinations in total (see Table 3-6).

Primer combinations		
Forward Primers	Reverse Primers	Primer mixes
BCR 1A F – BCR 5A F	ABL 1A R	1-5
	ABL 2A R	6-10
	ABL 3A R	11-15
	ABL 4A R	16-20
	ABL 5A R	21-25
	ABL 1B R	26-30
	ABL 2B R	31-35
	ABL 3B R	36-40
	ABL 4B R	41-45
	ABL 5B R	46-50
BCR 1B F – BCR 5B F	ABL 1A R	51-55
	ABL 2A R	56-60
	ABL 3A R	61-65
	ABL 4A R	66-70
	ABL 5A R	71-75
	ABL 1B R	76-80
	ABL 2B R	81-85
	ABL 3B R	86-90
	ABL 4B R	91-95
	ABL 5B R	96-100

Table 3-6: Simplex LR-PCR primer combinations used in screening for p190 cases indicated in Table 3-7, each of the 10 *BCR* forward primers are used with each of the 10 *ABL* reverse primers to give 100 primer mixes for screening. See Figure 3-1 for primer positions and Appendix II for sequences.

These 100 primers combinations were used to test 13 p190 ALL cases for whom no fixed cell suspensions were available for FISH analysis (Table 3-7).

Name	Breakpoint detected	Primer mix	Size of band	Breakpoint seq primer
Ital 1	Yes	Simplex 73	5.1 kb	BCR bt 2 (18020) F
Ital 2	Yes	Simplex 9	3.0 kb	ABL 4A R
Ital 3	No	NA	NA	NA
Ital 4	Yes	Mpx BCR 4A F + ABL C R	12.1 kb	NA
Ital 5	Yes	Mpx BCR 4C F + ABL D R	7.2 kb	BCR bt 3 (4496) F
Ital 6	Yes	Simplex 78	8.1 kb	BCR bt 2 (19316) F
Ital 7	No	NA	NA	NA
Ital 9	Yes	Mpx BCR 5A(b) F + ABL A R	9.2 kb	BCR bt 3 (19155) F
Ital 10	Yes	Simplex 100	4.5 kb	BCR bt 3 (24281) F
747	Yes	Mpx BCR 2C F + ABL A R	7.3 kb	ABL 4 (12550) R
P650	No	NA	NA	NA
P653	Yes	Mpx BCR 9C F + ABL C R	3.2 kb	BCR bt 3 (21368) F

Table 3-7: p190 cases screened by LR-PCR including primers used to detect and sequence the breakpoints (see Appendix II for primer details).

Initially, all amplified bands were checked by sequencing with the PCR primers to confirm that they were real (i.e. have one end from *BCR* and one end from *ABL*) and not an artefact. This sequencing yielded the breakpoint in case Ital 2, which fortunately lay close to the *ABL* primer. Of the remaining 12 cases, the breakpoints were successfully identified in three further cases (see Table 3-7). As an example, the LR-PCR product for Ital 6 is shown in Figure 3-16 A. Restriction enzyme digests were performed, using unique sites to ascertain the approximate position of the breakpoint. The results of the digests enabled primers to be designed that were much closer to the breakpoint (Figure 3-16 B). The product was then cloned and sequencing with M13 forward primer revealed the breakpoint sequence (Figure 16 C). The breakpoint in Ital 1 was sequenced in a similar manner by using restriction enzyme digests.

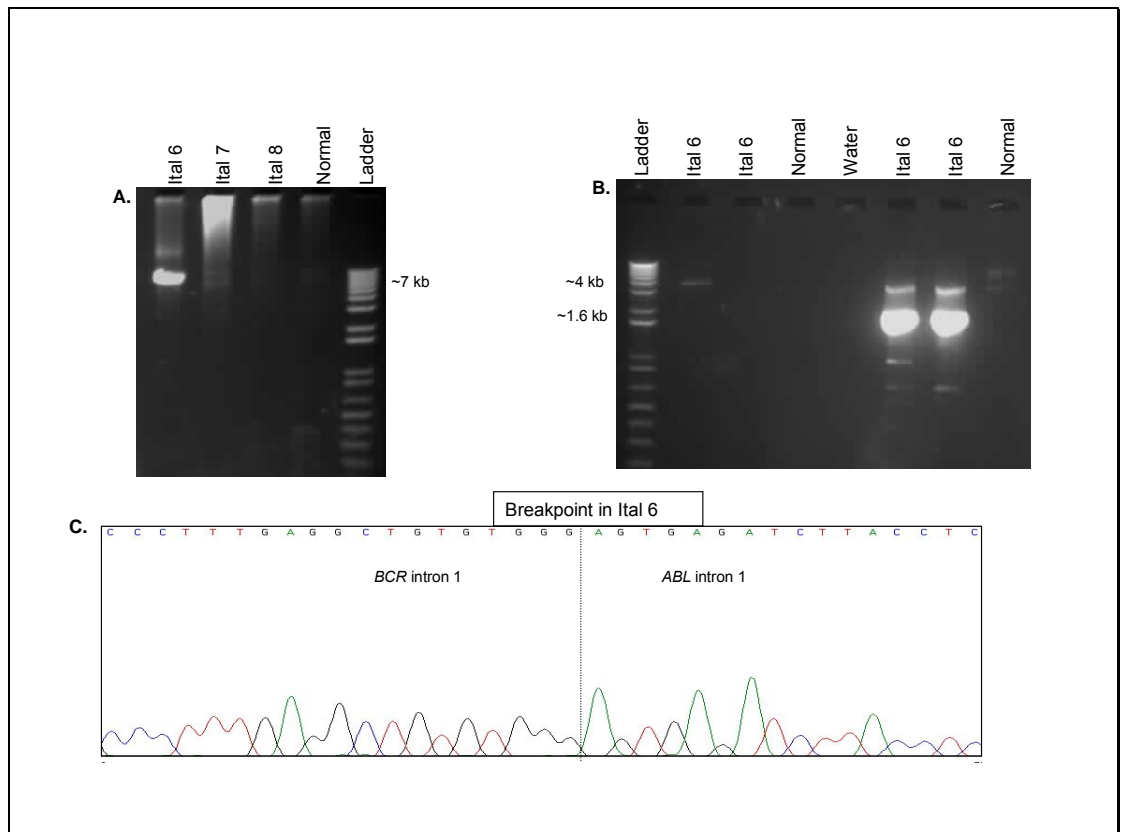


Figure 3-16: Simplex LR-PCR breakpoint detection and sequencing of Ital 6. (A) Amplification of Ital 6 breakpoint using primer mix 78 (see Table 3-7). (B) Amplification of breakpoint using primers designed from the results of restriction enzyme digests. (C) Electropherogram showing breakpoint sequence after cloning of smallest breakpoint band and sequencing with M13 forward primer.

Although most positive sequencing bands were very bright (see Figure 3-16 A), some breakpoints amplified weakly. Therefore all bands that did not appear to be a result of non-specific binding i.e. one faint discrete band present in only one patient for a particular primer combination (see Figure 3-17 A) were sequenced. Many bands were sequenced in this manner and, for one case (Ital 10), the breakpoint was identified. To confirm the breakpoint the PCR was repeated with the same primers but at a lower annealing temperature of 64°C to yield a much brighter band (see Figure 3-17 B). From this point onwards all LR-PCRs were run at this annealing temperature.

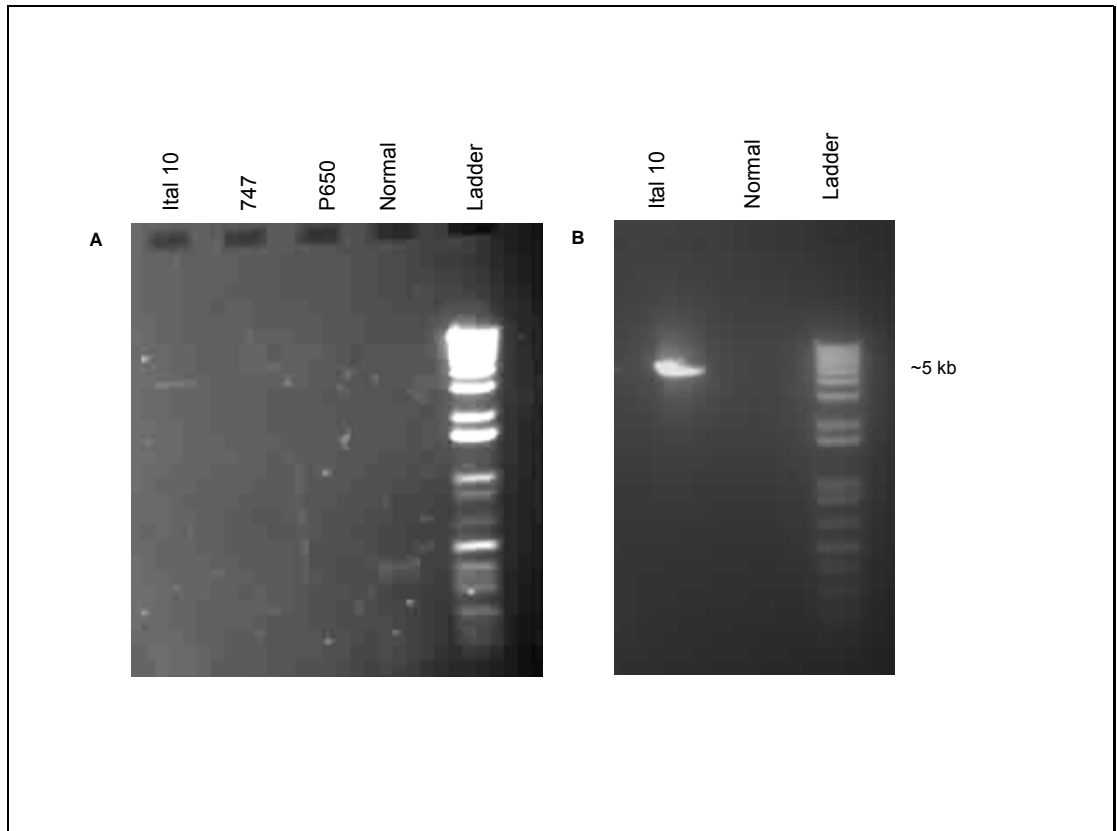


Figure 3-17: Simplex LR-PCR breakpoint amplification for Itai 10 using primer mix 100. (A). Agarose gels of LR-PCR annealing temperature at 66 °C and (B) at 64°C

Again, the amplified product for Itai 10 was relatively large - approximately 5 kb and quite weak. Such a weak product would not be easy to assess by restriction enzyme digests, so instead, to determine the approximate location of the breakpoint, 3 primers were designed at approximately 1.5 kb intervals 3' of the *BCR* PCR primer (*BCR* 5B F) and then used to sequence the band in Figure 3-17 B. If the sequencing failed, the breakpoint was assumed to be 5' of the failed primer and if sequencing worked with a particular primer but showed only *BCR* sequence, the breakpoint was assumed to be 3' of this primer and further primers designed and used for sequencing. In this way the location of the breakpoint was determined.

3.3.3.3 Multiplex LR-PCR on p190 samples

Although 4 out of 13 breakpoints were found using the simplex of 100 primer combinations, the additional 10 *ABL* primers (*ABL* C and D) designed to screen for the p210 breakpoints could be used on the p190 breakpoints in combination with the p190

specific *BCR* primers, to increase the coverage for breakpoints. However, another 10 *ABL* reverse primers would mean another 100 PCR reactions bringing the total number of reactions to 200 for each case. This would have been both very time consuming and expensive so the multiplex LR-PCR was again tested and optimised for screening for genomic breakpoints in p190 cases. This reduced the number of PCRs from 200 to 40. The multiplex was performed in exactly the same way as the p210 multiplex except each p190 *BCR* forward primer (of which there were 10, split into 2 groups of 5; *BCR* A and B) was combined with each *ABL* reverse primer mix A, B, C and D. Again, this technique was tested on the positive controls already identified using the simplex with a normal control included to ensure that there were no false positives (see Figure 3-18). Out of 5 positive controls tested, 4 amplified breakpoint bands of the correct size and one failed.

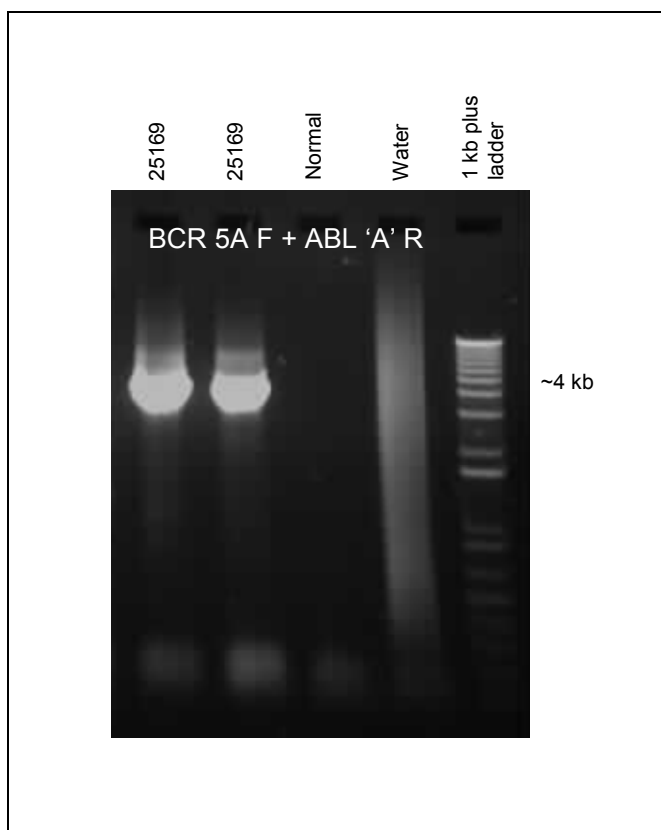


Figure 3-18: Test multiplex LR-PCR to detect 25169 positive control using *BCR* 5A F and *ABL* A R. The normal and water negative controls are both blank.

The multiplex using *BCR* A and B forward primers with the *ABL* A – D reverse primer mixes were therefore considered good enough for use in screening and were used to test a

further 25 samples that had been received (see Table 3-8) and also the 9 which had been previously screened using the simplex and had failed to give breakpoint bands (see Table 3-7) at the decreased annealing temperature of 64°C. This led to the identification of 10 new breakpoints.

ID Nbr	Breakpoint detected	LR-PCR primer mix	Size of band	Breakpoint sequencing primer
17847	No	NA	NA	NA
7853	No	NA	NA	NA
7847	No	NA	NA	NA
9463	Yes	Mpx BCR 3B F + ABL D R	4.3 kb	BCR 3B F
8069	No	NA	NA	NA
20763	Yes	Mpx BCR 5B F + ABL B R	4.0 kb	BCR bt 3 (23282) F
30023	No	NA	NA	NA
30583	No	NA	NA	NA
44095	Yes	Mpx BCR 3B F + ABL B R	8.7 kb	NA
44240	Yes	Mpx BCR 8C F + ABL A R	1.9 kb	BCR bt 3 (18256) F
50006	No	NA	NA	NA
25624	No	NA	NA	NA
27523	Yes	Mpx BCR 4B F + ABL B R	8.0 kb	BCR bt 3 (13468) F
28624	Yes	Mpx BCR 5B F + ABL A R	7.1 kb	BCR bt 3 (28590) F
28835	Yes	Mpx BCR 5A F + ABL C R	5.7 kb	BCR bt 3 (17550) F
28852	Yes	Mpx BCR 4B F + ABL A R	9.2 kb	BCR bt 2 (12440) F
11941	Yes	Mpx BCR 5C F + ABL B R	0.6 kb	ABL 1B R
14208	No	NA	NA	NA
14382	No	NA	NA	NA
14866	No	NA	NA	NA
15637	No	NA	NA	NA
19635	Yes	Mpx BCR 4B F + ABL C R	9.9 kb	BCR bt 3 (4699) F
20394	No	NA	NA	NA
21385	No	NA	NA	NA
21981	No	NA	NA	NA
24350	Yes	Mpx BCR 4A F + ABL C R	10.3 kb	BCR bt 3 (6038) F

Table 3-8: Cases screened using the p190 multiplex LR-PCR. Cases screened with *BCR* A, B and C forward primers and *ABL* A-D R multiplex mixes and *ABL* upstream multiplex mixes.

However, this still only gave a total of 19 breakpoints identified out of 43 cases with p190 ALL that were available for analysis. Therefore other techniques were investigated to find the multiplex negative breakpoints.

3.3.4 Long Range Bubble PCR

As the bubble PCR technique had been successfully used to amplify breakpoints in *FIP1L1-PDGFR*A positive CEL (see Chapter 4), this technique was also assessed for use in detection of p190 multiplex LR-PCR negative cases. As a proof of principle experiment, 6 bp blunt ended restriction enzymes were chosen to cut a 7 kb and a 10 kb fragment in *BCR* intron 1. Nested reverse primers were then designed at the 3' end of the fragments and the bubble PCR protocol performed as section 3.2.4.3 and Figure 2-1. Prior to use in bubble PCR, the primers were tested to make sure that they worked with the LR-PCR conditions (see Figure 3-19 A). The first step LR bubble PCR was initially tried at a 1 in 20 dilution and then this product was diluted again to a 1 in 20 dilution and used in the second step however, from the gel picture in Figure 3-19 B, it is clear from the smeary appearance that too much DNA was used in the PCR. Two more dilutions were tried: the DNA to be dispensed into the first step LR bubble PCR was diluted by 1 in 100 and then for the nested step, dilutions of the first step product were made of 1 in 200 (see Figure 3-19 C) and 1 in 2000 (see Figure 3-19 D). None of the normal controls gave the correct sized bands (see Figure 3-19 B, C, D) therefore LR bubble PCR was not pursued further.

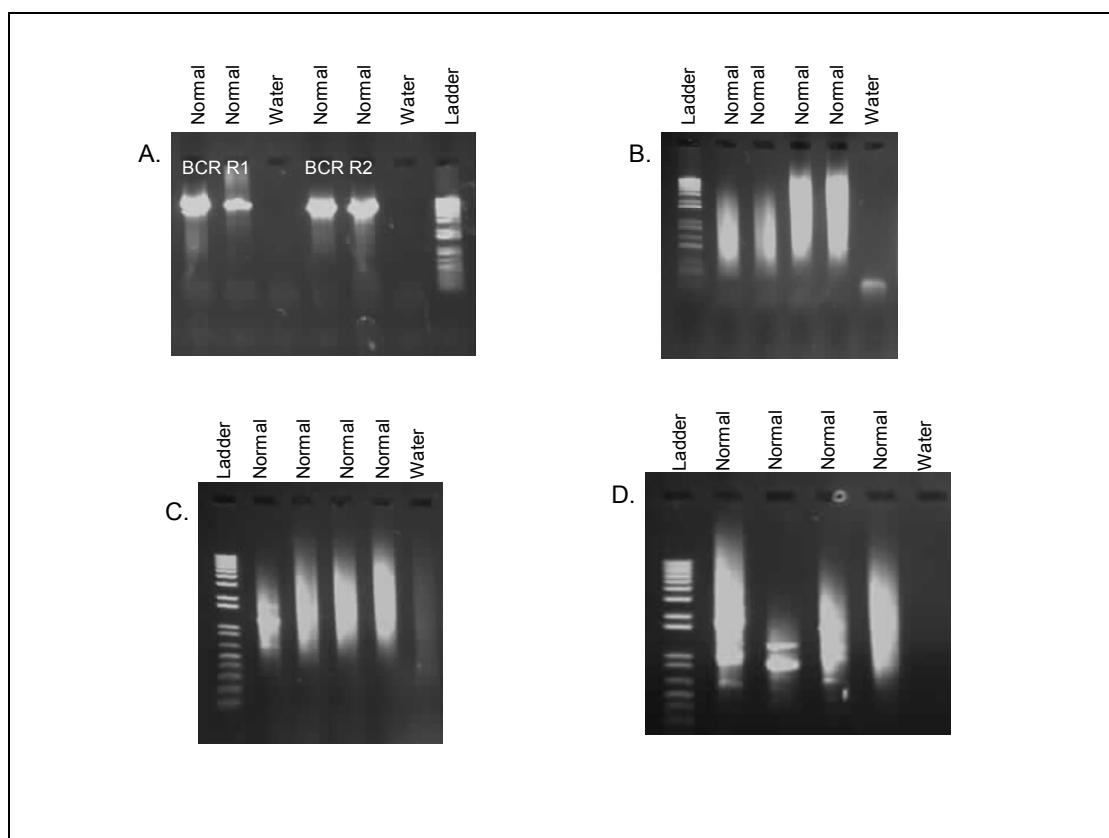


Figure 3-19: LR bubble PCR products. (A) LR-PCRs on normal control DNA, with *BCR* R1 and *BCR* R2 with appropriate *BCR* forward primers. (B) Nested LR bubble PCRs using bubble DNA dilutions of 1 in 20 in the first and nested step. (C) Nested LR bubble PCRs using bubble DNA dilutions of 1 in 100 in the first step and 1 in 200 in the nested step. (D) Nested LR bubble PCRs using bubble DNA dilutions of 1 in 100 in the first step and 1 in 2000 in the nested step.

3.3.5 Long Distance Inverse (LDI) PCR

One reason that the LR bubble PCR may not have worked is because it is necessary to use blunt ended ligation of the bubble oligo onto the digested DNA fragments. Blunt ended ligation is notoriously inefficient and, although bubble PCR works well for short PCRs, this inefficiency may have compromised the LR-PCRs. Therefore, LDI-PCR was attempted since this technique is known to work for amplification of large breakpoints (256).

To digest the DNA, 6 bp sticky ended restriction enzymes were used, specifically *EcoRI*, as it generated overlapping fragments that covered the m-bcr of *BCR* intron 1 of between 3-8 kb long (as determined by analysis of the sequence using DNA Strider). *EcoRI* also

cuts frequently in the *ABL* breakpoint region, thus increasing the likelihood of amplifying a breakpoint band within the size range of LR-PCR (see Table 3-9).

Position of EcoRI sites in <i>BCR</i> intron 1	Size of fragment	Name
35304 & 41820	6516	LDI 1
41820 & 44756	2936	LDI 2
44756 & 51448	6692	LDI 3
51448 & 54433	2985	LDI 4
54433 & 62665	8232	LDI 5
62665 & 70192	7527	LDI 6

Table 3-9: Location of EcoR1 sites in *BCR* intron 1 and the size of fragment for LDI-PCR amplification.

To amplify *BCR-ABL* breakpoints, nested forward and reverse *BCR* primers were designed in the 5' region of the *BCR* digested fragments (see Figure 3-2). Prior to use in LDI-PCR, these primers were tested on normal control DNA with the appropriate forward or reverse primers to ensure that they worked with LR-PCR conditions (see Figure 3-20 A). Initially, only two fragments were tested by LDI-PCR both on normal DNA and on positive control p190 ALL DNA (see Figure 3-20 B and C). The products were sequenced and all confirmed to be normal *BCR* or *BCR-ABL* breakpoint sequence as appropriate.

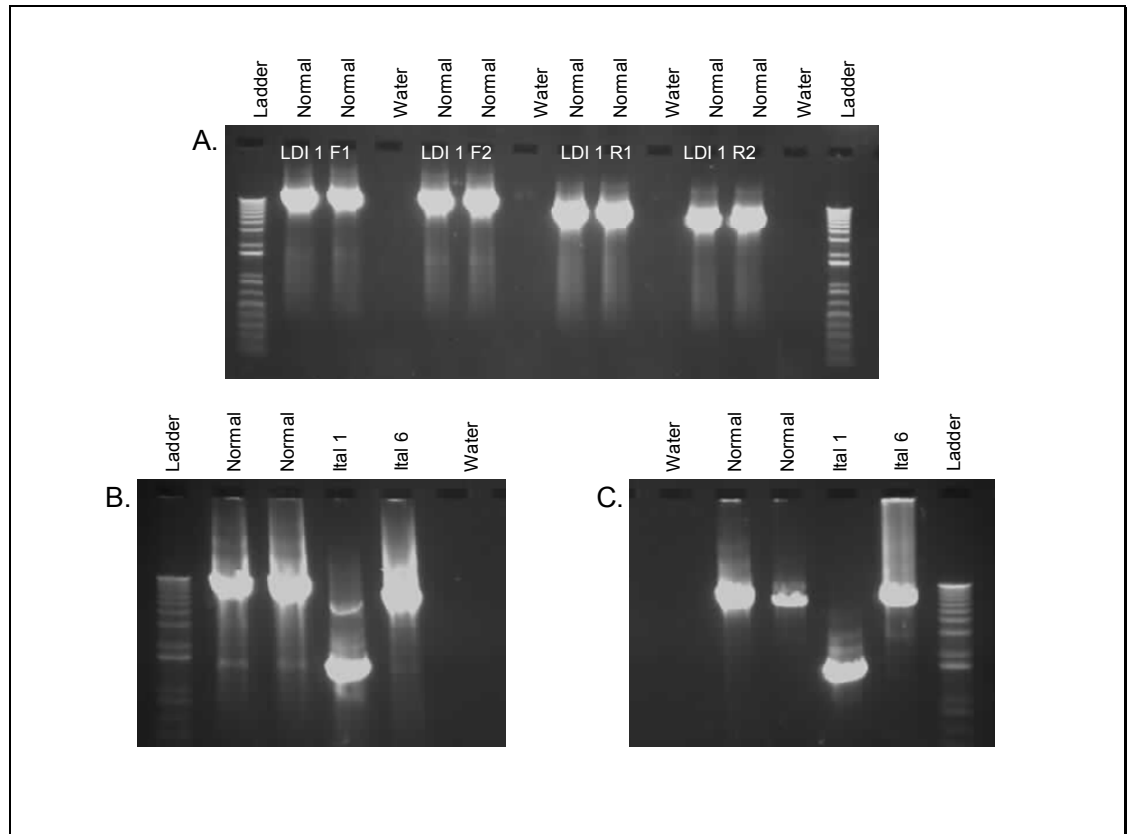


Figure 3-20: LDI-PCR products. (A) LR-PCRs using *BCR* LDI primers with the appropriate *BCR* forward and reverse primers on normal control DNA to check primers work correctly. (B) LDI-PCRs using LDI DNA dilutions of 1 in 20 in both first and nested PCRs. (C) LDI-PCRs using LDI DNA dilutions of 1 in 100 in both first and nested PCRs.

Due to the success of the first two fragments, the remaining 5 sets of primers were ordered and tested. LDI-PCR was then performed on 5 cases that were negative on the multiplex and for whom large quantities of DNA were available. However after sequencing all spurious bands, no breakpoint sequences were found and so this technique was not used any further. In retrospect, after further multiplex LR-PCR screening, Ital 5 should have given a band in LDI3 of 6269 bp but did not (see Figure 3-21).

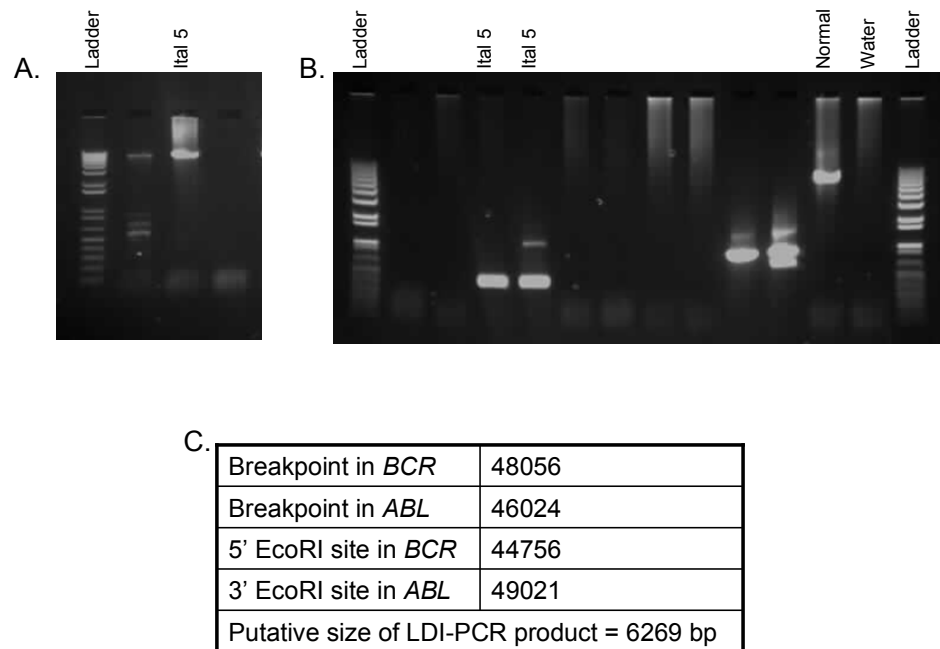


Figure 3-21: Multiplex LR-PCR breakpoint detection for Ital 5. (A) Agarose gel of amplified breakpoint band by multiplex LR-PCR. (B) LDI-PCR of Ital 5 with LDI3 which should have amplified the breakpoint band of (C) 6269 bp.

3.3.6 Multiplex LR-PCR with more primers

Due to the lack of success in finding more p190 breakpoints with alternative techniques, more multiplex LR-PCR primers were designed and used in screening. As breakpoints upstream of *ABL* exon 1b have been described in the literature and two cases in this dataset (one in a p190 ALL and a CML case see case Figures 3-26 and 3-13 respectively), 4 reverse primers upstream of *ABL* exon 1b were designed every 8 kb and used in multiplex with *BCR* A and B forward primers. In addition, all the *BCR* A - B forward and *ABL* A - D reverse primers were checked on UCSC Blat for SNPs and homology to other regions of the genome, even though this had been done when they were first designed. This resulted in 10 re-designed primers, 5 *BCR* primers and 5 *ABL* primers. These new *ABL* primers were then put into a multiplex with the remaining old primers and the new upstream *ABL* primers (see Appendix II) and tested on all remaining p190 multiplex LR-PCR negative cases.

One breakpoint, Ital 9, was amplified as a result (see Table 3-7 and Figure 3-22).

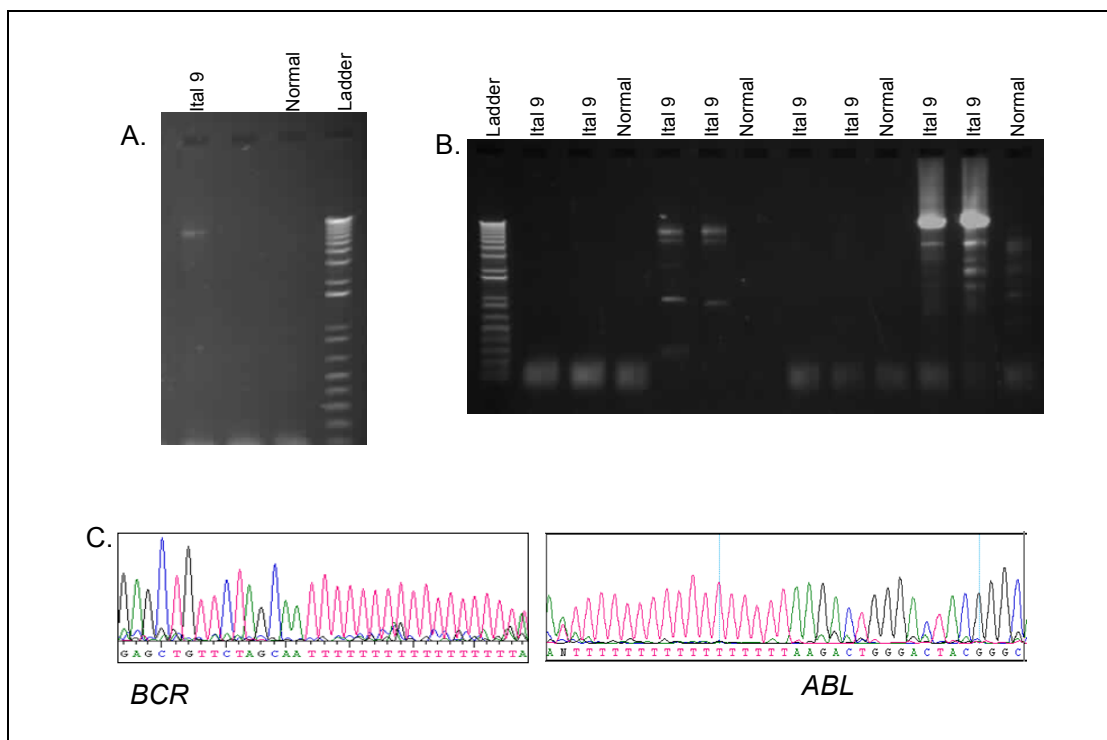


Figure 3-22: Multiplex LR-PCR breakpoint detection and sequencing for Ital 9. (A) Amplification of breakpoint band from multiplex LR-PCR screening with newly designed *BCR* 5A F and *ABL* A R primers. (B) Split apart multiplex PCR showing breakpoint bands in lanes 11 and 12 with the new *BCR* 5A F and the original *ABL* 4A R. (C) Electropherogram showing sequencing either side of the breakpoint, the precise identification of which was not possible due to the large run of Ts.

Upon comparison of the p210 and p190 breakpoints, it became apparent that one reason that there was a higher hit rate with the p210s was because of the smaller intronic breakpoint region in the M-bcr which allowed primers to be spaced up to 3 kb apart, whereas for p190 breakpoint the *BCR* primers were spaced every 8-10 kb. Therefore 11 more *BCR* primers, designated *BCR* C primers, were designed in the m-bcr of intron 1 (as no breakpoints had been found outside this region) and tested with the new multiplex *ABL* reverse primers. This screen was initially performed on 5 cases with large quantities of DNA and two further breakpoints were found for cases P653 and Ital 5 (see Table 3-8 and Figure 3-21). Upon cloning and sequencing of the amplified breakpoint bands, (see Figure 3-23 B – D) case P653 was found to have an insertion of approximately 3 kb derived from

chromosome 11p12 between the *BCR* intron 1 and *ABL* intron 1 breakpoints. This chromosome 11 material is from a gene desert and therefore, at least the forward breakpoint in this patient does not disrupt any additional genes. Searching of the Mitelman database (<http://cgap.nci.nih.gov/Chromosomes/Mitelman>) did not show any evidence for recurrent *BCR-ABL* translocations involving 11p. Further investigation of the cases karyotype revealed a three way translocation between chromosomes 9, 11 and 22 (59,XY,+2,+4,+5,+6,+7,+8,+8,t(9;22;11)(q34;q11;p11),+?10,+15,+16,+20,+21,+der(22)t(9;22;11)[9]/46,XY,del(5)(q15q33),del(7)(q22q34),add(11)(p15)[2]/ 46,XY[24]).

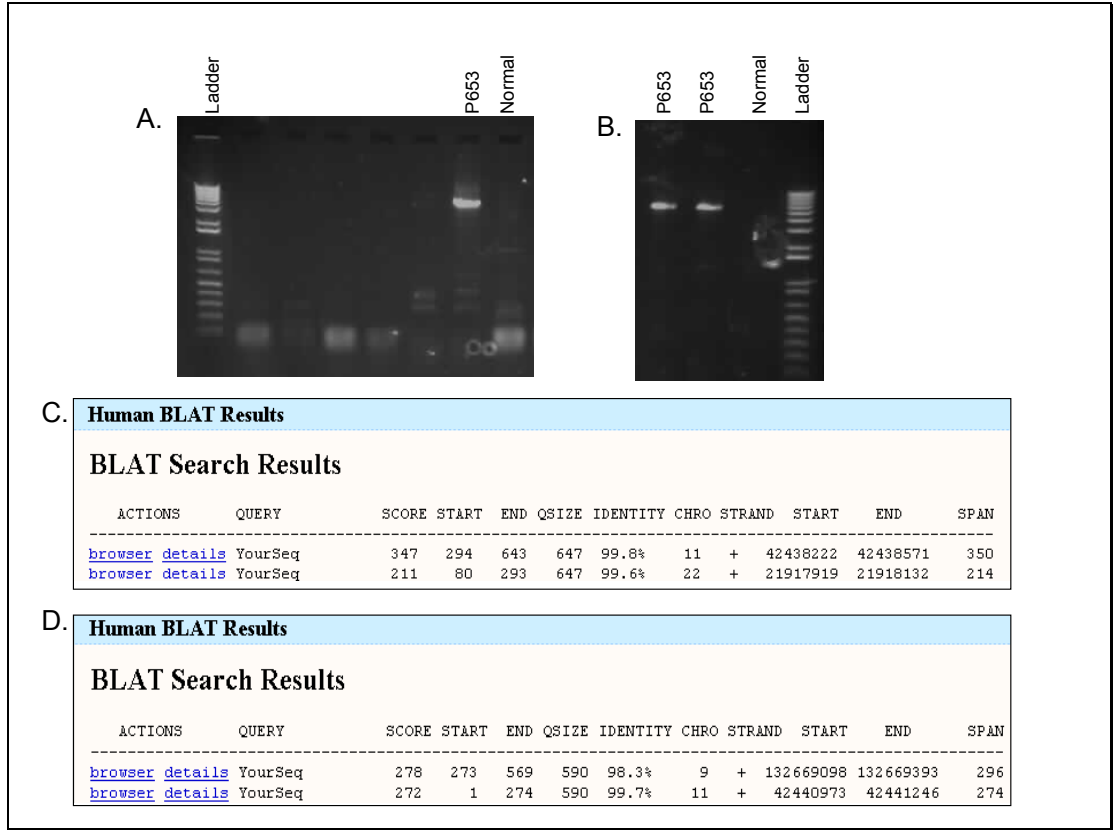


Figure 3-23: Multiplex LR-PCR breakpoint detection for P653, complex translocation. (A) Amplification of a breakpoint band in patient P653 using the new *BCR* 9C F and the original *ABL* C R mix. (B) Split apart breakpoint band amplified using *BCR* 9C F and *ABL* 4C R. (C) and D. BLAT search results from sequencing over the breakpoint, showing the insertion of chromosome 11 DNA from the complex translocation.

However, as 16 of the remaining p190 cases did not have enough DNA to perform the new screen, WGA was performed (see section 3.2.2). The WGA DNA was then tested by using LR-PCR with normal pairs of primers. It was found that the use of 20 ng input DNA per

PCR did not give bands for some of the WGA DNAs, so for these cases the amount was doubled to 40 ng per PCR, which proved to be more successful (see Figure 3-24 A and B).

Using the WGA DNA with the new *BCR* C primers and *ABL* multiplex primers, another 3 breakpoints were found for cases P747, 44240 and 11491 (see Tables 3-7 and 3-8 and Figure 3-24 C).

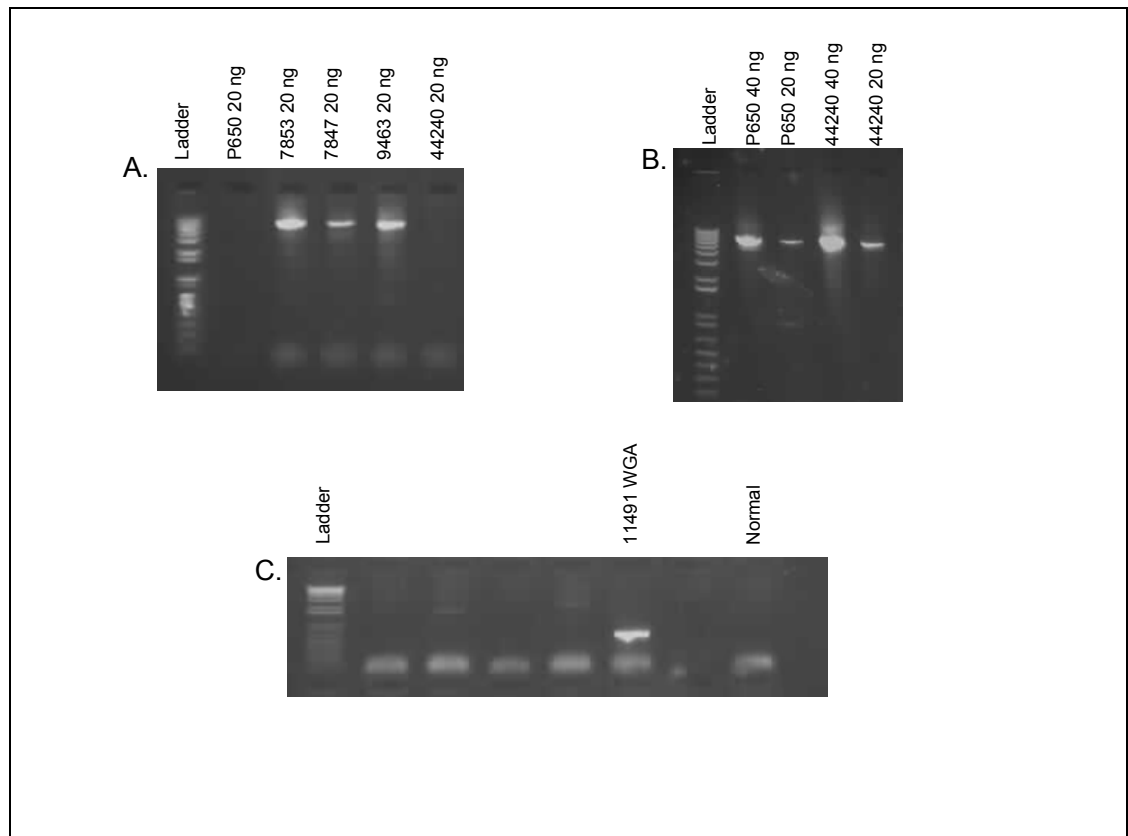


Figure 3-24: WGA LR-PCR products and multiplex LR-PCR breakpoint detection. (A) LR-PCR on WGA patient DNA at 20 ng per PCR using *ABL* forward and reverse primers to check the quality of the WGA DNA. (B) LR-PCR on WGA patient DNA at 20 ng and 40 ng per PCR using *BCR* forward and reverse primers to determine the best concentration of WGA DNA to use for screening. (C) Multiplex LR-PCR screen on WGA DNA, showing amplification of a breakpoint band in case 11491.

To delineate these breakpoints, *BCR* forward primers were designed at 1 kb intervals from the original *BCR* forward primer. Initially, a LR-PCR was performed with the new primers and, if a band was amplified, it was then sequenced with the forward primer. If the

breakpoint was not ascertained more primers would be designed, and the process repeated (see Figure 3-25).

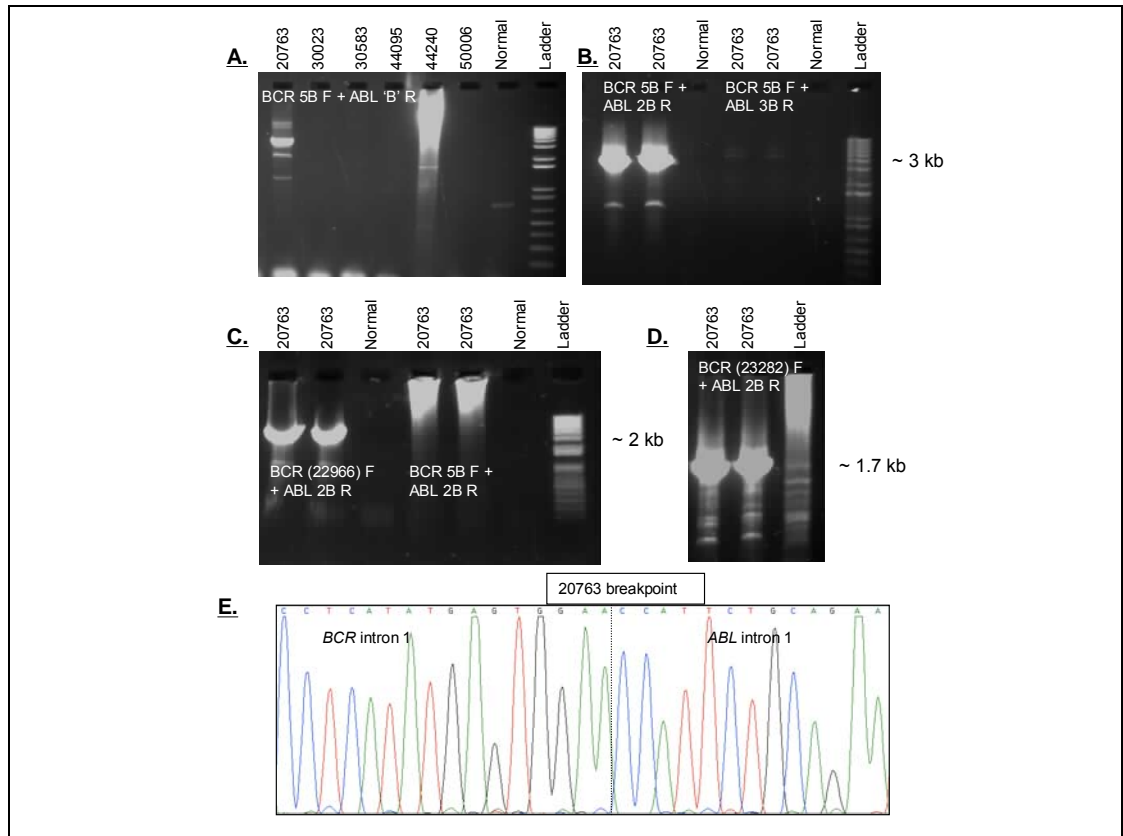


Figure 3-25: Multiplex LR-PCR breakpoint detection and sequencing for 20763. (A) Agarose gel of multiplex LR-PCR on 20763 using *BCR* 5A F and *ABL* reverse primer mix B to amplify the breakpoint. (B) Agarose gel of individual simplex LR-PCR. The breakpoint was amplified with *BCR* 5A F and *ABL* 2B R primer combination, confirming the breakpoint and ascertaining which *ABL* reverse primer to use in future PCRs. (C) and (D) Agarose gels of breakpoint bands using *BCR* primers designed 3' to the original *BCR* breakpoint primer. (E) Electropherogram of 20763 sequence obtained using *BCR* (23282) F.

However, subsequently, the original breakpoint bands in most cases were sequenced directly with new *BCR* forward primers and no further PCRs performed. Further sequencing was based on the results of the initial sequencing. In total, 43 p190 cases were screened (see Tables 3-5, 3-7 and 3-8) and 25 (58%) breakpoints were detected and sequenced (see Figure 3-26 and Appendix III).

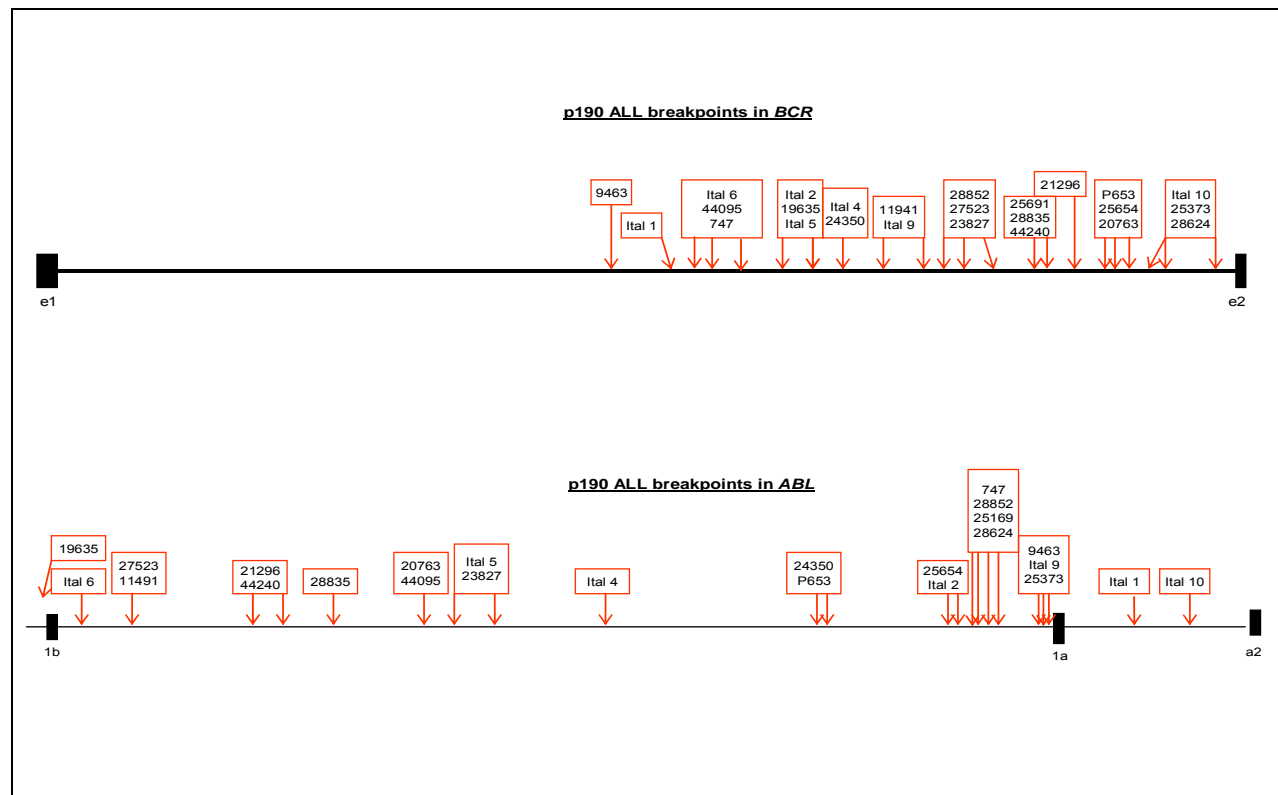


Figure 3-26: Maps showing *BCR* and *ABL* forward breakpoints in p190 ALL.

3.3.7 Reciprocal breakpoint detection and sequencing

Once the forward breakpoints (*BCR-ABL*) were sequenced, primers were designed to allow amplification by LR-PCR of the reciprocal breakpoints (*ABL-BCR*) on the derivative chromosome 9 using the appropriate *ABL* forward and *BCR* reverse primers flanking the breakpoint (see section 3.3.7.1 for an example). The sequence of both breakpoints is important as it may help to indicate how the translocations arose, by determining if there are any sequence signatures common to certain types of breakpoint repair.

3.3.7.1 Example of forward and reciprocal breakpoint determination for case Ital 10

The *BCR-ABL* breakpoint in Ital 10 was detected by use of BCR Bt 3 5B (22728) F and ABL 5B (16551) R (see Figure 3-27 A and Figure 3-17 for gel of forward breakpoint). The forward breakpoint was located to *BCR* Bt 3 at 24476 (NM_004327 at 67879) and *ABL* 5 at 13846 (NM_007313 at 138137) by sequencing using BCR Bt 3 (24281) F. Knowing the forward breakpoint in *BCR-ABL* enabled primers to be designed to detect the reciprocal breakpoint. Reciprocal primers were typically designed to amplify fragments of 4-8 kb if the translocation was balanced, thus allowing for moderate deletions which are commonly seen at reciprocal translocation breakpoints (see Figure 3-27 A for primer names). In this case, once the reciprocal breakpoint was amplified and was found to be the approximate predicted size of 4000 bp (actual predicted size 3898 bp if no deletions/insertions; see Figure 3-27 B), a further primer, ABL 5 (13568) F was designed to be close enough to the predicted reciprocal breakpoint (at ABL 5 13846) to sequence across it. This revealed a small deletion of 27 bp (see Figure 3-27 C for sequencing results).

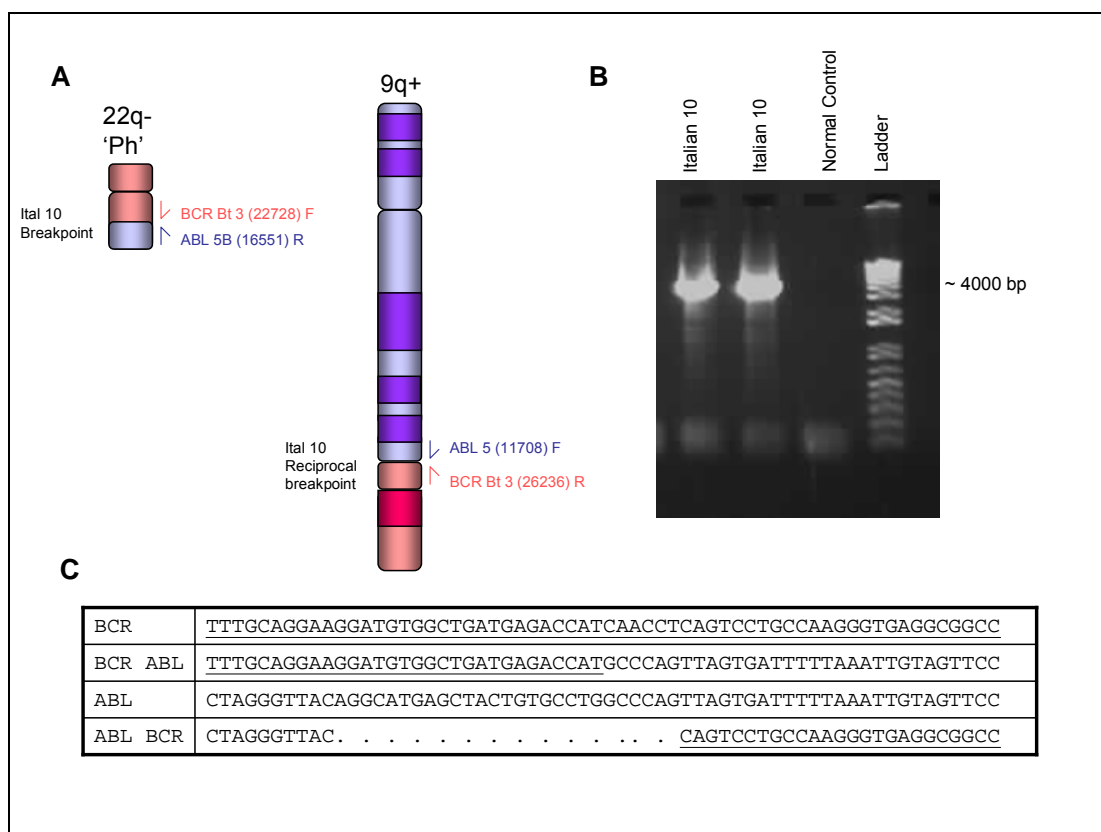


Figure 3-27: Reciprocal breakpoint detection using LR-PCR for Itai 10 p190 ALL. (A) Ideograms showing primers used to detect both the breakpoint and the reciprocal breakpoint respectively. (B) Agarose gel of Itai 10 reciprocal breakpoint LR-PCR product, amplified using *ABL* 5 (11708) F and *BCR* Bt 3 (26236) R primers. (C) Table showing sequence of both the breakpoint *BCR-ABL* and reciprocal breakpoint *ABL-BCR* in comparison to the normal *BCR* (underlined) and *ABL* sequences, the dots indicate a deletion in the reciprocal DNA.

3.3.7.2 Results of reciprocal breakpoint detection by LR-PCR

LR-PCR was used to screen for reciprocal breakpoints in all 82 cases in whom forward breakpoints had been found. This yielded 46 (56%) positive bands which were sequenced and confirmed. Of these, 16 were from cases with p210 CML (16/32; 50%), 10 with p210 ALL (10/25; 40%) and 20 with p190 ALL (20/25; 80%). See Table 3-10 for examples of different types of reciprocal breakpoints and Appendix III for all sequences. The number of cases with perfectly balanced translocations was 5 of the p190s, 2 of the p210 ALLs and 0 of the p210 CMLs. The median number of nucleotides deleted and the range were; 3 bp for the p190s (n=12; range 1-27), 37 bp for the p210 ALLs (n=6; range 3-7883) and 33 bp for the CMLs (n=13; range 1-985). Short stretches of micro-homologies (2-3 bp) were found in 4 of the p190s (25373, 20763, 22852 and 28624), 1 of the p210 ALLs (F17237) and 1 of

the p210 CMLs (1716). Duplications of sequences from either or both genes were found in 5 of the p190s (Ital 2 = 3bp, Ital 6 = 4 bp, Ital 9 = 4bp, Ital 4 = 5 bp and 19635 = 3 bp), 2 of the p210 ALLs (10258 = 7 bp and 10887 = 3 bp) and 7 of the p210 CMLs (248 = 3 bp, 470 = 3 bp, 403 = 4 bp, 286 = 173, 266 = 362 bp, 1716 = 534 bp).

44095	No deletion, duplication or insertion
<i>BCR</i>	<u>GGGGAAGTGGTTCCCGAGGAAGGAGGCAGAGGAGTGTAGATTTTGTAGAGAAGGCTAAGAGAGG</u>
<i>BCR ABL</i>	<u>GGGGAAGTGGTTCCCGAGGAAGGAGGCAGAGGATCTTCAGATATCTAAAAGAACTTTGTAAACA</u>
<i>ABL</i>	GGGGATGTTTCTGTTGTTGCTCCTTGCTGAGTATCTTCAGATATCTAAAAGAACTT . GTAACA
<i>ABL BCR</i>	GGGGATGTTTCTGTTGTTGCTCCTTGCTGAGT <u>AGTGTAGATTTTGTAGAGAAGGCTAAGAGAGG</u>
25373	Identical short stretches of homology
<i>BCR</i>	<u>TCGTCTTCAGATAAACAGCAAAACCGTTGGTTAAAAGCTGGTCTAACCACCTTATTTTGTGTGT</u>
<i>BCR ABL</i>	<u>TCGTCTTCAGATAAACAGCAAAACCGTTGGTTT</u> TACAAGACTATTTTTCAGCCTTTTTTACAAGA
<i>ABL</i>	GAAAGGGATTGCTGTTTTTTATTACCAGATTTTACAAGACTATTTTTCAGCCTTTTTTACAAGA
<i>ABL BCR</i>	GAAAGGGATTGCTGTTTTTTATTACCAGA . <u>TTAAAAGCTGGTCTAACCACCTTATTTTGTGTGT</u>
Italian 1	Duplication within one gene
<i>BCR</i>	<u>GGGACTCGTGTGGGGACAAGCTGAACAGTGTGCTGATCTCTGATCACCCTCAAGCCACAAGTG</u>
<i>BCR ABL</i>	<u>GGGACTCGTGTGGGGACAAGCTGAACAGTGTGTTAGCAGCCTAAAATGGTGGCTTTACCAGTTT</u>
<i>ABL</i>	AAAAGTAAGCTGTTGTTACTTTAAGCAAATATTTAGCAGCCTAAAATGGTGGCTTTACCAGTTT
<i>ABL BCR</i>	AAAAGTAAGCTGTTGTTACTTTAAGCAAATATTTAGATCTCTGATCACCCTCAAGCCACAAGT
Italian 6	Duplication within one gene, deletion within the other
<i>BCR</i>	<u>GTTCTGCCACCTGCCCTTTGAGGCTGTGTGGGCACGGTTCCTCTAGAGCTCCATGTCTAGCAG</u>
<i>BCR ABL</i>	<u>GTTCTGCCACCTGCCCTTTGAGGCTGTGTGGGAGTGAGATCTTACCTCTACAGAAATTTAAAAA</u>
<i>ABL</i>	AGGTCAGGAGTTCGAACCCAGCTGTGCAATCTAGTGAGATCTTACCTCTACAGAAATTTAAAAA
<i>ABL BCR</i>	AGGTCAGGAGTTCGAACCCA <u>TGGGCACGGTTCCTCTAGAGCTCCATGTCTAGCAG</u>
Italian 10	Deletion within both genes
<i>BCR</i>	<u>CTTTGCAGGAAGGATGTGGCTGATGAGACCATCAACCTCAGTCCTGCCAAGGGTGAGGCGGCCC</u>
<i>BCR ABL</i>	<u>CTTTGCAGGAAGGATGTGGCTGATGAGACCAT</u> GCCCAGTTAGTGATTTTAAATTGTAGTTCCT
<i>ABL</i>	TCTAGGGTTACAGGCATGAGCTACTGTGCCTGGCCAGTTAGTGATTTTAAATTGTAGTTCCT
<i>ABL BCR</i>	TCTAGGGTTAC CAGTCCTGCCAAGGGTGAGGCGGCCC

9463	Deletion within one gene
<i>BCR</i>	<u>TCCTTTTGAAGGTCATTGTTTTCTTAACAGAT</u> TGGTTATTTGGACTCTTACAAGTTCTCTGTA
<i>BCR ABL</i>	<u>TCCTTTTGAAGGTCATTGTTTTCTTAACAGAGCAATCCCAGGGCTTTAATATCCTTTTTGTA</u>
<i>ABL</i>	GTCTCAAAAAAAAAAGAAAAGAAATTAGAAACT CAATCCCAGGGCTTTAATATCCTTTTTGTA
<i>ABL BCR</i>	GTCTCAAAAAAAAAAGAAAAGAAATTAGAAACT . <u>TTATTTGGACTCTTACAAGTTCTCTGTA</u>
25373	Deletion of 1bp in one gene
<i>BCR</i>	<u>TCGTCTTCAGATAAACAGCAAAACCGTTGGTTAAAAGCTGGTCTAACCACCTTATTTTGTGT</u>
<i>BCR ABL</i>	<u>TCGTCTTCAGATAAACAGCAAAACCGTTGGTTT</u> TACAAGACTATTTTTCAGCCTTTTTCACAAGA
<i>ABL</i>	GAAAGGGATTGCTGTTTTTTATTACCAGATTTTACAAGACTATTTTTCAGCCTTTTTCACAAGA
<i>ABL BCR</i>	GAAAGGGATTGCTGTTTTTTATTACCAGA . <u>TTAAAAGCTGGTCTAACCACCTTATTTTGTGT</u>

Table 3-10: Forward and reciprocal breakpoints of 7 p190 ALL cases with the corresponding normal sequence. Short regions of homology are shown in bold, *BCR* derived sequences are underlined and deleted sequences are shown as dots.

3.3.8 Detection of deleted reciprocal breakpoints by MLPA

One possibility for the failure to amplify reciprocal breakpoints in many cases could be the presence of large deletions. Of the 36 cases that were negative for reciprocal breakpoints by LR-PCR, 30 had suitable sample quality to be tested by MLPA (see Table 3-11).

p190 ALL	MLPA results	p210 ALL	MLPA results	p210 CML	MLPA results
Italian 5	Failed	P851	Not deleted	NP241	<i>BCR</i> and <i>ABL</i> deleted
21296	Small <i>ABL</i> deletion	P682	Not deleted	E656	Not done
27523	Not deleted	22251	<i>BCR</i> deletion	1349	<i>BCR</i> and <i>ABL</i> deleted
24350	Not done	26219	Not deleted	1578	<i>ABL</i> deleted
P653	Not done	24375	Not done	1810	Not done
		11128	<i>BCR</i> and <i>ABL</i> deleted	1538	Failed
		F17685	<i>BCR</i> and <i>ABL</i> deleted	230	Not done
		F17730	Failed	440	Not deleted
		F18544	Failed	446	Not deleted
		F16604	<i>ABL</i> deleted	495	Not deleted
		F15010	Failed	545	Not deleted
		F22757	Failed	282	<i>BCR</i> and <i>ABL</i> deleted
		F11329	Not deleted	287	Not deleted
		F24559	Failed	291	Not deleted
		F20099	<i>ABL</i> and <i>EXOSC1</i> deleted	300	<i>BCR</i> deleted
				317	Not deleted

Table 3-11: MLPA results on p190 and p210 (ALL and CML) cases who were negative for reciprocal breakpoints by LR-PCR.

Of the 23 cases that gave informative results, 12 showed no deletions of the MLPA probes and therefore had no evidence of large deletions on the derivative chromosome and the remaining 11 (48%) showed deletions of one or more of the MLPA probes (see Figure 3-28). This correlated to 1/2 p190 cases, 5/9 p210 ALL cases and 5/12 p210 cases (see Table 3-11). The proportions of large deletions as determined by MLPA in each of the disease subtypes (including LR-PCR positive results but excluding failed or inadequate samples) were therefore; 5/28 (18%) for p210 CMLs, 5/19 (26%) for p210 ALLs and 1/23 (4%) for p190s. Identification of the reciprocal translocations by LR-PCR revealed that the number of perfectly balanced translocations was 5/23 (22%) for p190s, 2/19 (11%) for p210 ALLs and 0/28 (0%) for p210 CMLs.

1349		BCR1	ABL1	BCR16	FLJ31568-2	PROM12-2	ASS 4	BCR 2	ABL 6	ABL 11	EXOSC2-7
BCR1			1.95	2.8279372	2.0854723	2.103129	2.2943287	1.03630829	1.087400357	1.085074211	2.33086
ABL1		0.548634	1	1.096371752	1.127477221	1.1370224	1.207354002	0.9602639	0.9878858	0.986628	1.2399825
BCR16		0.4331119	0.912	1	1.028371279	1.0370774	1.107774101	0.511016	0.5362101	0.535063	1.1432384
FLJ31568-2		0.4795977	0.887	0.97241144	1	1.008466	1.071377744	0.4963179	0.5214168	0.5282814	1.175326
PROM12-2		0.4754823	0.879	0.964248165	0.991605118	1	1.062383655	0.4927462	0.5170396	0.5159335	1.108151
ASS 4		0.4425617	0.828	0.907627071	0.933377811	0.9412795	1	0.4638119	0.4866288	0.4856372	1.0430739
BCR 2		0.964963814	1.78	1.356886	2.0124053	2.823442	2.1560463		1	1.049301996	1.04705735
ABL 6		0.918624491	1.7	1.9649407	1.9178514	1.934088	2.0547434	0.953014494		1	0.997860819
ABL 11		0.921595951	1.7	1.8689387	1.9219629	1.938234	2.0591482	0.95057524	1.002143767		1
EXOSC2-7		0.4290771	0.794	0.870141457	0.894828483	0.9024041	0.958899322	0.4446562	0.4665786	0.4655805	1

1578	Exons	BCR1	ABL1	BCR16	FLJ31568-2	PROM12-2	ASS 4	BCR 2	ABL 6	ABL 11	EXOSC2-7
BCR1			1	1.084725367	1.096843435	1.945023	2.0297886	0.957583214	1.019337337	0.956051748	2.23246
ABL1		0.4771927	1	0.517623	0.5234057	0.9281505	0.969600243	0.4569517	0.4864203	0.4562209	1.065314
BCR16		0.921892333	1.93	1	1.01171854	1.793101	1.8712465	0.882788623	0.939719276	0.881376777	2.05089
FLJ31568-2		0.911707148	1.91	0.988951871	1	1.773291	1.8569728	0.873035461	0.929337106	0.871639213	2.03535
PROM12-2		0.5141328	1.077	0.5576329	0.5639232		1	0.492325	0.5240748	0.4915376	1.1477815
ASS 4		0.44286824	1.032	0.5344831	0.5498232	0.952239		1	0.471765	0.5821699	0.4710105
BCR 2		1.04429568	2.18	1.132774	1.145428945	2.931178	2.1196994		1	1.064489563	0.989400637
ABL 6		0.991029502	2.06	1.064147587	1.076035769	1.909125	1.9912825	0.939417384		1	0.93791497
ABL 11		1.045968487	2.13	1.124588951	1.147263668	2.634432	2.1230949	1.001601865	1.066194731		1
EXOSC2-7		0.4473362	0.933	0.4858877	0.4313153	0.871246	0.905215742	0.4289362	0.4565381	0.4282582	1

Figure 3-28: Peak areas obtained from MLPA exported to Microsoft Excel. The Excel spreadsheet was set up to assess the ratios of each peak relative to all other peaks for each patient and also to two healthy controls on each run. Ratios of 0.5 and 1.5 represent deletions or duplications respectively. For any given probe a horizontal red row and a corresponding vertical green column represent a deletion for that probe.

3.3.9 Statistical analysis

All statistical analyses were performed by Ru-Fang Yeh and Joe Wiemels. In addition to the breakpoints found in this study, 18 additional p210 CML forward breakpoints were amplified and sequenced using the LR-PCR multiplex described here but screened under the supervision of Junia Melo at the Hammersmith (see Appendix 1V for positions of breakpoints). For the full analysis see supplementary information.

3.3.9.1 Forward breakpoint locations

The Komolgorov-Smirnov test revealed that there were no significant differences in the distribution of forward breakpoints between the different subtypes of leukaemia (see Table 3-12).

Comparison Komolgorov-Smirnov test p-value	
<i>ABL</i> .ALL.vs.CML	0.967
<i>ABL</i> .p190.vs.p210	0.6
<i>ABL</i> .p190ALL.vs.p210 ALL	0.915
<i>ABL</i> .p190ALL.vs.p210 CML	0.637
<i>BCR</i> .p210ALL.vs.p210 CML	0.915

Table 3-12: Analysis of the breakpoint locations in the subtypes of leukaemia by the Komolgorov-Smirnov test. Any p-value below 0.05 reveals a significant difference.

3.3.9.2 Clustering of breakpoints

Analysis of the positions of the forward breakpoints in *BCR* for p210 ALL gave evidence for two distinct clusters of breakpoints (red line, Figure 3-29), whereas in p210 CML these, peaks are not so distinct (see blue line, Figure 3-29 A).

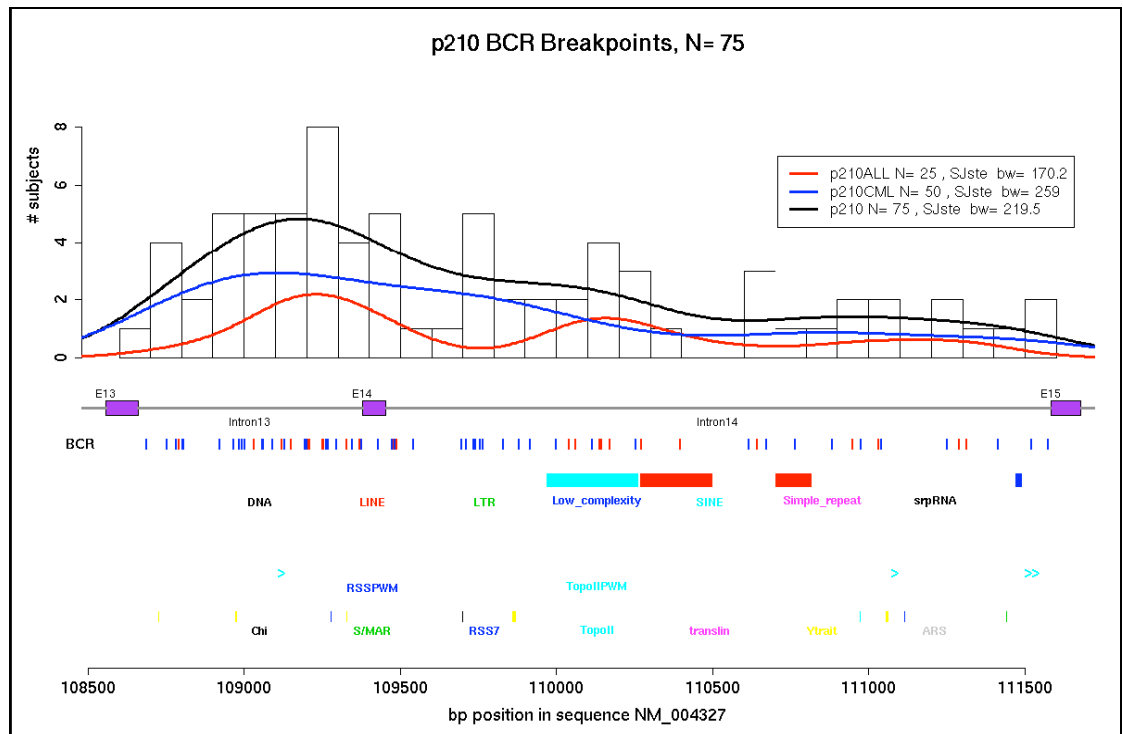


Figure 3-29: Analysis from clustering of p210 *BCR* breakpoints. Density curves were determined by using the Sheather-Jones solve-the-equation bandwidth. The coloured boxes below the breakpoints show the locations of different classes of repetitive elements as determined by RepeatMasker, and known motifs (exact match) using Fuzznuc or PWM. The peaks indicate clustering, which is particularly apparent with the p210 ALL breakpoints which shows 2 clusters.

The breakpoints in *BCR* for p190 ALL show a flat line and therefore no evidence of clustering around any specific region in these cases (see Figure 3-30), although it is striking that all breakpoints fall in the second half of *BCR* intron 1, previously defined as the m-bcr.

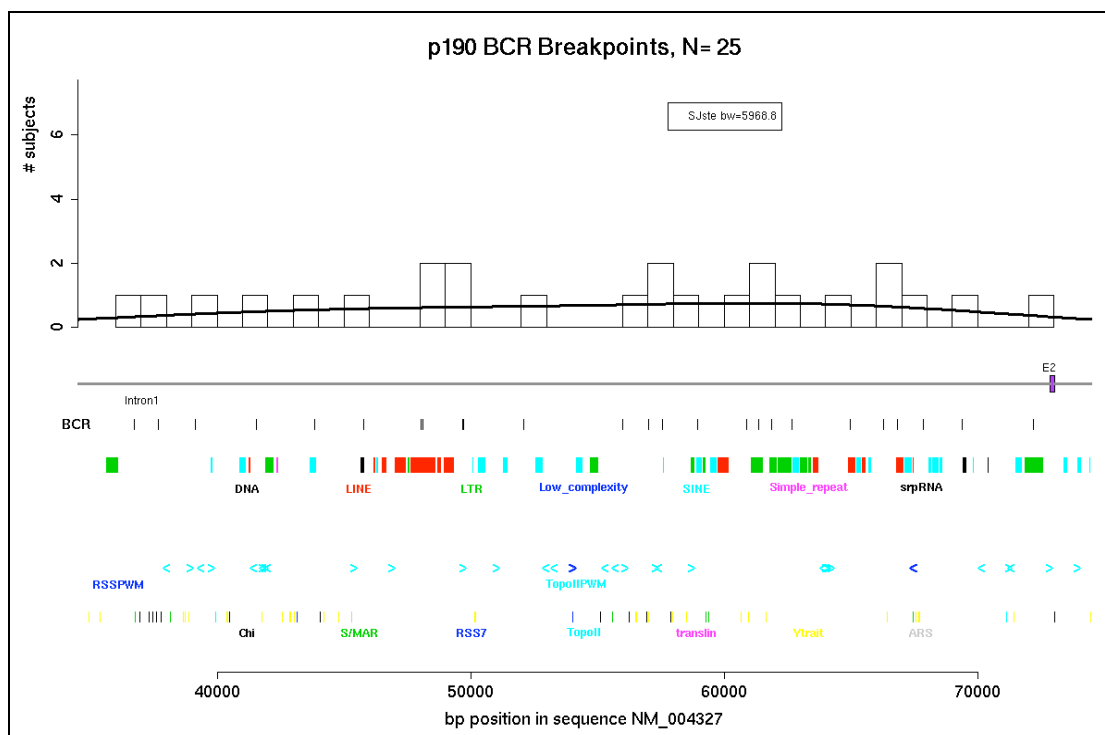


Figure 3-30: Analysis from clustering of p190 *BCR* breakpoints. Density curves were determined by using the Sheather-Jones solve-the-equation bandwidth. The coloured boxes below the breakpoints show the locations of different classes of repetitive elements as determined by RepeatMasker, and known motifs (exact match) using Fuzznuc or PWM. The flat line reveals that there is no evidence of clustering.

The breakpoints in *ABL* show that there is some evidence for clustering of breakpoints in the 3' end of *ABL* for p190 ALL and evidence of two clusters for p210 ALL, and although there appears to be no clustering for the p210 CML cases, all the cases combined show amplified clustering in two parts of *ABL* (see Figure 3-31 C).

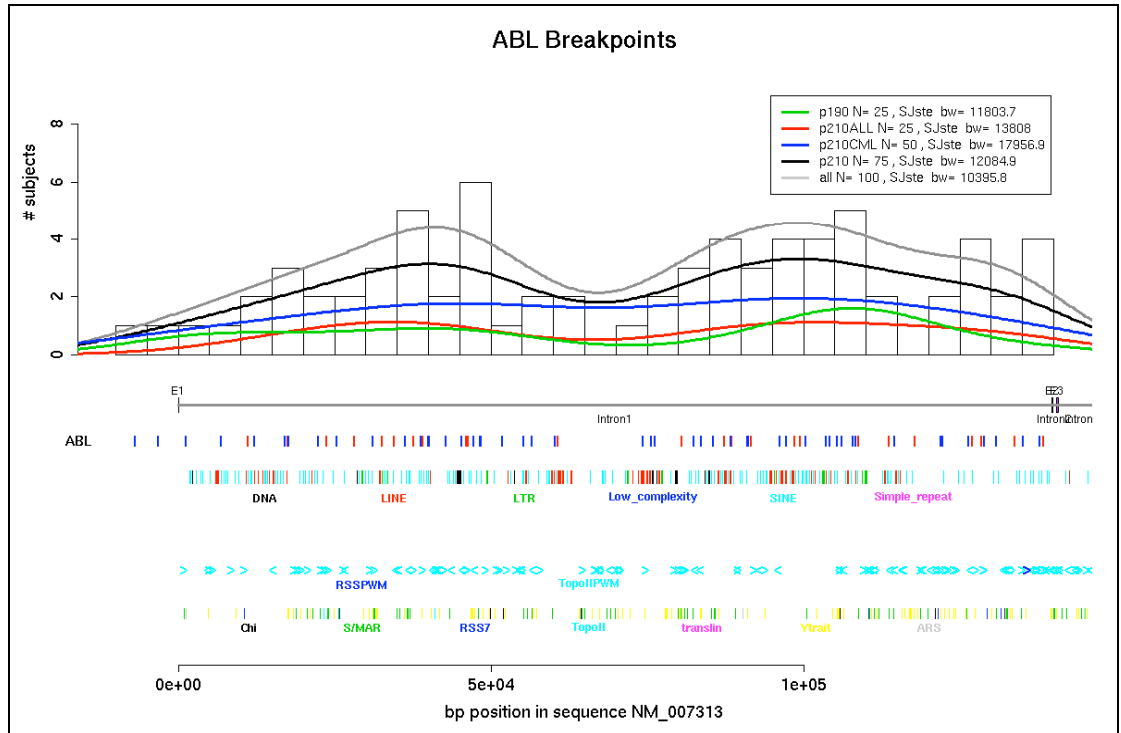


Figure 3-31: Analysis from clustering of p210 CML, ALL and p190 ALL *ABL* breakpoints. Density curves were determined by using the Sheather-Jones solve-the-equation bandwidth. The coloured boxes below the breakpoints show the locations of different classes of repetitive elements as determined by RepeatMasker, and known motifs (exact match) using Fuzznuc or PWM. The peaks may indicate some clustering in the p190 and p210 ALL breakpoints, conversely the flat line for p210 CML indicates no clustering.

In addition to clustering, the positions of the breakpoints in *BCR* were directly compared to those in *ABL* for all p210 and p190 breakpoints, however, no significant associations were found (see Figure 3-32).

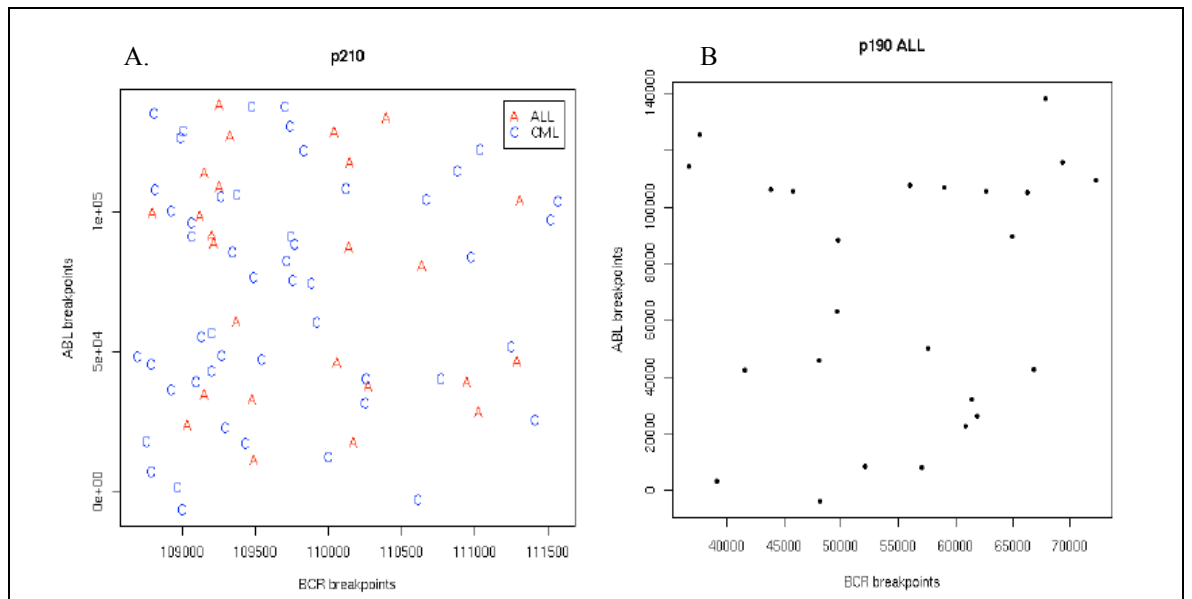


Figure 3-32: Positions of *ABL* versus *BCR* breakpoints in (A) p210 ALL and CML and (B) p190 ALL. If there were associations between the positions of the breakpoints in *BCR* and *ABL*, there would be clustering of data points on the graphs.

3.3.9.3 Proximity of breakpoints to sequence motifs

In order to check if the clustering of breakpoints was a result of their proximity to known sequence motifs, including RSSs, 50 bp of sequence flanking the breakpoints were compared to specific motifs. One significant association was found: a deficit of breakpoints falling in repeat regions for p210 CML *BCR* breakpoints and all p210 *BCR* breakpoints (see Figure 3-33 B and D).

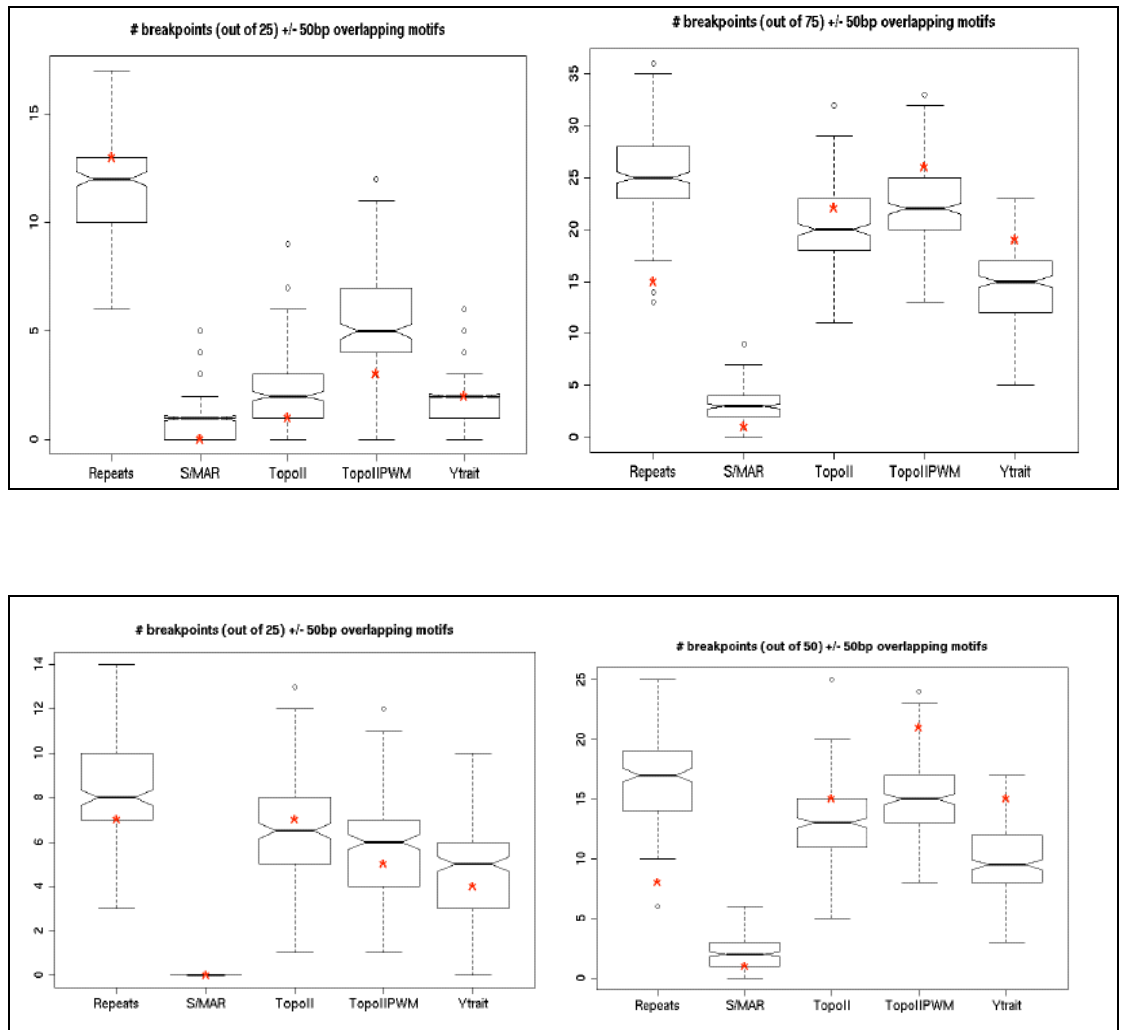


Figure 3-33: Proximity of breakpoints near sequence motifs. Box and whisker plots showing the percentage of breakpoints in BCR against sequence motifs (within 50 bp either side of the breakpoints). The red points are observed from the breakpoint data and the box plot is based on data generated from 200 simulations. Each simulation calculates the percentage of N breakpoints overlapping any of the motifs. (A) p190 ALL. (B) p210 ALL and CML. (C) p210 ALL. (D) p210 CML.

In addition to seeing if specified known motifs were associated with the breakpoints, undirected searches for motifs were performed to see if any unknown motifs were enriched 50 bp either side of the breakpoints. The only consistent result generated by this method was an association of CTGNNTTG with the *BCR* breakpoints in p210 ALL, which was found in 3 separate runs. This motif has no known function and so the significance of this finding is unclear.

3.3.10 Screening for a common sequence motif associated with recombination hot spots and genome instability in humans

A recent paper by Myers *et al.*, (260) identified a 13-mer degenerate sequence motif (CCNCCNTNNCCNC) that is enriched in short regions of the genome involved in meiotic crossover events, but importantly has also been found in disease causing recombination hotspots. Therefore screening was undertaken using Fuzznuc of the *BCR* and *ABL* genes and 50 bp either side of each of the *BCR-ABL* breakpoints (see Appendix IV). Screening of *ABL* revealed that the 13-mer motif occurs on average once every 4788 bp in the large *ABL* breakpoint region and once every 4952 bp in the whole *ABL* gene. Screening of the breakpoint regions revealed that only 2 CML cases (266 and 483) had *ABL* breakpoints that were located within 50 bp of a motif. Screening of *BCR* revealed a larger difference between breakpoint regions, the entire length of the *BCR* gene had a motif every 7615 bp, whilst the m-bcr had one every 4479 bp and the M-bcr had 2 motifs (NM_004327 at 108822 and 111061) giving it a frequency of one every 1511 bp. There were no motifs within 50 bp of p190 breakpoints but were 3/50 p210 CML and 3/25 p210 ALL (see Table 3-13). The previously identified 7-mer motif (CCTCCCT) has been linked to non-homologous DNA recombination and is part of the family of motifs from which the 13-mer degenerate motif is derived (261). This motif occurred every 1679 bp in the *BCR* gene as a whole, every 1707 bp m-bcr, but every 378 bp in the M-bcr. Furthermore, it was found in *BCR* within 50 bp of the breakpoint in 5/25 p210 ALL, 9/50 p210 CML and in 0/25 of the p190 ALL breakpoints (P=0.018; p210 versus p190; Fisher's exact test). The occurrence in the *ABL* breakpoint region was more than 4-fold lower, with a motif frequency of one every 7812 bp and none found to be within 50 bp of any of the breakpoints (see Table 3-13).

	Frequency of 13-mer motif (per bp)	Frequency of 7-mer motif (per bp)
<i>BCR</i> gene	7615	1679
m-bcr	4479	1707
M-BCR	1511	378
<i>ABL</i> breakpoint region	4788	7812
p190 ALL breakpoints	0/25 (0%)	0/25 (0%)
p210 ALL breakpoints	3/25 (12%)	5/25 (20%)
P210 CML breakpoints	3/50 (6%)	9/50 (18%)

Table 3-13: Frequencies of recombination motifs in the *BCR* and *ABL* breakpoint regions

3.3.11 Functional analysis of *BCR-ABL* breakpoints for cryptic RSSs

As previously mentioned, it is difficult to identify cryptic RSS sites by sequence homology alone because of the high degree of degeneracy these sites show. A rapid and robust *in vitro* assay that will give quantitative results is necessary to determine if RAG mis-recognition is responsible for the *BCR-ABL* fusion gene in some ALL cases.

One such assay was first described in 1987 by Hesse *et al.* (262), and was originally designed to study the mechanisms of how V(D)J coding and signal sequences were formed. This assay has subsequently been optimised and recently used to analyse the role of RAG mis-recognition in the lymphoid translocation breakpoints by substituting one of the consensus RSSs for common breakpoint cluster regions or known breakpoints, and thus demonstrating that some recurrent breakpoint clusters can act as RAG targets (207, 254, 263).

The method of Raghavan *et al.*, (263) utilises a plasmid which can replicate in either mammalian cells or *E.coli* cells. It confers ampicillin (Amp) resistance all the time but chloramphenicol (Cam) resistance is conditional on a RAG-mediated recombination event having taken place, enabling correct transcription of the *cat* gene. This is achieved by the presence of a stop sequence in the parental plasmid flanked by the two putative RSSs. If no recombination occurs, the stop sequence will prevent the expression of the downstream *cat* gene, however, if recombination takes place the stop sequence is excised and the *cat* gene is then placed under the control of a transcriptional promoter. Plasmids are constructed with the test sequence (in this study this will be the sequence immediately flanking the *ABL* or *BCR* breakpoints) plus an authentic 12 or 23 spacer RSS sequence. The plasmid containing the breakpoint sequence to be tested is transfected into the RAG producing Reh cell line and left for 48 hours to allow recombination events to take place. Plasmids are then recovered and transformed into *E.coli* for selection of recombination positive cells using Cam and Amp plates (see Figure 3-34). The number of Cam resistant colonies is a measure of the ability of the test sequence to act as a RAG recognition sequence. Use of this plasmid would have been ideal, however despite initial assurances from the authors that it was available, it proved impossible to obtain it from the authors and therefore making of a similar plasmid was undertaken (we already had the Reh cell line).

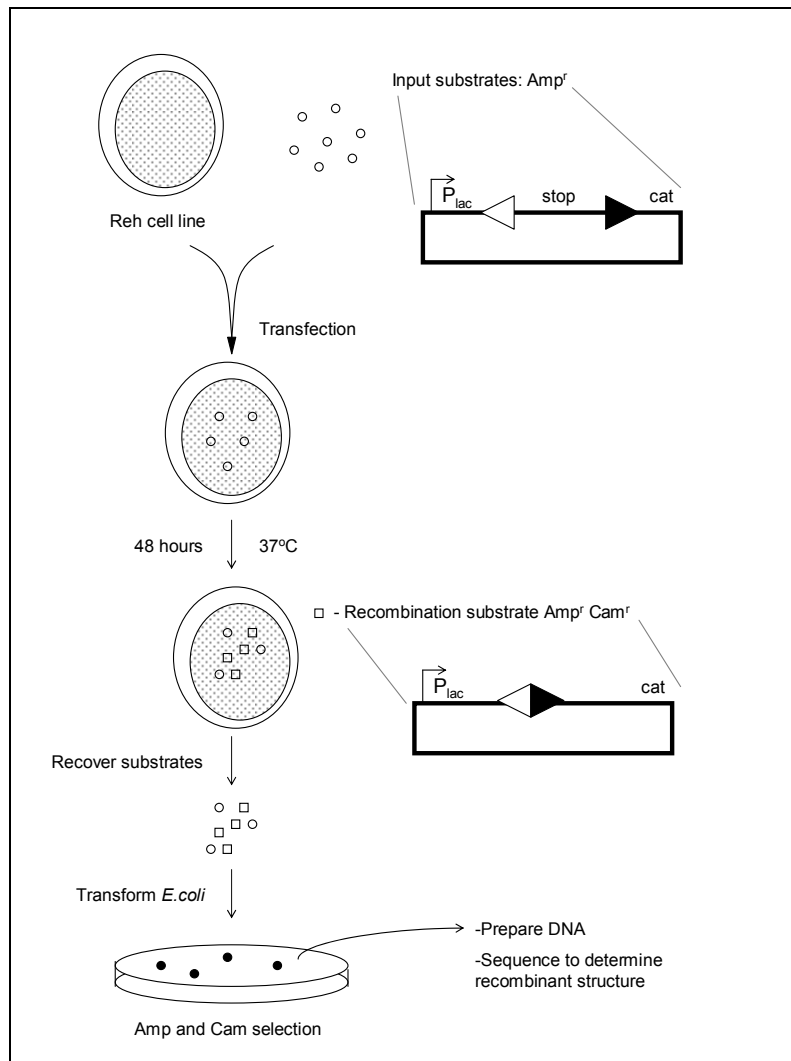


Figure 3-34: Modified from Raghavan *et al.*, (263), depicting the experimental procedure they used to detect V(D)J recombination. Circular plasmids are introduced into the human Reh cell line active for V(D)J recombination. After 48 hours, the episomes are harvested and transformed into *E.coli* for detection of recombinants on Amp plus Cam plates. The recombination is depicted between a consensus 12 (open triangle) and 23 (dark triangle) signal, leading to a signal joint formation. Stop denotes the prokaryotic bacteriophage lambda transcription terminator and the *E.coli* lac promoter is denoted as *Plac*.

3.3.11.1 Measuring RAG recombination using the pGlow-TOPO expression vector.

Initial experiments were performed using the pGlow-TOPO (Invitrogen) expression vector, whereby the strategy was to insert a promoter flanked by RSSs which would drive expression of the downstream *GFP* gene and therefore fluoresce upon transformation into *E.coli* cells. If a recombination event in the Reh cells excised the promoter, transcription of

the fluorescent *GFP* gene would not take place and the transformed *E.coli* would not fluoresce (see Figure 3-35).

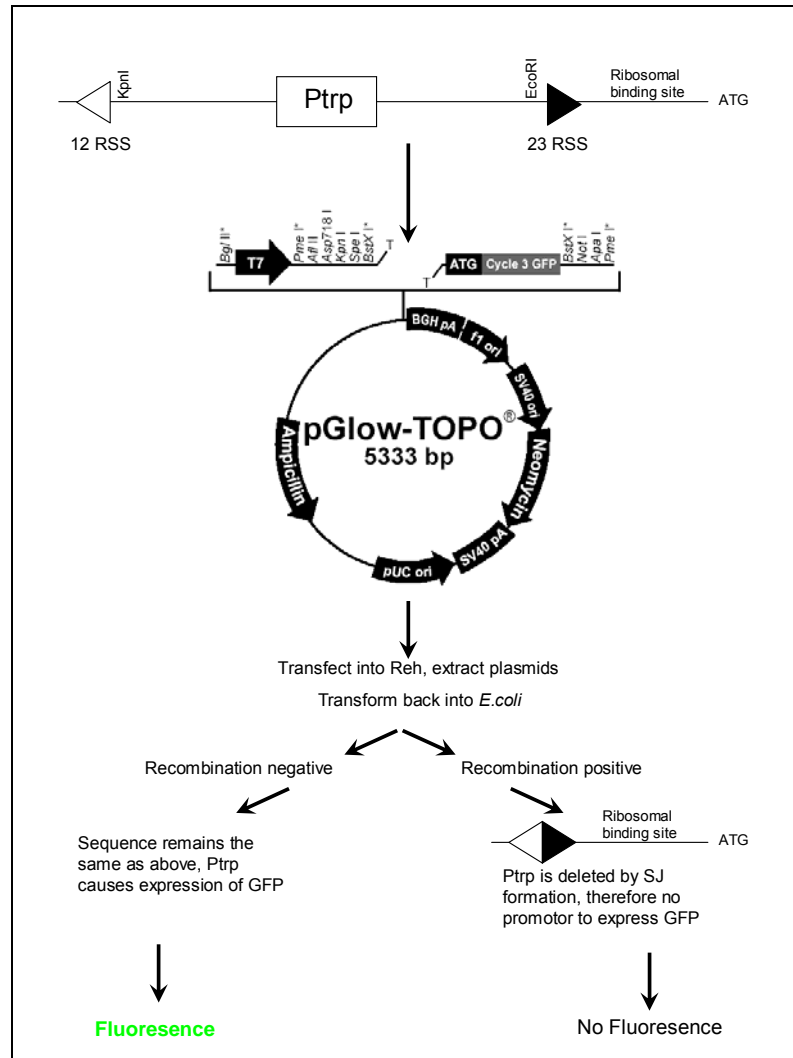


Figure 3-35: First strategy for determination of recombination using negative selection of the *GFP* gene. The tryptophan promoter denoted P_{trp}, flanked by consensus RSS sequences denoted by the open and dark triangles was cloned into the pGlow-TOPO vector. This should result in fluorescing *E. coli* colonies.

3.3.11.1.1 Cloning the P_{trp} RSS oligo into pGlow

Forward and reverse synthetic oligos of 140 bp long were designed to contain the tryptophan promoter (P_{trp}) flanked by the consensus RSS sequences as used by Raghavan *et al.*, 2001 (263) and were manufactured by Sigma (see Appendix II for sequence). In order to ligate the oligos into the pGlow vector, the P_{trp} RSS forward and reverse oligos

needed to be annealed so that they were double stranded therefore the protocol for annealing the bubble oligo was followed (see section 2.3.3). However, despite repeated attempts no recombinant clones were obtained so PCR primers were designed to enable PCR using the forward oligo as a template (see Figure 3-36 A). Even though the fragment was successfully cloned into the vector, as confirmed by restriction enzyme digest, it was difficult to ascertain if the *E.coli* colonies were glowing or not, due to the background fluorescence of the *E.coli* cells (see Figure 3-36 B).

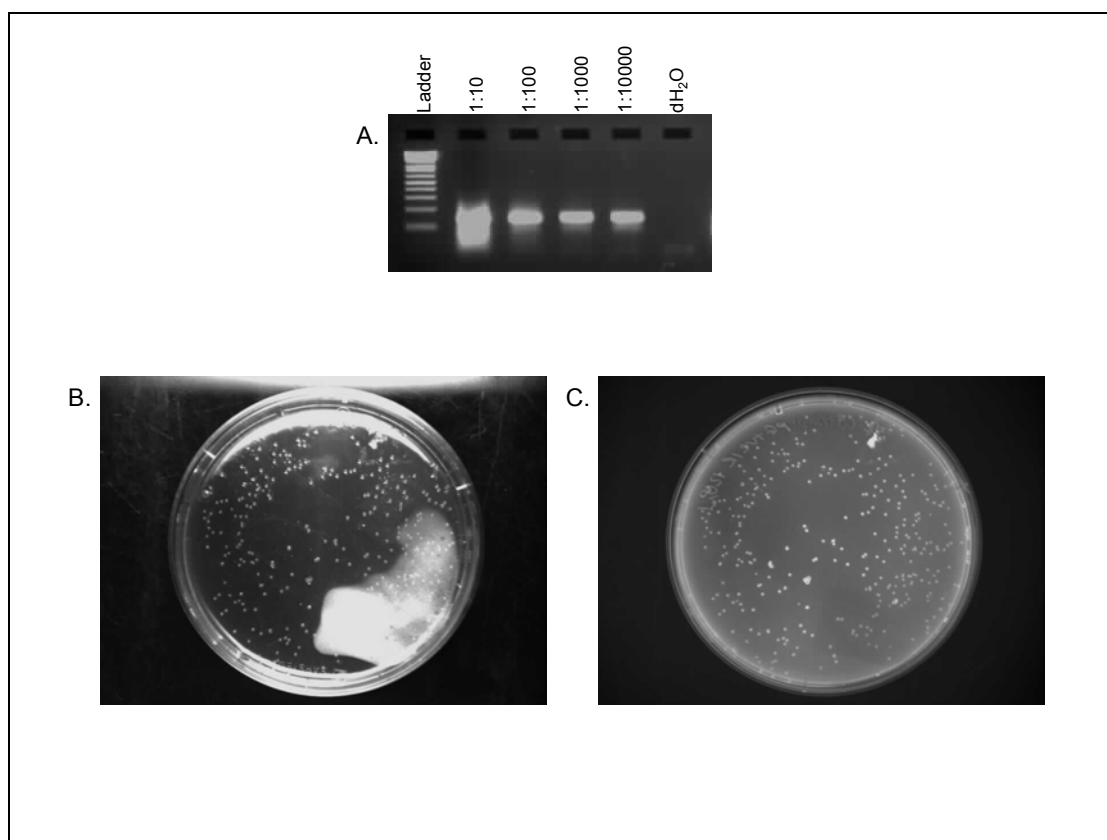


Figure 3-36: Cloning the *Ptrp* RSS oligo into pGlow. (A) Agarose gel of PCR amplified *Ptrp* RSS oligo. (B). *E.coli* transformed with PCR amplified *Ptrp* RSS oligo under white light and (C) the same plate under ultra violet light.

Another potential problem with this strategy was that it was based on negative selection whereby cells that were recombination positive would lose fluorescence. As recombination efficiencies are low when breakpoint regions are substituted for RSSs, an antibiotic selection strategy against the non-recombinants, whereby large numbers of transformed

E.coli cells can be plated out to find the few cells with double resistance, is likely to be preferable.

3.3.11.2 Measuring RAG recombination using pcDNA3.1/CT-GFP-TOPO vector

A second strategy, based on a positive selection and selection against the non-recombinants (as used in Raghavan *et al.*, 2001 (263)) was subsequently devised using the oop transcription stop sequence flanked by the RSS sequences (oop RSS) at the 5' end of the kanamycin (Kan) resistance gene. Any RAG recombination negative plasmids would retain the stop sequence and be resistant only to ampicillin, whereas recombination positive vectors will have the stop sequence excised and the colonies would be resistant to both ampicillin and kanamycin (see Figure 3-37).

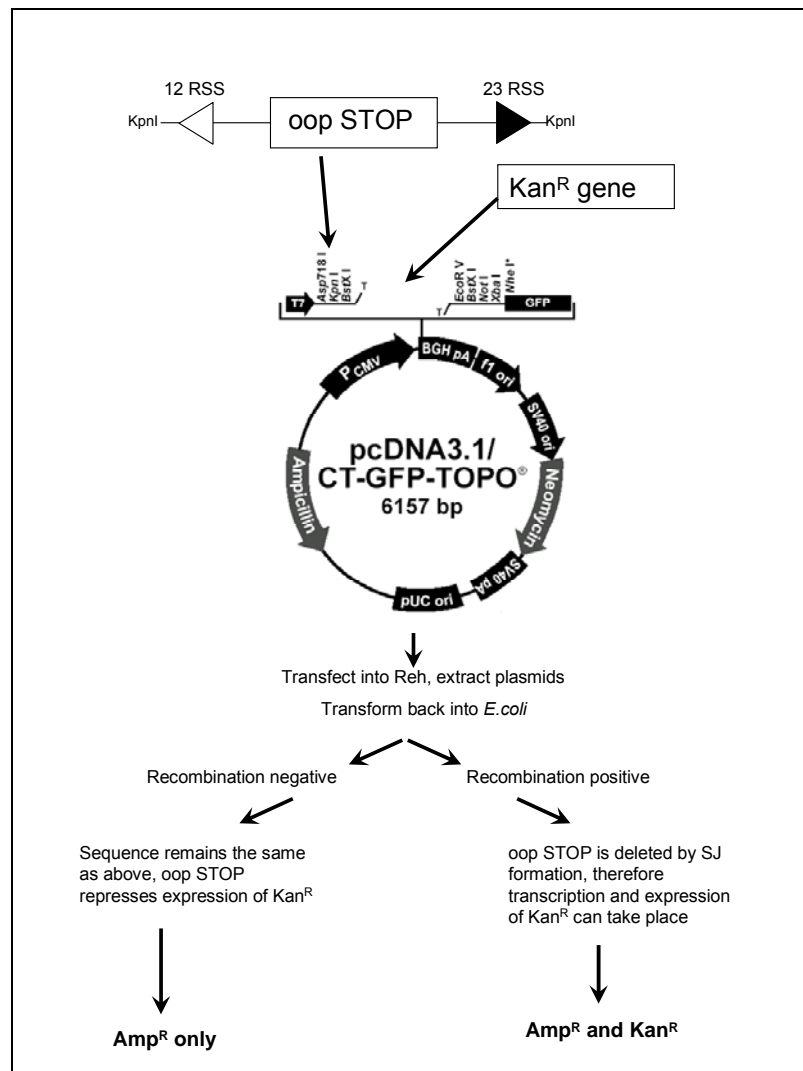


Figure 3-37: Second experimental strategy for determining illegitimate RAG recombination. The RSS flanked *oop* transcriptional terminator is cloned into the pcDNA3.1-TOPO vector between the T7 promoter and the *Kan* gene, therefore preventing transcription of the *Kan^R* gene and giving colonies upon ampicillin selection only. These vectors can then be transfected into the V(D)J positive Reh cell line for 48 hours, transformed back into *E. coli* and spread onto both ampicillin only and ampicillin plus kanamycin plates. Only recombination positive colonies will have both Amp^R and Kan^R. These colonies can then be sequenced.

3.3.11.2.1 Amplification and cloning of the *Kan^R* gene into pcDNA

To make the construct, the kanamycin resistance gene was amplified from the unligated XL-TOPO vector (Invitrogen) at dilutions of 1 in 50 and 1 in 5000, using High Fidelity PCR from the ORF to the end of the gene (from positions 1140 to 2076 of the XL-TOPO plasmid). This was cloned into the pcDNA3.1-TOPO vector (Invitrogen) (in a manner analogous to TOPO cloning in section 2.5) and upon transformation into competent *E. coli*

cells gave kanamycin resistant colonies. Six of these colonies were cultured and sequenced and all found to have the correct kanamycin insert. To ensure that the kanamycin resistance was a result of the cloned kanamycin gene and not inherent in the pcDNA3.1 plasmid, a negative control (i.e. without the kanamycin gene) was also performed and found to have no colonies on the kanamycin plates but many on the control ampicillin plate.

3.3.11.2.2 Amplification and cloning of the oop RSS fragment into pGlow TOPO

An oligo was designed and synthesised that encoded the bacteriophage lambda oop transcription stop site flanked by RSSs (see Appendix II for sequence). As there was a 250 bp limit on the size of the fragment that could be synthesised, the KpnI enzyme restriction sites necessary for cloning the oop RSS fragment into the pcDNA3.1 vector upstream of the Kan^R gene, were tagged onto the primers. Amplification of the fragment to make it double stranded for cloning was then performed using High Fidelity PCR (see Figure 3-38 A). The PCR product was then firstly cloned into the pGlow TOPO vector according to the protocol in section 2.5 (see Figure 3-38 B and C).

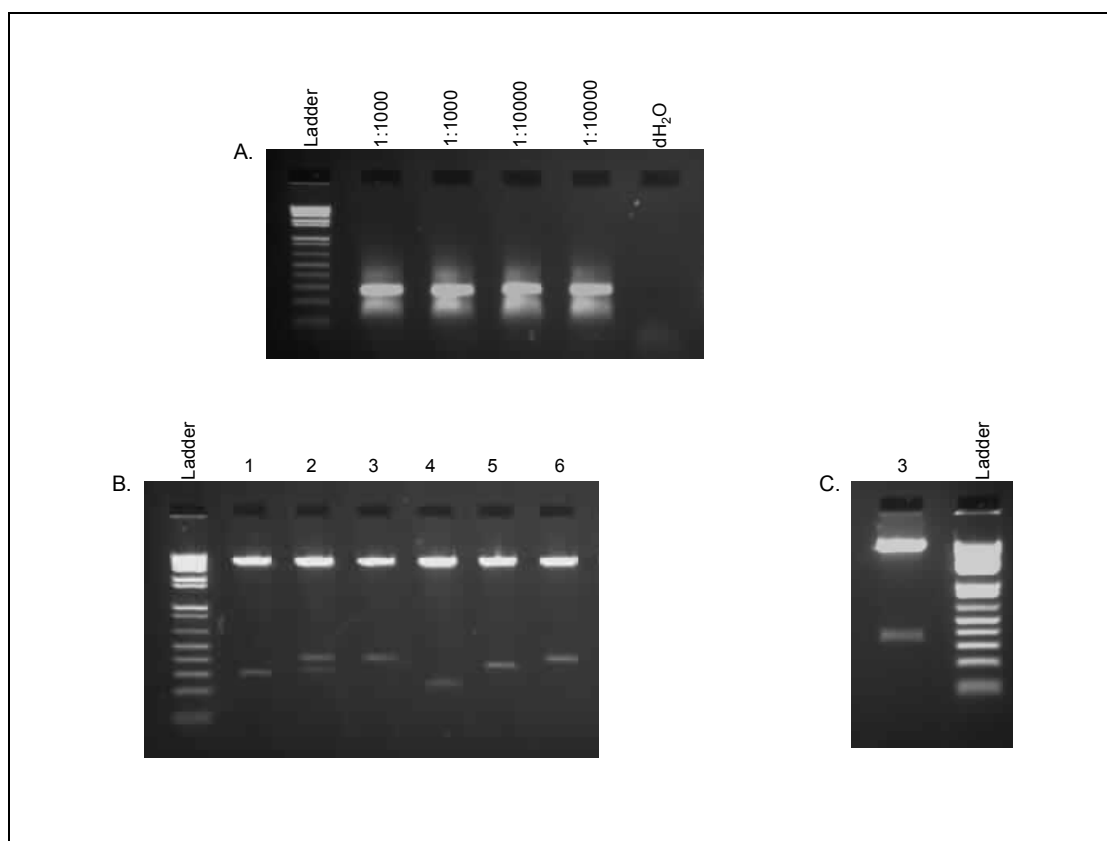


Figure 3-38: PCR and cloning of the RAG oop oligo. Agarose gel of the RAG oop oligo High Fidelity PCR product, at different dilutions of the original 100 μ M oligo (A). BglII digest of RAG oop ligated into the pGlow TOPO vector. The approximate size of the insert should be 438 bp and sequencing showed lane 3 to have the error free RAG oop insert (B). KpnI digest of product from lane 3 to confirm correct ligation and that the KpnI sites were intact for sub-cloning; the approximate size of the insert should be 330bp (C).

3.3.11.2.3 Sub-cloning oop RSS into pcDNA3.1 Kan^R

To sub-clone the oop RSS from the pGlow vector into the pcDNA Kan^R vector, 5 μ g of a sequenced pGlow oop RSS clone was digested with KpnI and the oop RSS fragment excised and purified from a crystal violet gel. This fragment was then ligated into a pcDNA Kan^R clone which had been KpnI digested, dephosphorylated using CIAP and crystal violet gel extracted. The two fragments were ligated at a 3:1 insert:vector ratio along with a vector only control, transformed into competent *E.coli* cells and plated out on ampicillin agar plates. As the vector only control showed a third less colonies than the ligated insert and vector, plasmid DNA was extracted from 20 colonies and digested with KpnI to check for the insert and 3 clones showed bands of the correct size. However, as the pcDNA3.1 and pGlow vectors are similar in size it was difficult to establish if the clones

were just carryover pGlow oop RSS clones and a second restriction digest was carried out with pcDNA3.1 Kan^R specific enzymes. Only clone 9 was pcDNA3.1 Kan^R with the oop RSS but upon sequencing it was found to be in the wrong orientation. Therefore plasmid DNA was extracted from 30 more colonies and digested with KpnI and a further 3 were found to have a correct sized insert, with again only one having the correct vector (see Figure 3-39 A and B) and upon sequencing this was found to be in the correct orientation.

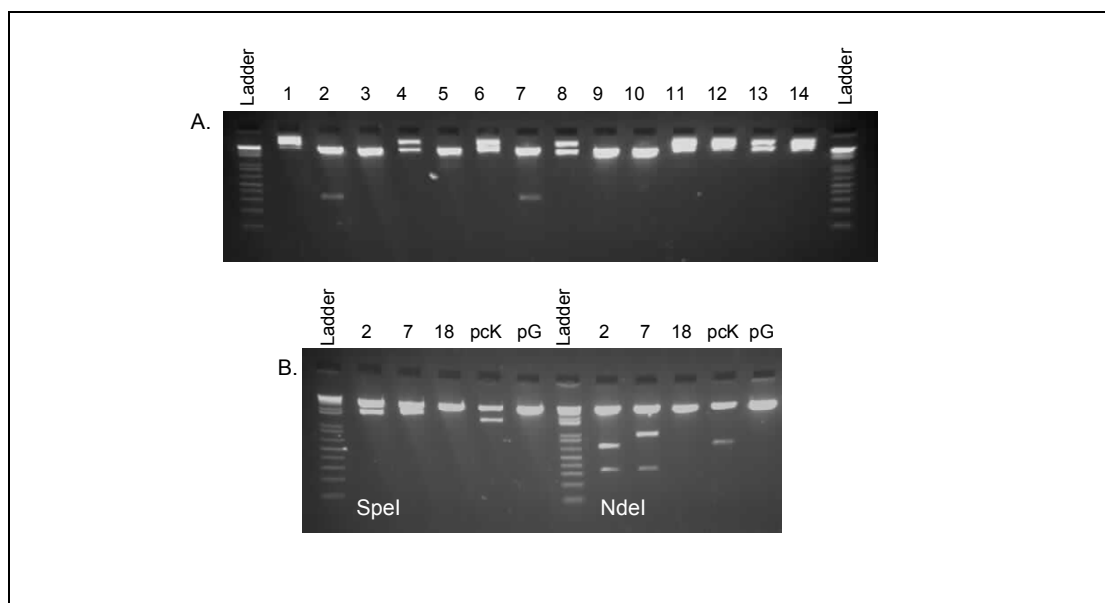


Figure 3-39: KpnI digest of the RAG oop oligo – pcDNA3.1 construct, the first 14 colonies: clones 2 and 7 show a band of the correct size for RAG oop ligation (A), as did clone 18 (not shown). To check that the RAG oop was in the correct vector, pcDNA3.1 CT TOPO with the Kanamycin gene (pcK), and not re-ligated into the original pGlow TOPO vector (pG), pcK specific digests, using SpeI and NdeI were performed, showing that clone 2 and clone 7 were both in the correct vector but clone 18 was not (B).

Before use in the cell line, it was necessary to check if the insert was able to stop transcription of the kanamycin gene, so competent *E.coli* cells were transformed with this clone and plated out on ampicillin only and kanamycin and ampicillin plates. If the insert worked correctly it would stop colonies growing on kanamycin plates, however, colonies were found on both plates and therefore the insert did not work correctly.

3.3.11.3 Measuring RAG recombination using pGlow-TOPO vector

The strategy in section 3.3.11.2 was thought not to work because of the strong CMV promoter in the pcDNA vector over-riding the oop stop sequence therefore the pGlow vector was used as the vector to be transformed into the Reh cells after addition of the correct DNA fragments (see Figure 3-40).

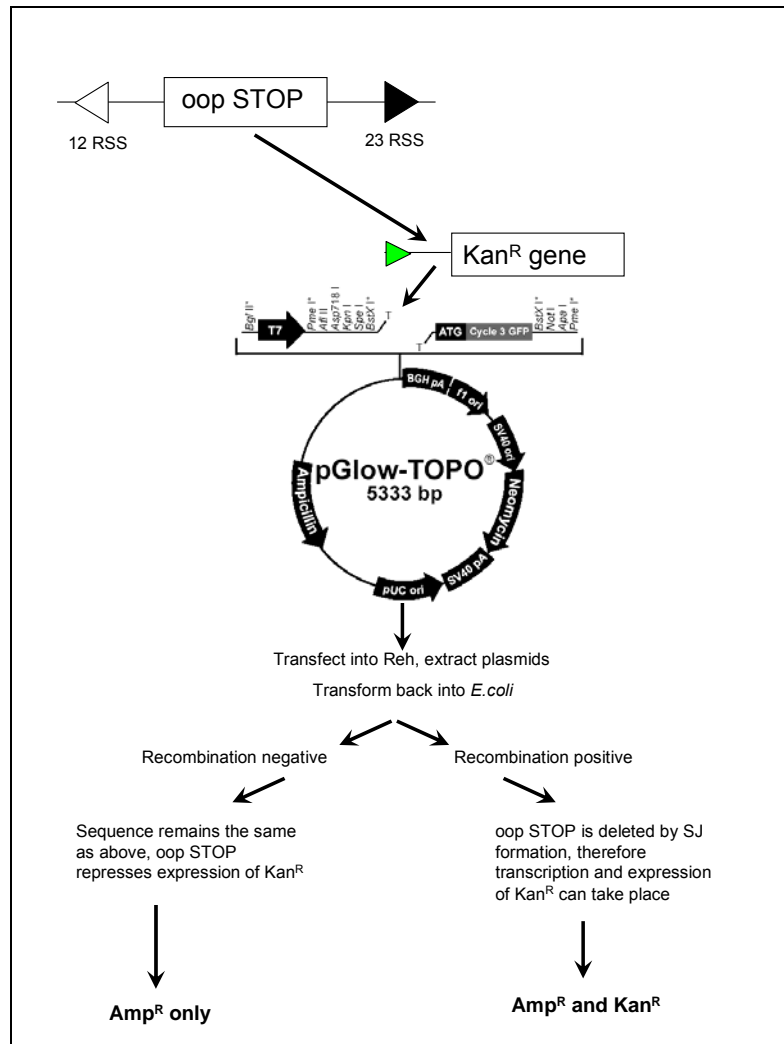


Figure 3-40: Third experimental strategy for determining illegitimate RAG recombination. The Kan^R gene is cloned into the pGlow-TOPO vector. This is then digested with XcmI restriction enzyme which specifically digests the plasmid between the Kan^R promoter and the Kan^R gene. The RSS flanked oop transcriptional terminator is then ligated in, preventing transcription of the Kan^R gene and giving colonies upon ampicillin selection only. These vectors can then be transfected into the V(D)J positive Reh cell line for 48 hours, transformed back into *E.coli* and spread onto both ampicillin only and ampicillin plus kanamycin plates. Only recombination positive colonies will have both Amp^R and Kan^R. These colonies can then be sequenced.

This strategy involved TOPO ligating the kanamycin gene with its own promoter into the pGlow vector. The unique XcmI site between the kanamycin promoter and gene was then utilised as the cloning site for the oop RSS fragment. However, as there was no XcmI site in the oop RSS fragment, blunt ended ligation was necessary. XcmI does not generate blunt ends so the vector, after XcmI digestion, was blunt ended by use of mung bean nuclease prior to dephosphorylation and crystal violet gel extraction of the vector. The oop RSS fragment was amplified using Pfu PCR which generates blunt ends, this was then crystal violet gel extracted and ligated into the prepared vector. This was tried several times unsuccessfully, by which time it became apparent that there was another group doing similar experiments on other lymphoid translocations from whom plasmids could be obtained.

3.3.12 Assessment of cryptic RSSs using the extra chromosomal recombination assay (plasmids provided by Professor Nadel (207))

This assay was originally designed to test the efficiency of V(D)J recombination between two authentic RSSs, which typically recombine at a high frequency, however when one of these authentic RSSs is replaced with a region containing putative cryptic RSSs, there is a decrease in the frequency of recombination and the appearance of other low level events such as; (a) non-specific recombination between the authentic RSS and a fortuitous RSS in either (i) the sequence flanking the test sequence or (ii) the plasmid core sequence or (b) break / repair (BR) recombination which is not mediated by V(D)J recombination and which makes up the bulk of the background in this assay (Marculescu *et al.*, (207) (see Figure 3-41).

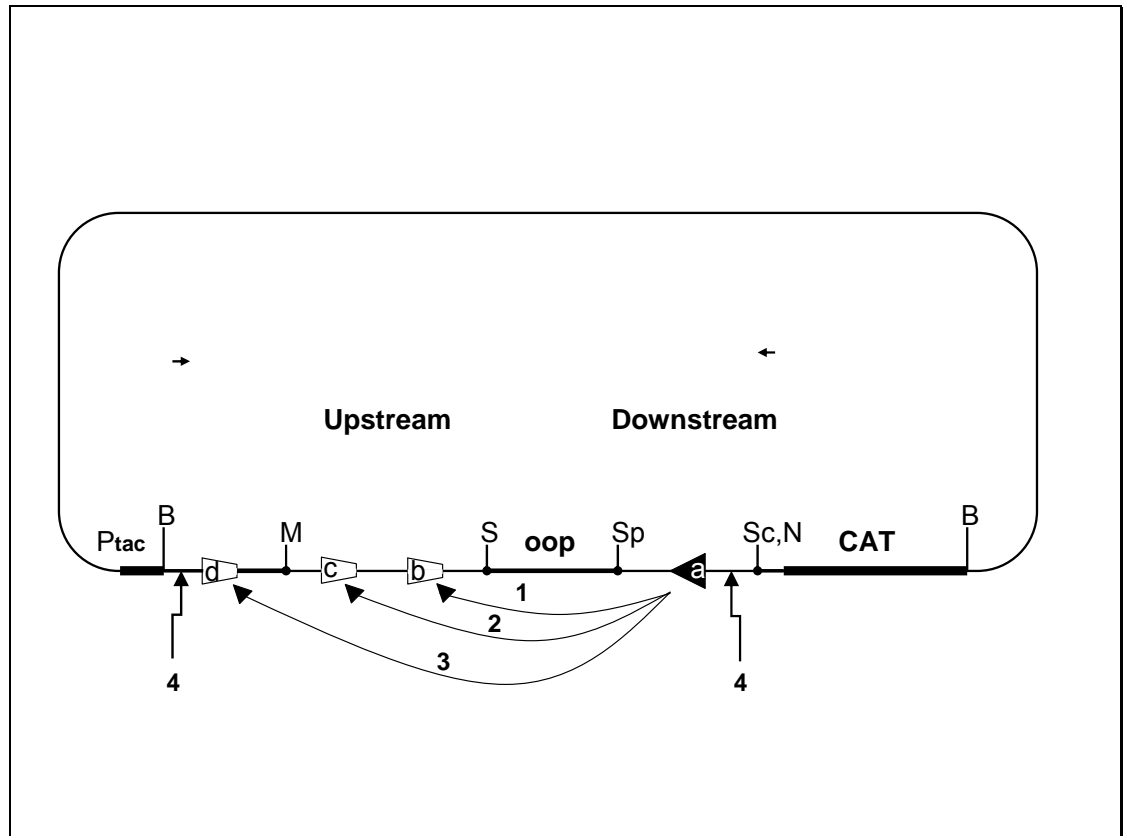


Figure 3-41: Recombination construct from Professor Nadel (not to scale). The upstream cassette is situated between the MluI (M) and Sall (S) sites. The downstream cassette is situated between the Spel (Sp) and SacII (Sc) or NotI (N) sites. The authentic RSS is shown as a triangle (a). Fortuitous RSSs are represented as trapezoids: b, fortuitous RSS observed at the *BCR* or *ABL* translocation breakpoints, c, other potential fortuitous RSS in the *BCR* or *ABL* junction fragments, d, other fortuitous RSS in core plasmid sequence. Pathways 1-3: V(D)J recombination; Pathway 4: break repair. The horizontal arrows indicate the location of the PCR screen primers. Modified from Marculescu *et al.*, (207).

In total, 5 test plasmids were received, 3 plasmids containing authentic RSS sites and test proto-oncogene sequences (RS7, RS8 and RS24) and 2 plasmids containing authentic/consensus RSS sites in both the upstream and downstream cassettes (RS39 and RS32) which were therefore suitable for use as positive controls to ensure the assay was working correctly (see Table 3-14). In addition, expression plasmids RAG1, RAG2 and TdT were received. All the plasmids were cloned into TOP10 cells, the plasmid extracted using the Endo-Free Maxi-prep kit (Qiagen) to ensure high quality DNA for transfections. The identity of all plasmids was confirmed by restriction enzyme digestion.

Name	RE	Upstream cassette			RE	oop STOP	RE	Downstream cassette			RE
RS7	(M)	Dβ1	220	◁▶ 160	(S)	STOP	(Sp)	TAL2	40	◁ 30	(N,Sc)
RS8	(M)	TAL1	190	▶ 190	(S)	STOP	(Sp)(Bg)	Dδ2	220	◁▶ 140	(N,Sc)
RS24	(M)	LMO2inv	50	▷ 50	(S)	STOP	(Sp)	Dδ2inv	140	◀▶ 220	(Bg,N,Sc)
RS39	(M)	D _H 3.10	40	▷ 320	(Bg)(S)	STOP	(Sp)	JH6m	190	◀ 200	(N,Sc)

Name	RE	Upstream cassette (1)			RE	Upstream cassette (2)			RE	oop STOP	RE	Downstream cassette			RE			
RS32	(M)	VκA2	90	▷	40	(N,Sc)	VκA27	100	▷	100	(S)	STOP	(Sp)	Jκ1	90	◀	25	(Sc)

Table 3-14: The locations of the upstream and downstream cassettes for use in the cell recombination assay. Open triangles represent 12 RSSs and closed triangles represent 23 RSSs.

D β 1 and D δ 2 were chosen to test the *BCR* and *ABL* junction sequences as although they did not have consensus RSSs (see Table 3-15), they are both authentic RSSs found *in vivo* that have been verified in the assay by Marculescu *et al.*, (207) to identify low efficiency cryptic proto-oncogene RSSs. Indeed, they also used D δ 2 to test the *BCL-2* major breakpoint cluster region to confirm that no functional cryptic RAG sites existed and, in addition, D δ 2 was also used to confirm that a cryptic RSS found by Lewis *et al.*, (264) had a recombination frequency of about 1% of a consensus RSS.

Heptamer and nonamer variations occur *in vivo* (265). Some have more effect than others, for example mutation of any of the first 3 bp of the heptamer (5' CAC) abolishes RSS activity, whereas substitutions in the nonamer have little effect as long as the A-rich core is maintained, as in D δ 2 (266). In addition, both D β 1 and D δ 2 contain a 12 and a 23 authentic RSS enabling any of the corresponding cryptic sequences in the rest of the *BCR* or *ABL* breakpoint sequences to be tested at the same time, and this also meant that fewer additional constructs needed to be made to test all putative combinations of recombination.

RSSs	Heptamer	Spacer	Nonamer
Consensus RSSs	CACAGTG		ACAAAAACC
Authentic RSSs			
D β 1	CACA <u>AT</u> G	23	ACAAAAACC
D β 1	CACA <u>AT</u> G ^a	12	ACAAAA <u>ACA</u>
D δ 2	CACAGTG	23	ACAAAA <u>ACT</u>
D δ 2	CACAGTG ^a	12	ACAAAA <u>ACT</u>
D _H 3.10	CACAGTG	12	ACAAAAACC
J _H 6 ^b	CACA <u>AT</u> G ^a	2 <u>2</u>	ACAAAAACC
V κ A2	CACAG <u>AG</u>	12	ACAAAAACC
V κ A27	CACAGTG	12	ACAAAAACC
J κ 1	CACAGTG ^a	23	ACAAAAACC

Table 3-15: Authentic RSSs used in plasmid constructs. Mismatches from the consensus sequence are underlined. a indicates sequence shown in reverse complement orientation. b indicates

In total, 11 p190 ALL, 11 p210 ALL and 12 p210 CML forward breakpoints had cryptic RSS sites as determined by the Fuzznuc screen, where the start of the forward or reverse cryptic heptamer fell within 10 bp of the breakpoint (see Table 3-16).

<u>Patient</u>	<u>Cryptic 12 RSSs</u>	<u>Distance breakpoint from RSS (bp)</u>	<u>Cryptic 23 RSSs</u>	<u>Distance breakpoint from RSS (bp)</u>
<u>p190</u>				
<u>Ital 1</u>	Reverse <i>BCR</i>	0, 2	Forward <i>BCR</i>	0
			Reverse <i>BCR</i>	2
<u>Ital 2</u>	Reverse <i>ABL</i>	2	Reverse <i>ABL</i>	1
<u>25654</u>	Reverse <i>BCR</i>	1	Reverse <i>BCR</i>	1
<u>Ital 6</u>	Forward <i>BCR</i>	1	Forward <i>BCR</i>	1
	Reverse <i>BCR</i>	4	Reverse <i>ABL</i>	6
			Reverse <i>BCR</i>	2, 4
<u>25169</u>	Forward <i>BCR</i>	0		
Ital 10	Reverse <i>ABL</i>	4	Reverse <i>ABL</i>	4
Ital 4	Reverse <i>ABL</i>	6	Reverse <i>ABL</i>	6
28835			Forward <i>ABL</i>	9
19635	Forward <i>ABL</i>	6		
27523	Reverse <i>BCR</i>	10		
P747			Forward <i>BCR</i>	10
<u>p210 ALL</u>				
<u>P851</u>	Forward <i>ABL</i>	1	Forward <i>ABL</i>	1
<u>F17730</u>	Forward <i>BCR</i>	10	Forward <i>BCR</i>	0
<u>F24559</u>			Forward <i>BCR</i>	1
22251	Reverse <i>BCR</i>	4	Forward <i>ABL</i>	7
11129			Forward <i>ABL</i>	7
6875	Reverse <i>ABL</i>	1		
F22666			Forward <i>ABL</i>	5
F17685	Reverse <i>BCR</i>	5		
F16604	Forward <i>BCR</i>	6	Forward <i>BCR</i>	6
F14472			Reverse <i>ABL</i>	2
F20099	Forward <i>BCR</i>	7		
<u>p210 CML</u>				
<u>266</u>	Forward <i>ABL</i>	8		
<u>386</u>	Forward <i>BCR</i>	1	Forward <i>BCR</i>	1

			Forward <i>ABL</i>	2, 6
NP241	Reverse <i>BCR</i>	10	Reverse <i>ABL</i>	10
	Reverse <i>ABL</i>	10		
1538			Reverse <i>BCR</i>	7, 9
1578			Reverse <i>BCR</i>	4
			Forward <i>ABL</i>	9
1349	Reverse <i>ABL</i>	3	Forward <i>BCR</i>	5
			Reverse <i>BCR</i>	4
316	Forward <i>BCR</i>	2	Forward <i>BCR</i>	2
440	Forward <i>BCR</i>	4, 8		
446	Forward <i>ABL</i>	5		
545	Reverse <i>ABL</i>	10	Reverse <i>ABL</i>	10
282	Reverse <i>ABL</i>	6	Reverse <i>ABL</i>	6
291			Forward <i>BCR</i>	8

Table 3-16: Results of the EMBOSS Fuzznuc screen on all 82 cases. Only the cases shown had cryptic RSS sites that were within the specified parameters (see section 3.2.7) that were 10 bp or closer to the breakpoints. Those in bold and underlined were chosen to use in the recombination assay.

Due to the complexity of the protocol for testing breakpoints for cryptic RSSs and the limited time due to the delay in receiving the plasmids, 5 p190 ALL, 3 p210 ALL and 2 p210 CML breakpoints were selected for functional analysis based on the following criteria; proximity of cryptic RSS site to breakpoint, to have a mixture of both *ABL* and *BCR* breakpoints and to have a mixture of 12 and 23 cryptic RSSs (see Table 3-17).

RSSs	Heptamer	Spacer	Nonamer	Distance from breakpoint (bp)
<u>Consensus RSSs</u>	CACAGTG	12/23	ACAAAAACC	
<u>Fortuitous RSSs</u>				
<u>p190 ALLs</u>				
Ital 1 <i>BCR</i>	CAC <u>I</u> G <u>T</u> T ^a	12	ACAC <u>G</u> AG <u>T</u> C	2
Ital 1 <i>BCR</i>	CAC <u>A</u> CTG ^a	12	<u>C</u> C <u>A</u> C <u>A</u> CGAG	0
Ital 1 <i>BCR</i>	CAC <u>I</u> G <u>T</u> T ^a	23	ACAG <u>C</u> TGAA	2
Ital 1 <i>BCR</i>	CAC <u>A</u> CTG	23	<u>C</u> C <u>A</u> CAGCTG	0
Ital 2 <i>ABL</i>	CACAG <u>C</u> A ^a	12	<u>A</u> A <u>A</u> CACCTC	2
Ital 2 <i>ABL</i>	CACAG <u>C</u> A ^a	23	<u>G</u> CTGTAAAA	1
25654 <i>BCR</i>	CACAG <u>A</u> G ^a	12	<u>C</u> <u>C</u> <u>C</u> <u>C</u> CACCC	1
25654 <i>BCR</i>	CACAG <u>A</u> G ^a	23	<u>T</u> CATGGCTG	1
Ital 6 <i>BCR</i>	CACG <u>G</u> T <u>T</u>	12	<u>C</u> CATGTCTA	1
Ital 6 <i>BCR</i>	CACAG <u>C</u> C ^a	12	<u>T</u> GGCAGAA <u>C</u>	4
Ital 6 <i>BCR</i>	CACG <u>G</u> T <u>T</u>	23	AGAGCATGC	1
Ital 6 <i>BCR</i>	CAC <u>A</u> CAG ^a	23	AG <u>C</u> <u>C</u> <u>C</u> <u>C</u> CTCC	2
Ital 6 <i>BCR</i>	CACAG <u>C</u> C ^a	23	<u>C</u> <u>C</u> <u>C</u> <u>C</u> CT <u>C</u> <u>C</u> <u>C</u> <u>C</u>	4
25169 <i>BCR</i>	CAC <u>C</u> TCA	12	<u>G</u> CTGAAA <u>A</u> C	0
<u>p210 ALLs</u>				
P851 <i>ABL</i>	CACAG <u>A</u> G	12	ACATTATAC	0
P851 <i>ABL</i>	CACAG <u>A</u> G	23	<u>T</u> TAGGATT <u>T</u>	0
F17730 <i>BCR</i>	CAC <u>I</u> GG <u>T</u>	12	<u>T</u> GAAACCAG	10
F17730 <i>BCR</i>	CAC <u>T</u> TCG	23	<u>G</u> AAACCAG <u>C</u>	0
F24559 <i>BCR</i>	CAC <u>T</u> TCG	23	<u>G</u> AAACCAG <u>C</u>	1
<u>p210 CMLs</u>				
266 <i>ABL</i>	CAC <u>C</u> TTG	12	<u>G</u> CTGGGATT	8
386 <i>BCR</i>	CAC <u>T</u> TTG	12	AGGTGGAT <u>C</u>	1
386 <i>BCR</i>	CAC <u>T</u> TTG	23	<u>T</u> TGAGCTCA	1

Table 3-17: Cryptic RSS sites at the vicinity of the *BCR* or *ABL* breakpoint for cases selected to be tested. Mismatches from the consensus sequence are underlined, ^a indicates that the reverse complement orientation is shown.

3.3.12.2 Additional constructs required for testing *BCR-ABL* breakpoint cryptic RSSs

As each cryptic RSS found in each breakpoint could either be used as a signal joint or a coding joint in a RAG recombination event and it is the formation of both of these that defines V(D)J recombination, it was necessary to test the selected breakpoints for all the putative combinations of cryptic RSSs (see Figure 3-42) by using the authentic RSSs Dδ2

or D β 1. As these authentic RSSs contained both 12 and 23 RSSs and D δ 2 was already supplied with the downstream cassette in both forward and inverse (RS8 and RS24 respectively) orientations, it was necessary to make only two additional constructs (D δ 2 forward and inverse in the upstream cassette) giving a total of 5 test constructs for each breakpoint junction to be examined (see Table 3-18).





Name	RE	Upstream cassette			RE	oop STOP	RE	Downstream cassette			RE
RS3	(M)	D δ 2	220	 	140	(S)	STOP	(Sp)	BCR/ABL	 	(N,Sc)
RS2	(M)	D δ 2 _{inv}	140	 	220	(S)	STOP	(Sp)	BCR/ABL	 	(N,Sc)
RS7	(M)	D β 1	220	 	160	(S)	STOP	(Sp)	BCR/ABL	 	(N,Sc)
RS8	(M)	BCR/ABL		 		(S)	STOP	(Sp)(Bg)	D δ 2	220  	140 (N,Sc)
RS24	(M)	BCR/ABL		 		(S)	STOP	(Sp)	D δ 2 _{inv}	140  	220 (Bg,N,Sc)

Table 3-18: The locations of the upstream and downstream cassettes for use in the cell recombination assay. RS7, RS8 and RS24 were received from Professor Nadel, whilst RS3 and RS2 were made by inserting the D δ 2 and D δ 2_{inv} cassettes into the upstream cassette in order to test every combination of RSS.

In order to make the two new test constructs RS2 (D δ 2 inverse in the upstream cassette) and RS3 (D δ 2 in the upstream cassette), primers were designed to be as close to the original primers used by Marculescu *et al.*, (207) as possible and tagged with either MluI or SalI in order to be able to be cloned into the upstream cassette (see Appendix II for primer sequences). Due to possible sequence variants in normal control DNA, dilutions of each of the appropriate plasmids (RS24 and RS8) were used as templates for amplification with High Fidelity PCR. The PCR products (see Figure 3-43 A) were then cloned into the TOPO pCR4 vector and the plasmid DNA extracted. Each clone was digested with EcoRI to ensure the correct sized insert (see Figure 3-43 B). Once each clone was sequenced to ensure no PCR errors had been introduced, 5 μ g was digested with MluI and SalI, crystal violet extracted, cleaned and quantified with the NanoDrop.

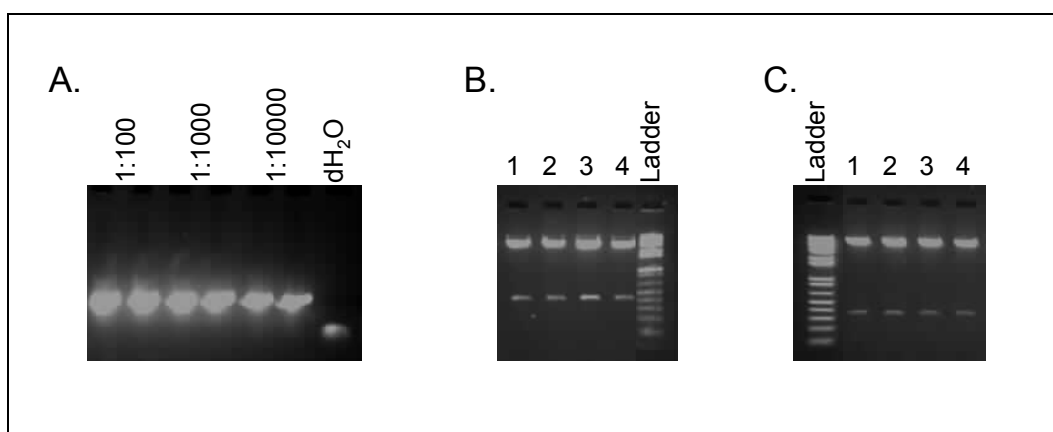


Figure 3-43: Agarose gels showing the process of making construct RS2. (A) PCR of D2inv from various dilutions of RS24. (B) EcoRI digest of D δ 2inv from pCR4 TOPO of clones 1-4. C. MluI and SalI digest of D2inv from the core plasmid.

Concurrently, the parent vector, RS24, was prepared for the inserts by removal of the LMO2inv upstream cassette by MluI and SalI digestion (see Table 3-18). The RS2 (D δ 2inv) and RS3 (D δ 2) inserts were then ligated into the parent vector and DNA from individual colonies sequenced with the core plasmid specific Rag F and Rag R primers to ensure correct insertion into the parent plasmid (see Figure 3-44). For example, the RS2 Rag F sequence was compared to both the core plasmid sequence using NCBI BLAST 2 sequences (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>) (Figure 3-44 A) and the RS24 (D δ 2inv) Rag R reverse complement sequence using UCSC BLAT (<http://genome.ucsc.edu/cgi-bin/hgBlat>) (see Figure 3-44 B and C). Finally, RS2 Rag R

sequence was compared to RS24 Rag R sequence (see Figure 3-44 D) to ensure insertion of correct D δ 2inv sequence in the correct place.

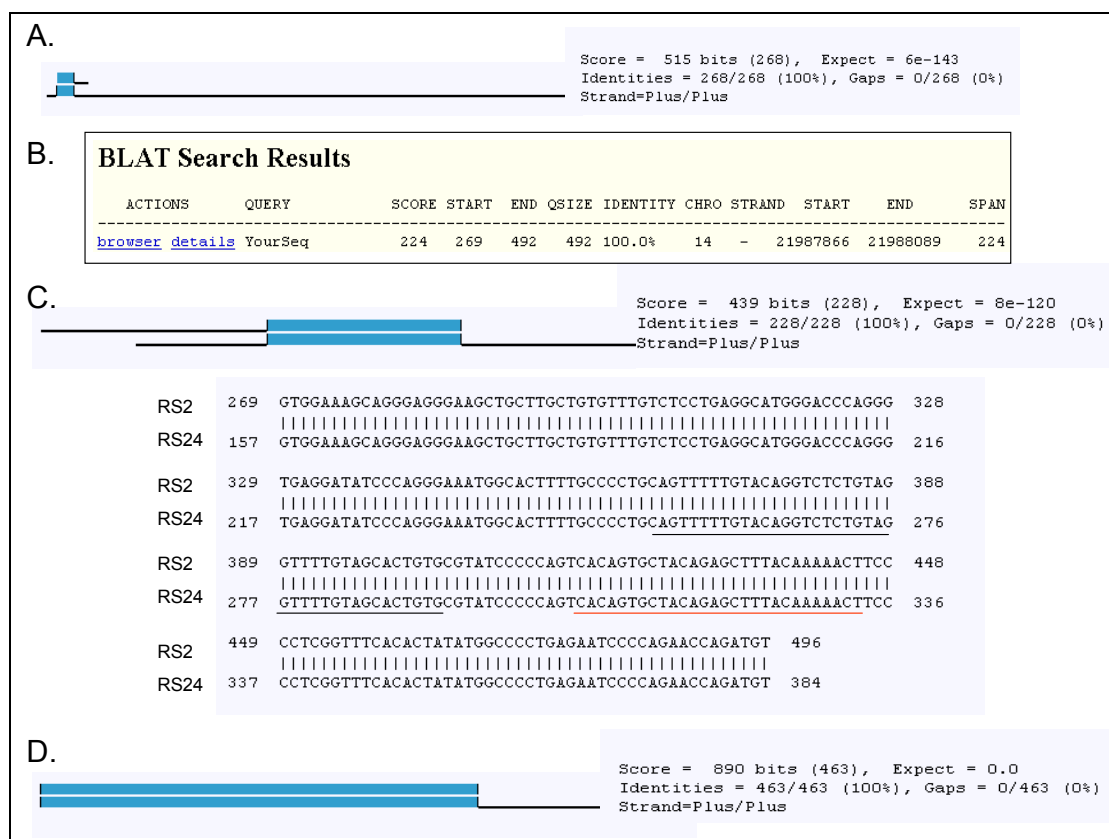


Figure 3-44: Sequencing of RS2. (A) Comparison of RS2 Rag F sequencing with the core plasmid sequence using NCBI 'compare 2 sequences', showing that the first 268 bp of RS2 (top line) are identical to the plasmid core sequence (bottom line). (B) UCSC BLAT results from sequencing of RS2 with Rag F, showing insertion of D δ 2inv (located on chromosome 14) adjacent to the plasmid core sequence starting at 269 bp. (C) Comparison of RS2 Rag F sequencing with RS24 Rag R sequencing (reverse complemented), showing identical sequence. Underlined in black is the inverse 23 RSS and underlined in red is the forward 12 RSS (see Table 3-18). (D) Comparison of RS2 and RS24 Rag R sequencing.

Once RS2 and RS3 were confirmed to be correct by digests and sequencing, the downstream cassettes were excised using a SpeI and SacII double digest. The downstream cassette for RS7 and the upstream cassettes for RS8 and RS24 were also excised ready for ligations to breakpoint fragments.

3.3.12.3 Making constructs for testing patient breakpoints.

As each RSS can potentially be used as either a coding joint or a signal joint, there are 2 constructs required for each type of RSS (12, 23, forward or reverse) and, as many of the cases chosen had more than one type of RSS at or near the breakpoint there were 31 constructs that needed to be made to test all permutations of the selected 10 breakpoints (see Figure 3-45).

Vector	RSS	Upstream cassette	Downstream cassette	Tests for:
<u>p190 ALL's</u>				
Ital 1 RS8	D δ 2			R 12 SJ
Ital 1 RS3	D δ 2			R 12 CJ
Ital 1RS24	D δ 2inv			R 23 SJ
Ital 2 RS8	D δ 2			R 12 SJ
Ital 2 RS3	D δ 2			R 12 CJ
Ital 2 RS24	D δ 2inv			R 23 SJ
Ital 2 RS2	D δ 2inv			R 23 CJ
25654 RS8	D δ 2			R 12 SJ
25654 RS3	D δ 2			R 12 CJ
25654 RS24	D δ 2inv			R 23 SJ
25654 RS2	D δ 2inv			R 12 CJ
Ital 6 RS2	D δ 2inv			F 12 SJ
Ital 6 RS24	D δ 2inv			R 23 CJ
				R 23 SJ
				F 12 CJ
Ital 6 RS8	D δ 2			F 23 CJ
Ital 6 RS7	D β 1			R 12 SJ
				R 12 CJ
25169 RS2	D δ 2inv			F 23 SJ
25169 RS24	D δ 2inv			F 12 SJ
<u>p210 ALL's</u>				
P851 RS2	D δ 2inv			F 12 SJ
P851 RS24	D δ 2inv			F 12 CJ
P851 RS7	D β 1			F 23 SJ
P851 RS8	D δ 2			F 23 CJ
F17730 RS7	D β 1			F 23 SJ
F17730 RS8	D δ 2			F 23 CJ
F24559 RS7	D β 1			F 23 SJ
F24559 RS8	D δ 2			F 23 CJ

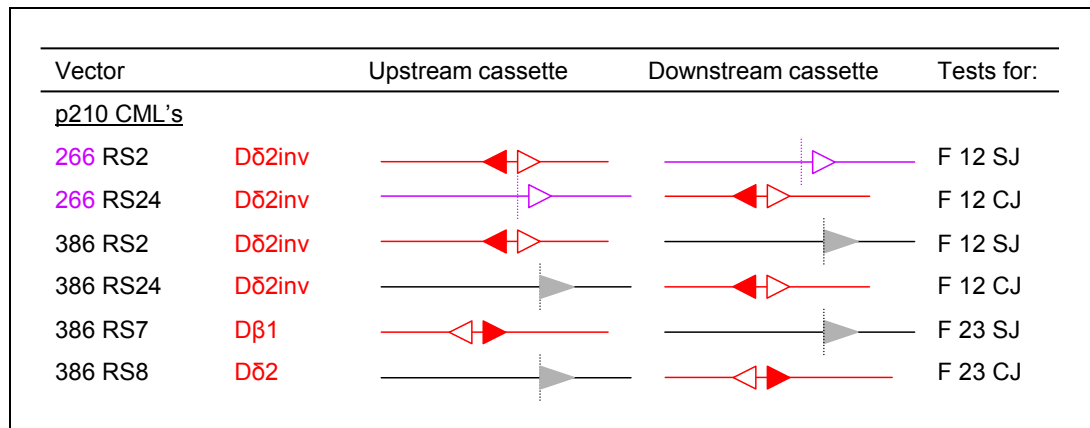


Figure 3-45: Constructs required to test RSS activity in the 10 cases. Authentic RSS sites (in red) and patient breakpoint flanking cryptic RSS sites in black (*BCR* cassettes) and purple (*ABL* cassettes) patient breakpoint locations are denoted by the vertical dotted lines (to scale, except 386 breakpoint region which should be ~7cm). RSSs with 12 bp spacers are open triangles and 23 bp spacers are solid black (*BCR*) and red (authentic RSSs) triangles; those that can be either 12 or 23 bp RSSs are solid grey (*BCR*) and solid purple (*ABL*) triangles. For clarity, cryptic RSSs in the patient sequences are not shown if over 10bp from the breakpoint.

To make each construct, each *BCR* or *ABL* breakpoint region was amplified from the patient DNA from which it was originally derived (to ensure that there was no exclusion of SNPs particular to each patient which might effect the results) using High Fidelity PCR (see Figure 3-46 A). Tagged onto the 5' end of the primers was the appropriate enzyme sequence to allow cloning into the parent 'core' plasmid. In addition, the minimum region flanking the breakpoint was approximately 100 bp either side with a maximum of approximately 300bp either side (see primer lists in Appendix II) depending on where primers could be designed. The fragments were then cloned into TOPO pCR4, the resultant extracted plasmid clones were digested with MluI/SalI or SpeI/SacII as appropriate (see Figure 3-46 B) and sequenced with M13 F and M13 R primers. Once the inserts were confirmed to be mutation free compared to the PCR amplified and originally amplified breakpoint sequence, the insert was ligated into the appropriate parent vector (RS2, RS3, RS7, RS8 or RS24 see Figure 3-45). All final test constructs were checked by digests (see Figure 46 C) and sequencing both the upstream and the downstream cassettes (see Figure 46 D and E).

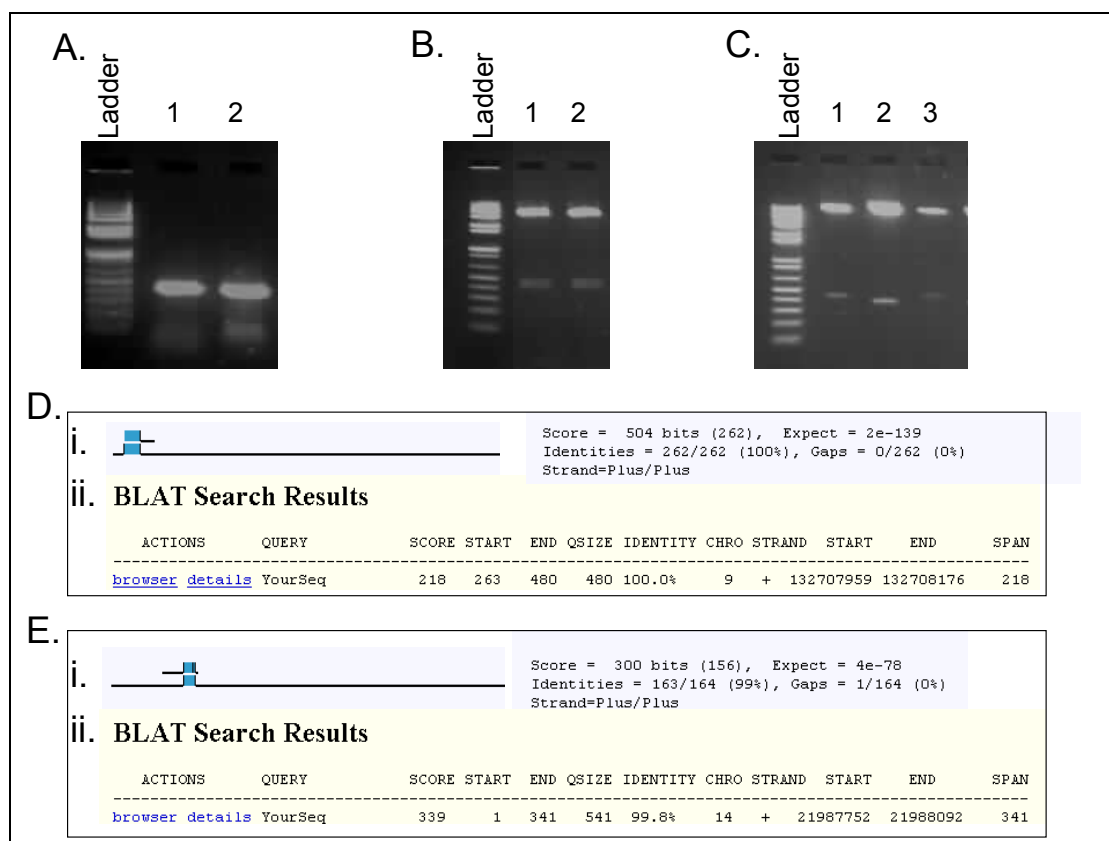


Figure 3-46: Construction and verification of construct P851 RS8. (A) High fidelity PCR of the P851 *ABL* breakpoint region in duplicate using tagged *Mlu*I forward and *Sall* reverse primers. (B) *Mlu*I and *Sall* digest of the *ABL* fragment from pCR4 TOPO vector, clones 1 and 2. (C) *Mlu*I and *Sall* digest of the *ABL* fragment from the core plasmid, clones 1 and 3 have the correct *ABL* insert, clone 2 has retained the original upstream RS8 cassette. (D) Rag F sequencing of clone 1. (i) Comparison of the P851 RS8 vector to the plasmid core sequence showing identical sequence for the first 262 bp. (ii) BLAT results showing chromosome 9 *ABL* sequence from 263 bp onwards. (E) Rag R sequencing of clone 1. (i) Comparison of P851 *ABL* vector to the plasmid core sequence. (ii) BLAT results showing chromosome 14 Dδ2 sequence.

3.3.12.4 Optimising FuGene transfections

As the protocol received from Professor Nadel did not detail the amount of FuGene transfection reagent to use in the assay and as this is likely to vary between NIH-3T3 cell lines, it was necessary to establish the optimal volume of FuGene. To estimate transfection efficiency, different amounts of FuGene reagent were used with pEFLacZ (see section 3.2.9.3). The volumes of FuGene found to give the highest transfection efficiency of between 5-10% were 6, 7 and 8 μ l.

To ensure that the cell culture assay was working correctly and to further optimise the volume of FuGene, the positive control RS39 plasmid was transfected into the NIH-3T3 cells using 30, 35 and 40 μ l of FuGene (these volumes represent the same ratios as with the pEFLacZ but scaled up for use in the cell culture assay which requires a greater number of starting NIH-3T3 cells and 10 μ g of plasmid DNA in total). In addition, a negative control with all the plasmids but no FuGene was also included. The resultant transfected TOP10 cells were plated on both ampicillin only and chloramphenicol plus ampicillin plates. The negative control worked well with only approximately 20 colonies on each plate. The positive control plates were confluent with colonies on the ampicillin only plates indicating that the transfections had worked and that the plasmids had propagated in the NIH-3T3 cells. The remaining cells that were plated on the chloramphenicol plus ampicillin plates showed that 40 μ l of FuGene gave the most recombined colonies at an average of 582 (compared to an average of 250 per plate from Marculescu *et al.*, (207) using electroporation) and therefore this amount was used for all further transfections. The number of colonies on ampicillin plus chloramphenicol plates for plasmids RS8, RS32 and RS39 was also greater than that reported by Marculescu *et al.*, (207) (see Table 3-19), presumably due to the improved transfection efficiency with FuGene.

Name	Colonies A	Colonies AC	Colonies ACM
RS8	TNTC	97	~18
RS32	TNTC	~6000	~3700
RS39	TNTC	560	~250
No FuGene	24	1	-

Table 3-19: Comparison of assay results with those published by Marculescu *et al.*, (207). The number of colonies were counted on each plate; A = ampicillin, AC = ampicillin and chloramphenicol and ACM = ampicillin and chloramphenicol. TNTC = too numerous to count.

Although the numbers of colonies on each plate compared to the negative (no FuGene) control were a good indication that the assay had worked, to definitively prove this was the case 5 colonies from each plasmid were amplified by PCR and sequenced using Rag F and Rag R primers (see Figure 3-47 A). In concordance with Marculescu *et al.*, (207), V(D)J recombination was seen all of the RS32 and RS39 colonies, indicated by the formation of

coding joints (see Figure 3-47 B) and also showing features of RAG recombination by insertion of non-templated nucleotides (see Figure 3-47 C to E). However, no V(D)J recombination was seen in the 5 colonies for RS8, probably because this vector only has a 5% recombination efficiency and 5 colonies were not enough to see this. However as the other 2 positives gave 100% recombination efficiency with high colony numbers, and upon discussion with Professor Nadel, this was considered to be a high enough sensitivity to proceed with patient constructs.

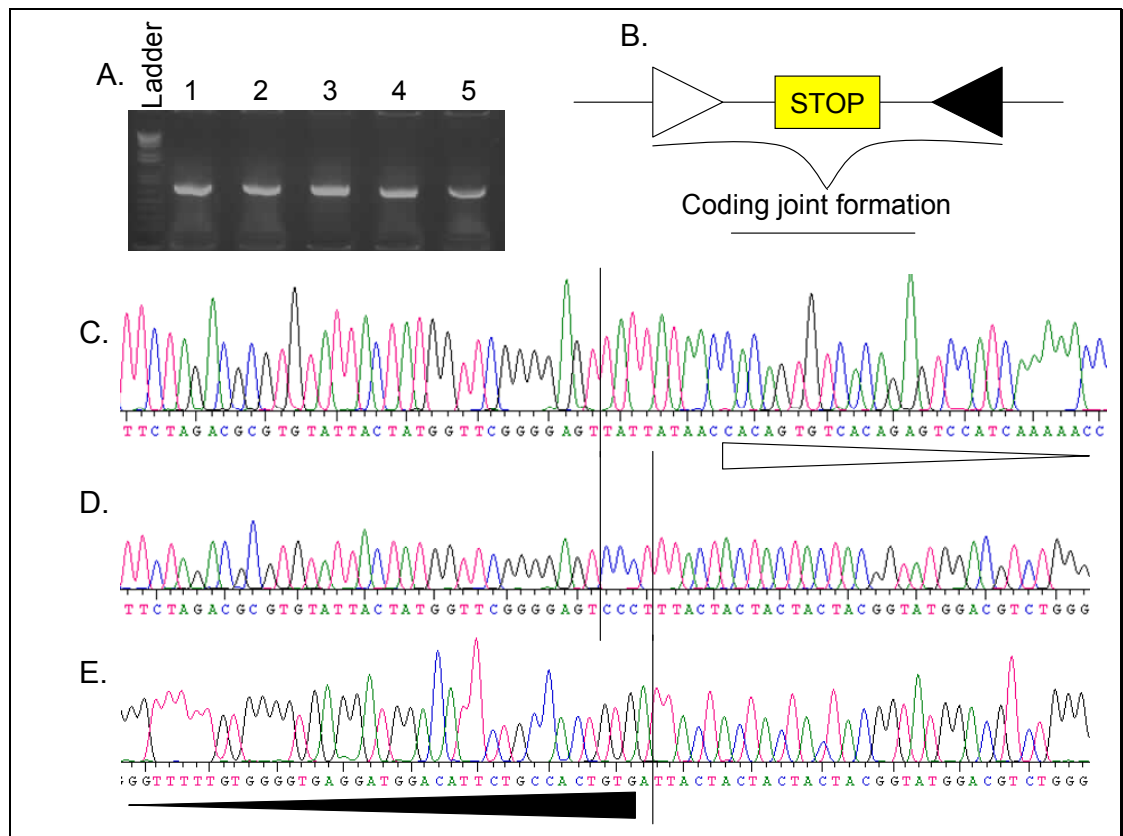


Figure 3-47: Analysis of the RS39 positive control. (A) Agarose gel of five RS39 colonies amplified by PCR from the cell culture assay. (B) Schematic showing coding joint formation if V(D)J recombination occurs. (C) and (E) Electropherogram of un-recombined RS39 plasmid sequences showing the authentic 12 RSS (open triangle beneath the sequence) and 23 RSS (solid triangle below the sequence). (D) Electropherogram of RS39 colony 1 (from gel A) showing V(D)J recombination, the vertical lines show where the breaks occurred and also flank the inserted of non-templated nucleotides which are a feature of V(D)J coding joints.

3.3.12.5 Cell culture assay with *BCR* / *ABL* test constructs.

Once the correct clones were identified through sequencing, clones were re-grown from glycerol stocks and the plasmid DNA extracted using the PureLink HQ mini-prep kit (Invitrogen) and only 2.5 µg was required for each transfection. The quality of this DNA was initially tested by extracting the positive control RS39 plasmid and comparing it to the original RS39 in the cell culture assay to ensure that it did not affect transfection efficiency. The numbers of ampicillin and chloramphenicol colonies produced at the end were similar between the two methods of plasmid extraction and therefore this technique was used subsequently for extraction of the test plasmids for which less DNA was required. Once the test constructs were re-grown and extracted, they were re-sequenced with Rag F and Rag R to ensure that they were correct.

All *BCR* / *ABL* test constructs were used in the cell culture assay at least once, except 3 constructs; Ital 1 RS3, 25169 RS2 and 266 RS2 (see Figure 3-45) which failed to ligate into the core plasmids and could not be repeated due to time constraints. Each set of transfections was run with a positive control to ensure the assay was working with a high enough sensitivity and a negative (no FuGene) control. Again as there was only a limited time, only a small number of transfections could be performed and thus the colony numbers were somewhat variable (see Table 3-20). Importantly, all positive controls worked well with colony numbers routinely double those reported by Marculescu *et al.*, (207). PCR and sequencing revealed that most of the recombination events between the vectors containing two authentic RSSs (RS32 and RS39) were indeed mediated by RAG (85-100% see Table 3-20).

Vector	No. T ^a	No. clones ^b	No. A ^c	Specific RSS ^d	Other RSS ^e	BR ^f
Controls						
RS8	3	360 (120)	5	0	0	5
RS39	4	1996 (499)	40	34 (85%)	1 (2.5%)	5 (12.5%)
RS32	3	19500 (6500)	28	28 (100%)	0 (0%)	0 (0%)
<u>p190 ALLs</u>						
Ital 1 RS8	3	194 (65)	43	0 (0%)	16 (37%)	27 (63%)
Ital 1 RS3	<i>Failed to ligate insert into parent vector RS3.</i>					
Ital 1 RS24	2	135 (68)	42	0 (0%)	5 (12%)	37 (88%)
Ital 2 RS8	1	15	14	0 (0%)	0 (0%)	14 (100%)
Ital 2 RS3	1	175	24	0 (0%)	12 (50%)	12 (50%)
Ital 2 RS24	2	124 (62)	35	0 (0%)	10 (28%)	26 (72%)
Ital 2 RS2	3	190 (63)	34	0 (0%)	14 (41%)	20 (59%)
25654 RS8	1	30	15	0 (0%)	3 (20%)	12 (80%)
25654 RS3	1	39	19	0 (0%)	5 (26%)	14 (74%)
25654 RS24	2	212 (106)	48	0 (0%)	1 (2%)	33 (98%)
25654 RS2	2	48 (24)	41	0 (0%)	18 (44%)	23 (54%)
Ital 6 RS2	2	46 (23)	38	0 (0%)	8 (21%)	30 (79%)
Ital 6 RS24	1	50	24	0 (0%)	1 (4%)	15 (96%)
Ital 6 RS8	2	311 (156)	39	2 (5%)	5 (13%)	32 (82%)
Ital 6 RS7	2	154 (77)	31	3 (10%)	1 (3%)	27 (87%)
25169 RS2	<i>Failed to ligate insert into parent vector RS2.</i>					
25169 RS24	1	61	20	0 (0%)	4 (20%)	16 (80%)
<u>p210 ALLs</u>						
P851 RS2	1	180	21	4 (19%)	7 (33%)	10 (48%)
P851 RS24	1	98	12	0 (0%)	2 (17%)	10 (83%)
P851 RS7	2	137 (69)	43	1 (2%)	1 (2%)	41 (96%)
P851 RS8	2	265 (133)	47	0 (0%)	21 (45%)	26 (55%)
F17730 RS7	2	62 (31)	33	0 (0%)	5 (15%)	28 (85%)
F17730 RS8	1	78	20	0 (0%)	4 (20%)	16 (80%)

F24559 RS7	2	132 (66)	37	0 (0%)	4 (11%)	33 (89%)
F24559 RS8	3	159 (53)	41	0 (0%)	8 (20%)	33 (80%)
<u>p210 CMLs</u>						
266 RS2	<i>Failed to ligate insert into parent vector RS2.</i>					
266 RS24	1	55	18	0 (0%)	1 (6%)	17 (94%)
386 RS2	1	77	9	0 (0%)	5 (56%)	4 (46%)
386 RS24	1	316	21	0 (0%)	1 (5%)	20 (95%)
386 RS7	1	46	20	0 (0%)	0 (0%)	20 (100%)
386 RS8	2	475 (238)	48	0 (0%)	7 (15%)	31 (85%)

Table 3-20: Results of transfections with positive control constructs and test constructs containing authentic RSS sites and *BCR / ABL* breakpoint sequences (see Figure 3-45 for details of constructs). a is the total number of independent transfections, b is the total number of clones obtained (average number of colonies per transfection), c total number of clones analysed (by PCR and sequencing), d is V(D)J-mediated recombination between two authentic RSSs or between one authentic RSS and the fortuitous RSS identified at the breakpoints in vivo, e is V(D)J-mediated recombination at other fortuitous sites (within the sequence flanking the breakpoint or in the core plasmid), f is break repair (BR)-mediated recombination (defined as not mediated by V(D)J). Excluded from the table are those that failed to PCR.

The majority of the breakpoints tested (26/28) had a higher level of BR (with an average of 70%) than non-specific V(D)J recombination (with an average 19%) (see Table 3-20). Colonies arising through BR were not mediated by RAG and were scored if one of the following were found: (a) breaks joining the plasmid core sequence together, completely excising the upstream and downstream sequences and the intervening OOP transcription stop site (see Figure 3-48 A), (b) a break in the test *BCR / ABL* sequence and a break in the sequence flanking the RSS (see Figure 3-48 B) (c) a break in the plasmid core and (i) a break in the sequence flanking the authentic RSS (see Figure 3-48 C) or (ii) a break in the test *BCR / ABL* breakpoint flanking sequence (see Figure 3-48 D) and (d) if the sequencing did not directly show a BR event but read over both of the authentic RSS sites. This was also considered to be a BR event as it was not possible for these to be mediated by V(D)J recombination (not shown).

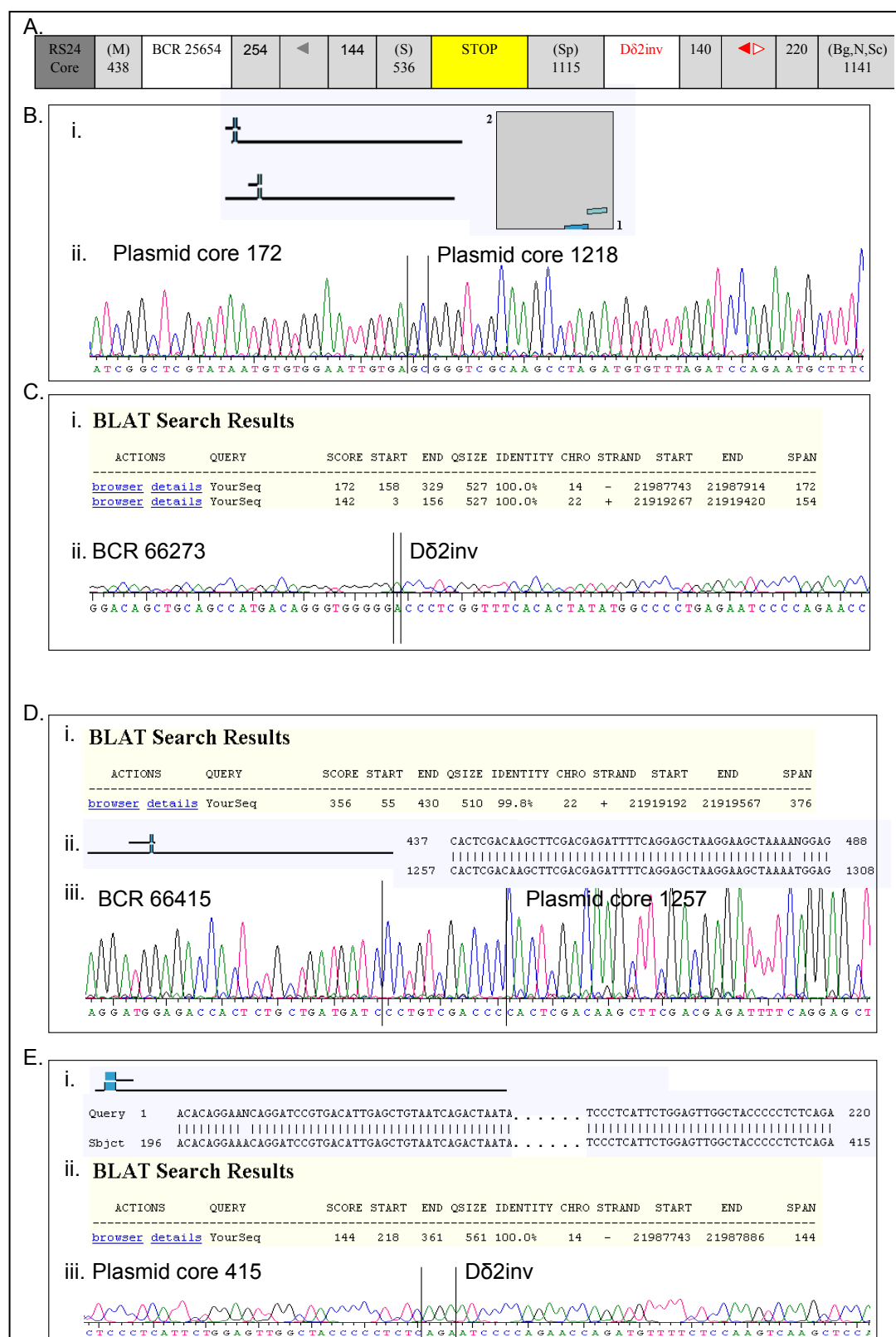


Figure 3-48: BR events in construct 25654 RS24. The vertical black lines mark the breakpoints in each region. (A) Table showing structure of 25654 RS24 construct. (B) BR joining the plasmid core sequence together excising the OOP stop signal. (i) Comparison of colony sequence to plasmid

core sequence. ii. Electropherogram of colony sequence, showing overlapping sequence between the two plasmid regions. (C) BR joining of test DNA (*BCR*) to a site other than the RSSs in D δ 2inv. (i) BLAT of colony sequence results (chromosome 22 = *BCR*, chromosome 14 = D δ 2inv. (ii) Electropherogram showing colony sequence with a *BCR* break at position 66273 (breakpoint at 66292) fused to D δ 2inv 3' of both 12 and 23 RSSs; breaks indicated by the vertical lines with an inserted A nucleotide. (D) BR joining test DNA (*BCR*) to plasmid core DNA. i. BLAT of colony sequence (chromosome 22 = *BCR*) sequence to position 430 then (ii) comparison to plasmid core sequence shows plasmid core DNA from position 437 onwards. (iii) Electropherogram of the breakpoint region, insertion of unknown sequence between breakpoints marked by vertical lines. (E) BR joining plasmid core DNA to a site other than the RSSs in D δ 2inv. (i) Identical sequence between colony sequence and core plasmid DNA up to position 220. (ii) From position 218 onwards, BLAT results show chromosome 14 sequence (D δ 2inv). (iii) Electropherogram of breakpoint showing plasmid core sequence at position 415 to D δ 2inv 3' of both RSSs with a 3 bp homology marked by the vertical lines.

There was also a variety of non-specific V(D)J recombination, with 143/168 (85%) utilising fortuitous RSS sites within the plasmid core DNA (see Figure 3-49 A-D).

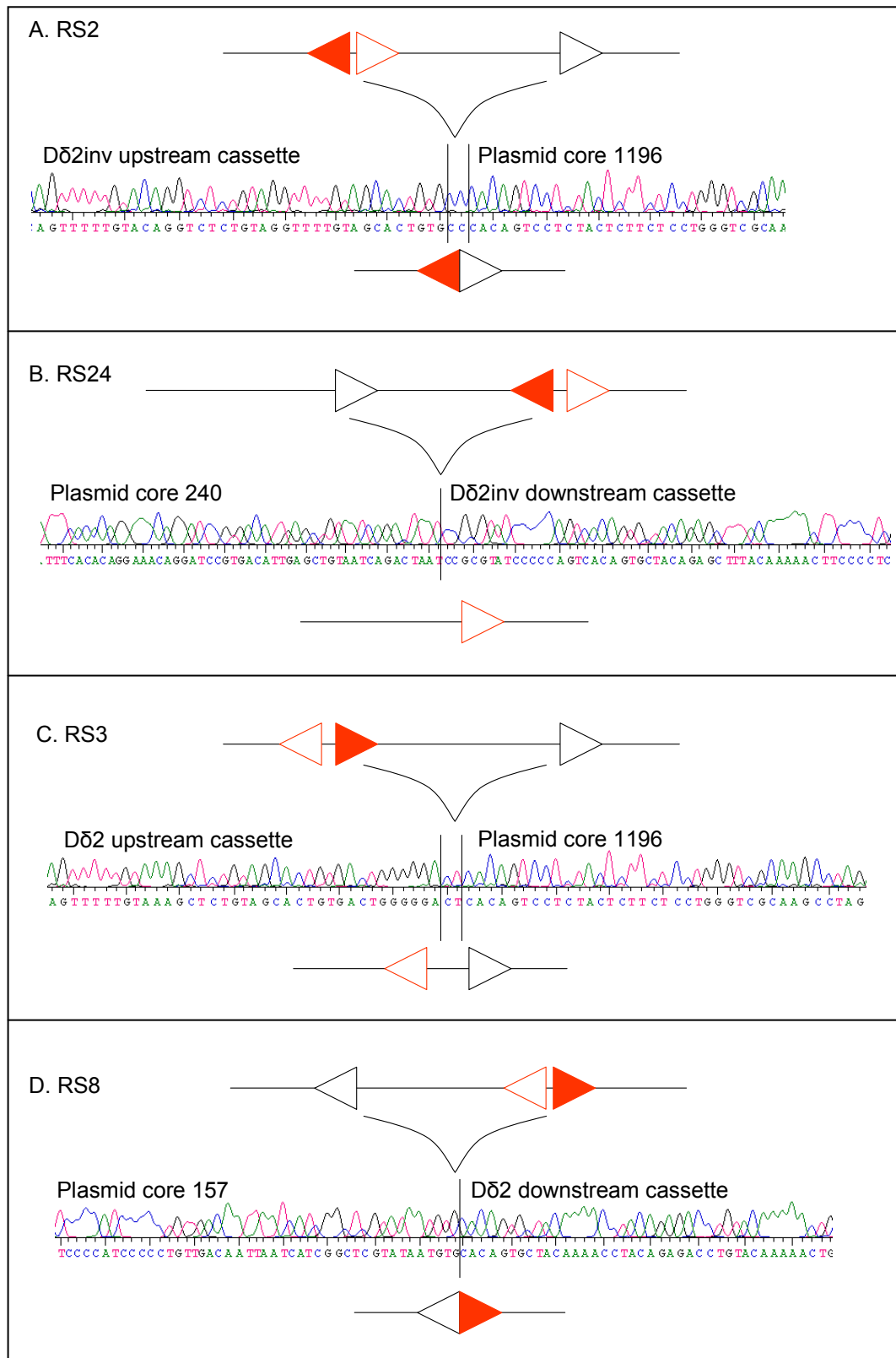


Figure 3-49: Examples of colonies with 'non-specific' V(D)J recombination between authentic RSSs and fortuitous RSSs in the plasmid core. Empty and solid red triangles represent authentic 12 and 23 RSSs, respectively, and empty black triangles represent fortuitous 12 RSSs within the plasmid

core DNA. The black vertical lines mark the breakpoints in each of the sequences. (A) RS2 signal joint formation with the plasmid core showing 2 non-templated nucleotides (CC) inserted at the breakpoint. (B) RS24 coding joint formation with the plasmid core DNA. (C) RS3 hybrid joint formation with the plasmid core DNA, with 2 non-templated nucleotides (CT) at the breakpoint. (D) RS8 signal joint formation with the plasmid core DNA.

There were a smaller number of ‘non-specific’ V(D)J recombination events 25/168 (15%) utilising fortuitous sites other than the breakpoint in the test sequence in both *BCR* (see Figure 3-50) and *ABL* breakpoint regions (see Figure 3-51).

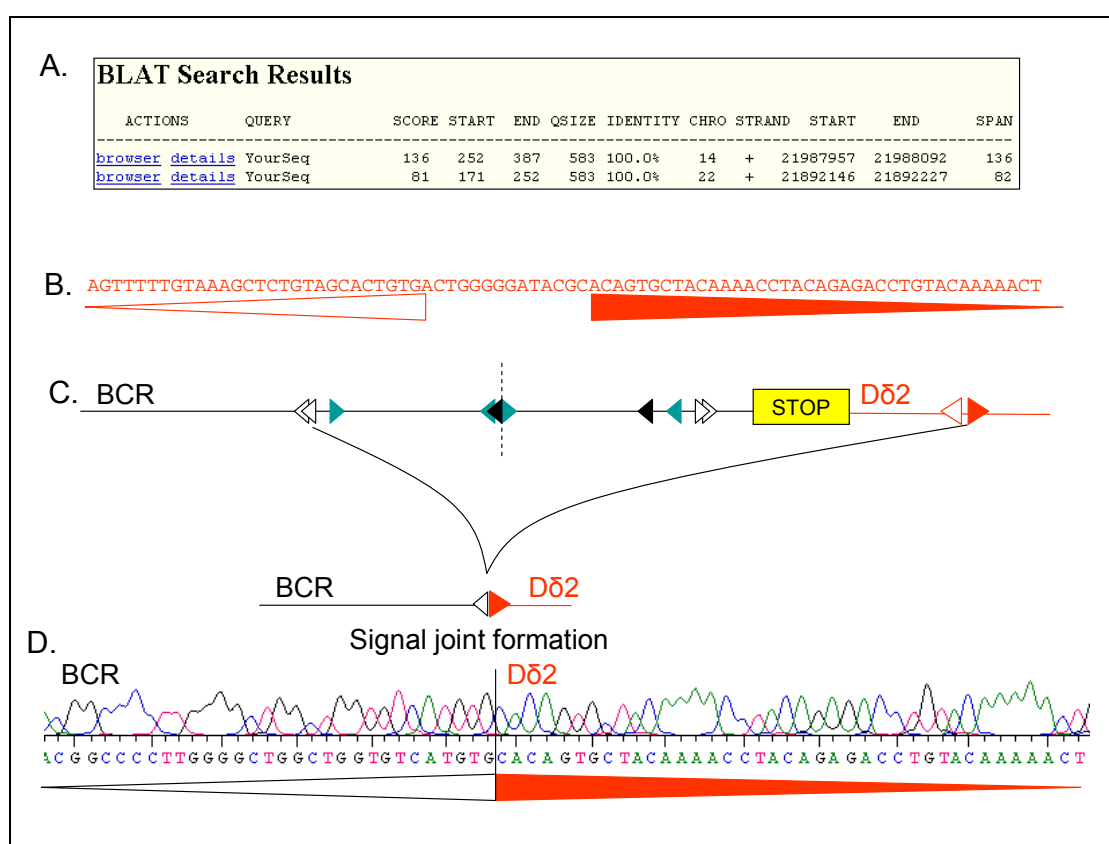


Figure 3-50: Itai 6 RS8 ‘non-specific’ V(D)J recombination found in one colony from the cell culture assay. (A) BLAT results of colony sequencing. (B) Dδ2 RSS sequence. (C) Schematic representation of the signal joint formation, Itai 6 *BCR* recombination cassette is to scale with all the cryptic RSSs represented by triangles (12 RSSs are white, 23 RSSs are black and 12/23 RSSs are green). The authentic RSSs in Dδ2 are represented by red empty or solid triangles (12 and 23 RSSs respectively). (D) Electropherogram of the colony sequencing with the breakpoint marked by the vertical black line.

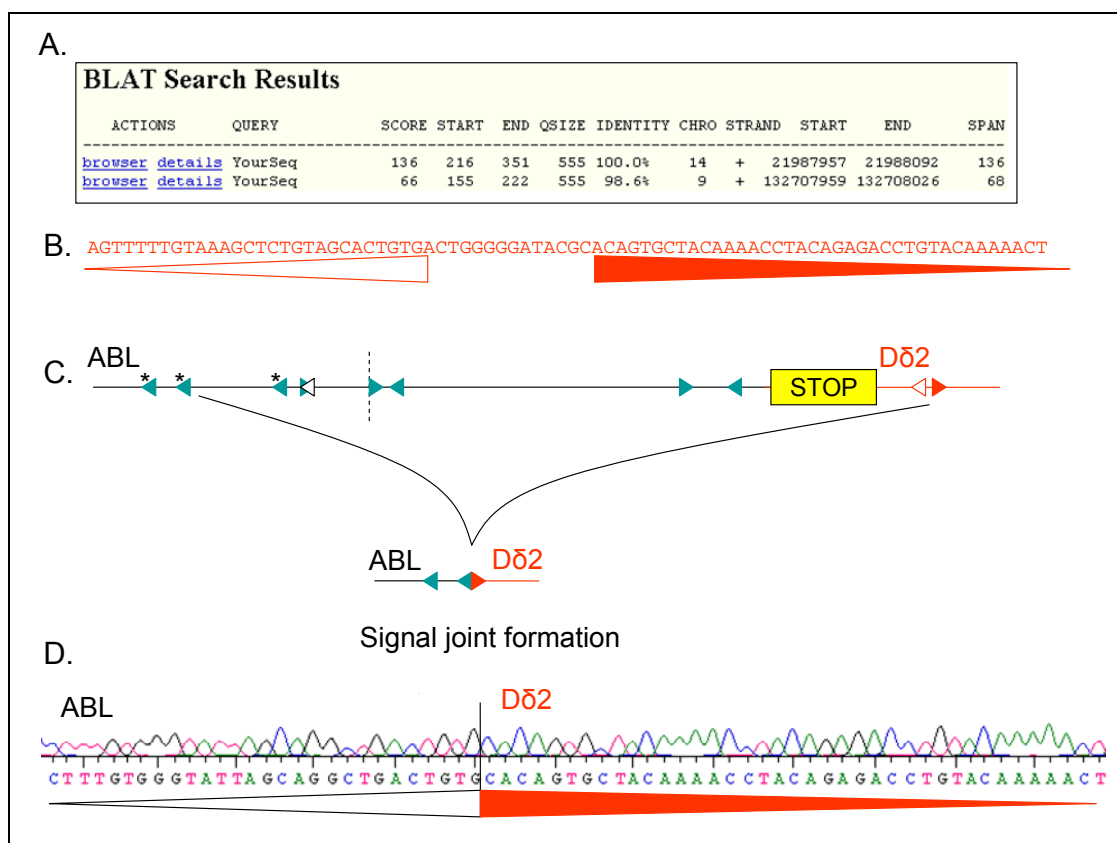


Figure 3-51: P851 RS8 ‘non-specific’ V(D)J recombination found in 13 colonies from the cell culture assay, utilising 3 different cryptic 12 RSSs (asterisked). (A) BLAT results of colony sequencing. (B) Dδ2 RSS sequence. (C) Schematic representation of the signal joint formation, P851 ABL recombination cassette is to scale with all the cryptic RSSs represented by triangles (12 RSSs are white and 12/23 RSSs are green). The authentic RSSs in Dδ2 are represented by red empty or solid triangles (12 and 23 RSSs respectively). (D) Electropherogram of the colony sequencing with the breakpoint marked by the vertical black line.

However, some of the transfections yielded no RAG mediated (neither specific nor non-specific) recombination and 8/28 of the breakpoints tested (29%) had fewer than 10% RAG mediated recombination, implying that these transfections may not have worked efficiently. Indeed, the average frequency of non-specific V(D)J recombination using construct RS8 was 22%, compared to 70% using the same authentic RSS by Marculescu *et al.*, (207).

3.3.12.5.1 Identification of ‘specific’ RSSs at the breakpoints in two patients

The results in Table 3-20 show that most of the breakpoints tested did not have functional ‘specific’ RSSs at or near the breakpoints. Of the 10 cases tested, two were found to have

low level functional RSSs at the breakpoints. The first case was the p190 ALL patient Ital 6 which had several RSSs all, within 10 bp of the breakpoint (see Table 3-17). Three colonies (of 31 tested; 10%) from the Ital6 RS7 construct (see Figure 3-45) were found to have a break at the 12/23 RSS 4 bp upstream of the *BCR* forward breakpoint (see Figure 3-52 A). As the 23 RSS from D β 1 was used to form the coding joint, the 12 RSS from the *BCR* breakpoint was the active RSS (see Figure 3-52 B and C). Furthermore, the sequence from the colonies also revealed non-templated nucleotides which are a common feature V(D)J coding joints (see Figure 3-52 D). Upon comparison of the RAG mediated breakpoint with the forward and reciprocal breakpoints for this case (see Figure 3-52 A), the breakpoint was found to be exactly at the reciprocal *ABL-BCR* breakpoint junction (see Figure 3-52 D).

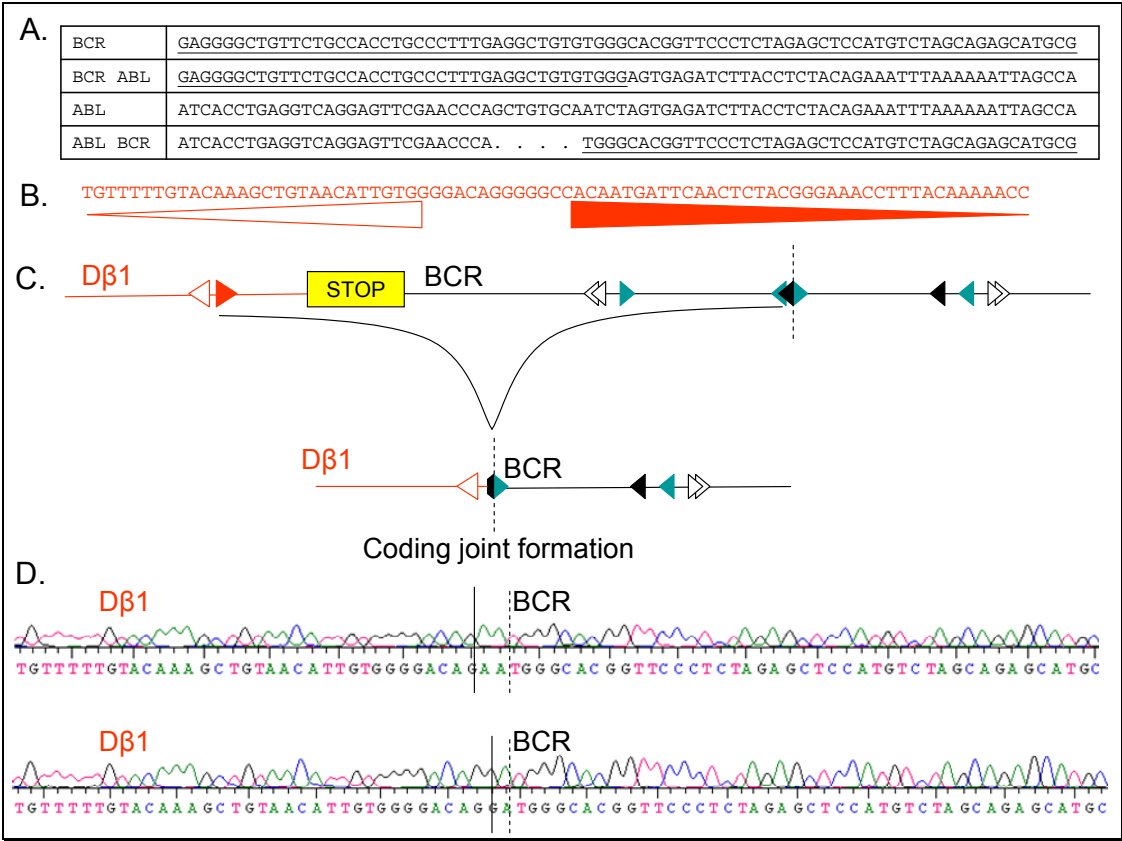


Figure 3-52: Ital 6 RS7 ‘specific’ V(D)J recombination found in 3 colonies from the cell culture assay. (A) Ital 6 forward and reciprocal breakpoint sequences. (B) D β 1 RSS sequence. (C) Schematic representation of the coding joint formation, Ital 6 *BCR* recombination cassette is to scale with all the cryptic RSSs represented by triangles (12 RSSs are white, 23 RSSs are black and 12/23 RSSs are green). The authentic RSSs in D β 1 are represented by red empty or solid triangles (12 and 23 RSSs respectively). (D) Electropherogram from colony sequencing with the

breakpoint in D β 1 marked by the vertical black line, the breakpoint in *BCR* marked by the dotted black line. The intervening nucleotides are non-templated.

Two colonies from the Ital 6 RS8 construct (see Figure 3-45) also showed ‘specific’ V(D)J recombination by formation of a signal joint utilising the same 12 RSS in the colonies with the coding joint formation (see Figure 3-52) at the exact same location (see Figure 3-53). The formation of a coding joint and a signal joint is a definitive feature of V(D)J recombination. There was one additional colony with this construct that showed a signal joint with a fortuitous RSS, 55 bp 5’ of the breakpoint (see Figure 3-53 and Figure 3-50), but this did not give rise to a coding joint in the Ital 6 RS7 construct and there were no other V(D)J recombination breaks involving any of the other cryptic RSSs in any of the other Ital 6 constructs.

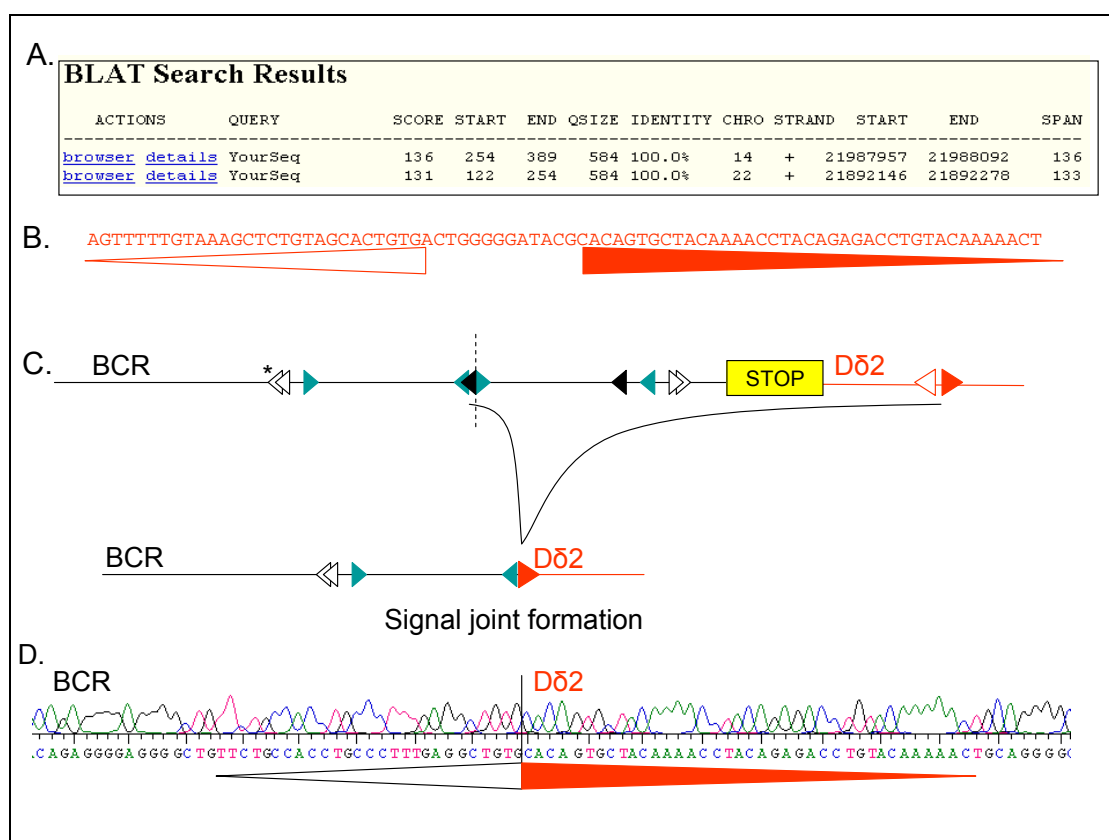


Figure 3-53: Ital 6 RS8 ‘specific’ V(D)J recombination found in 2 colonies from the cell culture assay. (A) BLAT results from colony sequencing. (B) D δ 2 RSS sequence. (C) Schematic representation of the signal joint formation, Ital 6 *BCR* recombination cassette is to scale with all the cryptic RSSs represented by triangles (12 RSSs are white, 23 RSSs are black and 12/23 RSSs are green). The authentic RSSs in D δ 2 are represented by red empty or solid triangles (12 and 23

RSSs respectively). (D) Electropherogram from colony sequencing showing the breakpoint in D δ 2 marked by the vertical black line.

The second patient to show a breakpoint ‘specific’ RSS was the p210 ALL patient P851. One colony in the P851 RS7 construct showed a RAG mediated recombination event utilising the 12 RSS to form a hybrid joint (see Figure 3-54 C). Comparison of the breakpoint junction to the breakpoint locates it exactly to the breakpoint (see Figure 3-54 A and D). This construct was used to test for the 23 forward signal joint (see Figure 3-45) but picked up the 12 forward RSS instead, there were no other V(D)J recombination events within other *ABL* cryptic RSSs for this construct.

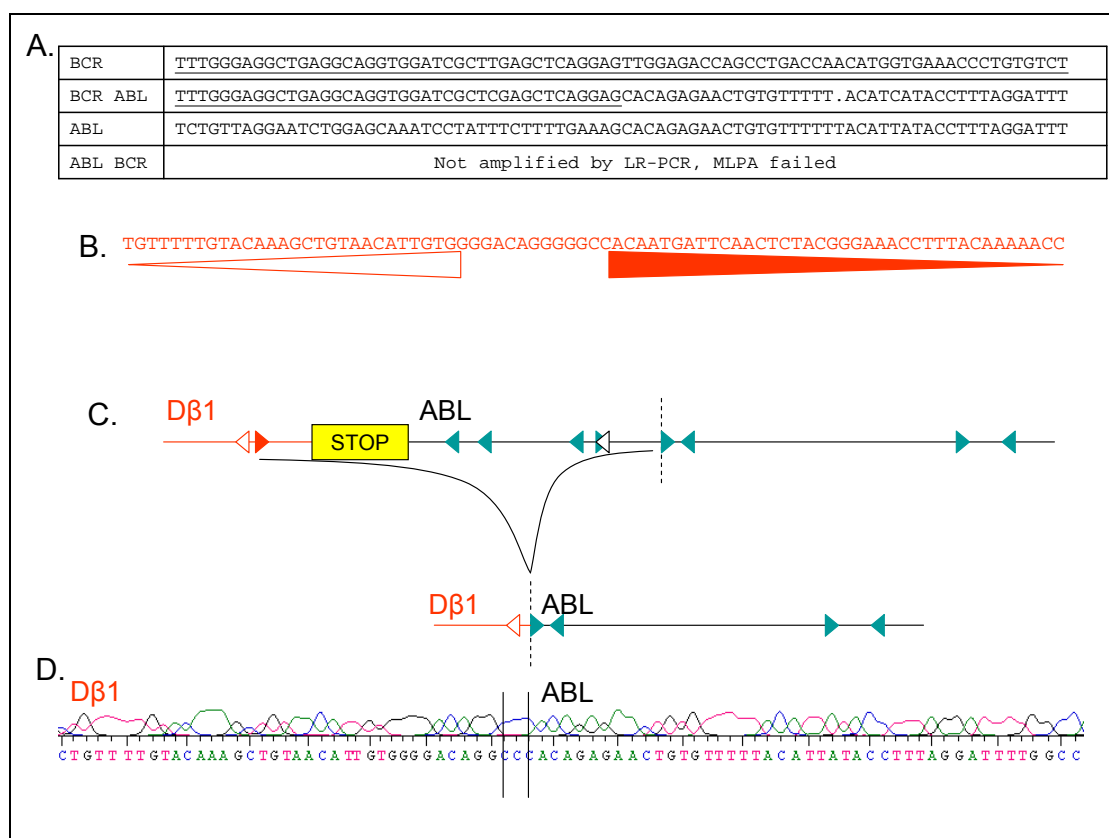


Figure 3-54: P851 RS7 ‘specific’ V(D)J recombination found in 1 colony from the cell culture assay. (A) P851 forward breakpoint sequence. (B) D β 1 RSS sequence. (C) Schematic representation of the hybrid joint formation, P851 *ABL* recombination cassette is to scale with all the cryptic RSSs represented by triangles (12 RSSs are white and 12/23 RSSs are green). The authentic RSSs in D β 1 are represented by red empty or solid triangles (12 and 23 RSSs respectively). (D) Electropherogram from colony sequencing showing the breakpoint in D β 1 marked by the vertical black lines and insertion of non-templated nucleotides CC.

Four colonies using the P851 RS2 construct (see Figure 3-45) were also found to utilise the same 12 RSS located at the breakpoint to make a signal joint. Three colonies were found to have the a signal joint exactly at the breakpoint (see Figure 3-55) and a further colony was found to have a signal joint with a 4 bp deletion in *ABL*. Again no other RSSs either upstream or downstream were found to be utilised in RAG mediated recombination. However, the P851 RS24 construct should have yielded a coding joint with this active 12 RSS, but only two colonies were found to have non-specific RSS with the plasmid, the remaining 10 colonies all showed BR.

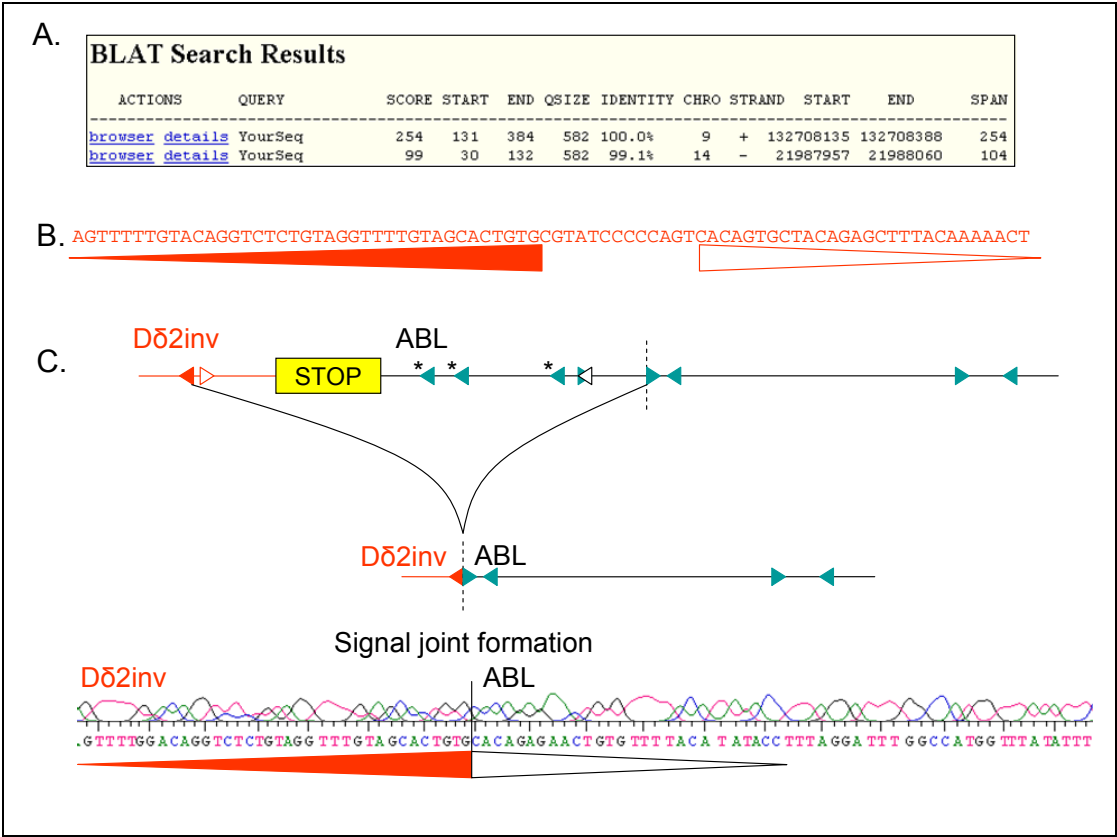


Figure 3-55: P851 RS2 'specific' V(D)J recombination found in 4 colonies from the cell culture assay. (A) BLAT of colony sequence. (B) Dδ2inv RSS sequence. (C) Schematic representation of the signal joint formation, P851 *ABL* recombination cassette is to scale with all the cryptic RSSs represented by triangles (12 RSSs are white and 12/23 RSSs are green). The authentic RSSs in Dδ2inv are represented by red empty or solid triangles (12 and 23 RSSs respectively). (D) Electropherogram from colony sequencing with the breakpoint in Dδ2 marked by the vertical black line.

In addition, 13 of the colonies from the P851 RS8 construct were found to have inverse 12 RSSs that were used in V(D)J signal joint formation with the authentic 23 forward RSS (see Figure 3-55 and 3-51).

3.4 Discussion

In order to successfully treat leukaemia, it is important to understand its pathogenesis. Since the discovery of the Ph chromosome nearly 50 years ago (19), many advances have been made in the understanding of CML and Ph positive ALL that have resulted in the development of novel therapies. Perhaps most notably, the identification the *BCR-ABL* fusion gene (20) and the demonstration that the fusion results in constitutive activation of the ABL tyrosine kinase moiety (21, 31) that has led to the rationally designed ATP mimic imatinib. Imatinib prevents constitutive activation of ABL by preventing the binding and therefore phosphorylation of signalling substrates to BCR-ABL thus halting aberrant cell proliferation (130). However, imatinib, despite being effective on chronic phase Ph positive CML has a decreased efficacy in Ph positive ALL (142). Although both diseases harbour the Ph chromosome, there are distinct differences with over two-thirds of Ph positive ALLs having the p190 BCR-ABL fusion (24) compared with over 99% of CMLs that have p210 BCR-ABL (23). This, along with evidence from recent studies (243, 244) highlights the possibility that CML and ALL fusions may be disease specific and thus may arise by different mechanisms. Such a mechanism might be aberrant RAG recombinase activity (207, 254, 263, 267) in lymphoid precursors specific to ALL. To investigate the hypothesis that illegitimate RAG recombination may give rise to p190 ALL, both the genomic DNA forward and reciprocal breakpoints were amplified and sequenced in order to check for RAG specific signal sequences by homology searches and functional RAG substrate assays. This approach required controls and therefore p210 CML and p210 ALL breakpoints were also analysed. In addition, differences in clustering of breakpoints in *ABL* between p210 Ph positive leukaemia and p190 ALL may also be expected and this was examined using statistical analysis.

3.4.1 Detection of forward *BCR-ABL* breakpoints

In order to facilitate detection of genomic breakpoints for both p210 and p190 positive diseases, several long template PCR-based techniques were evaluated. Use of the long range PCR (LR-PCR) kit (Roche) after extensive optimisation of conditions by changing buffers, template DNA concentration, enzyme concentration etc, was found to be more robust than the High Fidelity PCR kit (Roche). This is mainly because LR-PCR is designed to amplify larger DNA templates and was therefore more reliable at amplification of templates up to 12 kb in size, which in turn necessitated fewer primers for screening. This kit also yielded much brighter bands even though less template was used. This was important as there was only a limited amount of DNA available for many cases.

Initially, simplex PCR was used to screen all cases for breakpoints and although this was successful in identifying a number of breakpoints, it soon became clear that a substantial proportion of breakpoints, especially the p190 patients, were not being picked up. Therefore the number of screening primers in both *ABL* and the m-bcr were increased to adequately cover the region within which breakpoints can occur. However, using all the *BCR* and *ABL* primer combinations (not including the *ABL* upstream primers) to screen one p210 or p190 patient with simplex LR-PCR would require 20-40 or approximately 225 individual PCRs, respectively (see Figure 3-1 for primer locations). However the simplex was easily converted to a multiplex LR-PCR as all PCRs were already optimised to work under the same conditions.

3.4.1.1 Detection of p210 ALL and CML genomic breakpoints

Although the simplex PCR worked well for p210 cases amplifying breakpoints in 6/7 in the initial test set, the multiplex LR-PCR was more efficient as a preliminary screen, reducing the number of PCRs from 20 to 4. However, initial testing yielded breakpoints for only 2 of 3 p210 ALL positive controls and therefore it was possible that some breakpoints may be missed. Indeed, preliminary screening of newly received p210 cases, revealed a detection rate of 52% (12/23) in p210 ALL and 73% (22/30) in p210 CML. The remaining multiplex negative cases were then screened with the simplex LR-PCR which detected a further 26% (6/23) of p210 ALL and 20% (6/30) of p210 CML breakpoints. The cases that were still negative were then screened with the *ABL* upstream and redesigned primers but

no further breakpoints were found. This indicates that although the multiplex was a good rapid method of detecting the majority of p210 breakpoints, it may miss about 20% of cases that would be detected by simplex PCR. It is also possible that other factors might have contributed to the initial multiplex failures, such as variability in the PCR blocks, a failed well or missed aliquoting of DNA.

In total, using the simplex and multiplex screens, 25/32 (78%) p210 ALL and 32/32 (100%) p210 CML breakpoints were amplified and sequenced (see Figure 3-56), giving the combined screen a very high detection rate. Of those that failed it is possible that the breakpoint in some may have been located in such a way that the breakpoint amplicon may have been too big, although the 10 additional *ABL* primers (C and D) were designed to overcome this. It is also possible that liquid handling errors such as failing to put DNA in all wells for PCR may have contributed to some failures, but as each case was done at least in triplicate (by multiplex, the new multiplex with upstream *ABL* primers and simplex) this was unlikely to occur in the same patient for the same primer on three separate occasions, in addition for each LR-PCR all cases had to give a band with normal control primers to ensure the assay was working correctly.

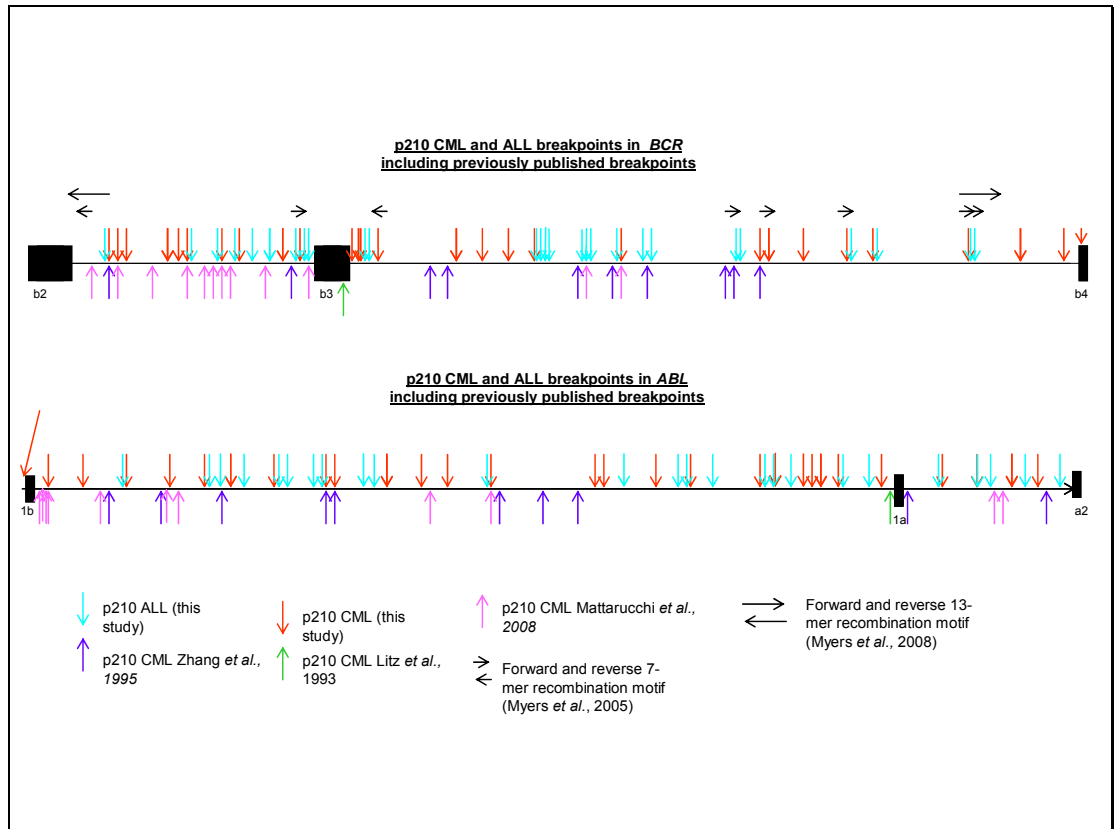


Figure 3-56: Breakpoints in p210 CML and p210 ALL compared to other previously published breakpoints

3.4.1.2 Detection of p190 ALL genomic breakpoints

The identification of p190 breakpoints is more difficult than p210 breakpoints due to the large size of the m-bcr within *BCR* intron 1 and indeed breakpoints in only two have been sequenced to date (214, 216). The initial strategy was to sub-localise the breakpoints using fosmid FISH, prior to LR-PCR with a limited number of primers flanking the predicted fused fosmids. Indeed, this proved successful in identifying 3 of 5 breakpoints, however, the remaining 2 breakpoints were not elucidated until screening with LR-PCR, as the 5' predicted fosmids were actually incorrect (see Table 3-5). This may have been because the background in some cases was quite high and, in addition, ALL cytogenetic suspensions tend to have few metaphases. This together with the fact that fosmids are small probes (approximately 40 kb) means that if the breaks were close to the end of the probe the resultant small signal may not be observed and/or ascribed correctly. It also became apparent that cytogenetic material could not be obtained for all cases and so LR-PCR was the preferred strategy to detect further breakpoints.

Like the p210 cases, the p190 ALL cases were initially screened using the simplex LR-PCR, however the first screen (with 5 *BCR* forward and 10 *ABL* reverse primers) detected no breakpoints. Five more *BCR* primers were designed, resulting in 100 primer combinations for each case, which were tested on a larger number of cases and which yielded 4/13 (31%) breakpoints. In an attempt to increase the detection rate for both p210 and p190 breakpoints, 10 further *ABL* reverse primers were designed. At this point the number of primer combinations for the simplex PCR would have been 200 for each case and was therefore impractical, both in terms of the amount of time it would take to perform, but also with the amount of DNA this would use for each case. By combining each *BCR* forward with 5 *ABL* reverse primers the number of PCRs was reduced to from 200 to 40. Initial testing of the p190 multiplex yielded breakpoints in 4/5 positive controls and therefore this was used to screen further p190 patients. This first multiplex detected 9/25 (36%) of breakpoints in newly received cases and 1/9 (11%) of the simplex negative cases, which was a small improvement on the simplex PCR. This meant that 14/37 (38%) breakpoints had been detected with the multiplex, with a further 5/6 from the fosmid FISH patients giving a total of 19/43 (44%) (see Figure 3-26).

Once a putative breakpoint band had been amplified by the multiplex, the band was confirmed by doing a split apart PCR of the primers in the particular multiplex mix and sequencing with the PCR primers. Most confirmed bands were large (i.e. several kb) with the possibility of the breakpoint occurring anywhere within it, and so several techniques were used to sequence across the breakpoint, in particular digests of the amplified band and further PCRs with primers designed every 1 kb to decrease the size of the breakpoint band prior to sequencing. However, the quickest and most efficient way to sequence across the breakpoint was to design successive primers approximately every 1 kb along the breakpoint region and use the original breakpoint band to sequence from. If a sequencing reaction failed with a particular primer it was assumed to be because this was where the breakpoint was and therefore another primer upstream or downstream, as appropriate, was designed. This strategy was continued until the breakpoint was sequenced.

The numbers of breakpoints identified and rates of detection were still relatively low by comparison with the p210 multiplex, therefore other PCR-based techniques were tested. Firstly, due to the success of bubble PCR in amplifying *FIP1L1-PDGFR*A breakpoints (see Chapter 4), a long range bubble PCR was devised (see section 3.3.4). Unfortunately, this did not work on the normal controls, possibly because bubble PCR relies on relatively inefficient blunt end ligation of the bubble oligo onto the digested DNA fragments. Although this works well for short amplicons, the efficiency may be too low for the longer amplicons required for *BCR-ABL*. Alternatively, the pre-defined bubble oligo specific primers may not have been suitable for LR-PCR, despite the fact that the long range primers were designed to the same annealing temperature as the bubble specific primers. Although it would have been possible to redesign the bubble PCR to incorporate sticky ended ligation, another one sided PCR-based technique called LDI-PCR (256) was trialled, which has the advantage of routinely using sticky ends for ligation. Despite working well on several positive controls, LDI-PCR failed to work on any of the 5 cases tested. The multiplex LR-PCR successfully amplified one of these LDI-PCR negative breakpoints (Ital 5), thus indicating that the LDI-PCR had failed, rather than the breakpoints in the test cases were intrinsically unamplifiable by this technique (see section 3.3.5). This may have been due to incomplete digestion of the genomic DNA, presence of a SNP abolishing or creating a restriction site, or perhaps preferential amplification of a smaller, non-specific fragment. Neither of the one-sided techniques were pursued further as they required a relatively large amount of DNA and were labour intensive.

Failure of these alternative techniques prompted a return to the LR-PCR multiplex. As one upstream *ABL* p190 breakpoint (19635) had already been identified with the multiplex and there were examples of upstream breakpoints in the literature (225, 268), 5 upstream *ABL* primers were designed and at the same time, all the other multiplex primers were checked again on UCSC BLAT to make sure no further SNPs had been identified and that they had no homology to other regions of the genome. As a result of this analysis, 10 primers were re-designed. All previous negative cases were screened using the upstream primers and the newly re-designed primers with the rest of the original multiplex primers (see section 3.3.6). This meant that if any liquid handling errors had occurred in the first screen, giving rise to false negative results, each sample was tested again. However, the re-screening led to the identification of only 1 further breakpoint with one of the re-designed primers and

no further breakpoints were found to be upstream of *ABL*. Therefore a total of 20/43 (46%) had been found, which by comparison to 86% of p210 breakpoints was poor. Comparison of these multiplexes also revealed that another difference between them was the frequency of the *BCR* primers, the smaller p210 breakpoint region effectively had a *BCR* primer every 2-3 kb, whereas the p190 region only had one every ~8 kb. The *ABL* primers were also 8-10 kb apart meaning that for *BCR-ABL* sequence junctions, the maximum distance between the p210 primers was 10-12 kb whereas between p190 primers it could be up to 20 kb and therefore not amplifiable. Consequently 11 more *BCR* forward primers were designed in the m-bcr, as no breakpoints had been found upstream of this region, concurring with other probe based studies of the p190 breakpoint region (214, 269). As some cases were running low on DNA, 5 cases with ample DNA were screened first, and 2 breakpoints were amplified. Sequencing of one of these breakpoints, P653 revealed an insertion of approximately 3 kb of chromosome 11p, which was later found to be a consequence of a complex karyotype. The break did not involve any genes in chromosome 11 and was not found in association with additional *BCR-ABL* breakpoints according to the Mitelman database.

Due to the success of the new *BCR* primers, 2 further negative cases were tested and, in addition, 16 cases for whom only very small amounts of DNA were available, which were therefore whole gene amplified (WGA) prior to screening. As a result 3 further breakpoints were amplified, all from WGA DNA. Although all the LR-PCR normal controls worked well after WGA (once the correct concentration of DNA had been determined) it may be possible that some breakpoints were missed due to decreased quality of the DNA. WGA uses strand displacement to amplify DNA and therefore only gives rise to strands of about 10 kb in length which may decrease the ability to amplify long breakpoint products.

In total, therefore, 25/43 (58%) of p190 breakpoints were amplified compared to 57/64 (89%) of the p210 breakpoints and only 2 p190 breakpoints that have been sequenced in two separate studies (214, 216). Taking the false negative rate of approximately 20% in detecting breakpoints with the p210 multiplex, the discrepancy between p210 and p190 detection rates are almost accounted for since, unlike the p210 multiplex, it was impractical to screen each of the 18 multiplex negative p190 cases with the approximately

200 PCRs that constitute all the simplex PCR combinations in the m-bcr. In addition as over 20 breakpoints were amplified and sequenced for each subset of disease, this was enough to establish whether there was any significant clustering of breakpoints.

3.4.2 Detection of reciprocal *ABL-BCR* breakpoints

In addition to statistical analysis of forward breakpoints to determine if there were any significant clusters, it was also important to look at sequence signatures that are elucidated by sequencing the reciprocal breakpoints, such as micro-homologies, deletions, insertions and duplications in order to try and establish the mechanism of recombination (223, 224, 232, 270). To this end, in addition to the forward breakpoints, the reciprocal breakpoints were also sequenced (see section 3.3.7 and Appendix III). In theory reciprocal breakpoints should be easier to locate if completely balanced however, in practice some have large deletions (125, 127, 128). Therefore firstly primers were designed to flank the reciprocal breakpoint based on the position of the forward breakpoint for use in LR-PCR. Using this method 46/82 (56%) of the reciprocal breakpoints were amplified. These were much easier to sequence as the approximate location of the reciprocal breakpoints could be estimated, based on the location of the forward breakpoints and sequencing primers designed accordingly. Since the primers were designed to amplify reciprocal fragments of 2-5 kb, it was assumed that all failures were a consequence of deletions, however it is possible that a small proportion failed for technical reasons.

The results from the 46 reciprocal breakpoints amplified by LR-PCR revealed significant differences depending on disease category: 16/32 (50%) were p210 CML, 10/25 (40%) were p210 ALL, and 20/25 (80%) were p190. The greater proportion of p190 reciprocal breakpoints that were amplified concurs with experiments performed by Melo *et al.*, (124, 271) where expression of the *ABL-BCR* transcript, which is lost in cases with large deletions, was found in 65% of CML, 1/2 p210 ALL and 7/7 p190 ALL patients. This implies that perhaps the p190 ALLs are less likely than the p210 cases to have large derivative 9 deletions, which may reflect different mechanisms by which these translocations arise.

To investigate this further, the LR-PCR negative cases were screened with MLPA (see section 3.3.8) as this method can be used on DNA to detect large deletions on the derivative 9 chromosome (128, 257) which are known to arise in 10-15% of CML cases (125, 127, 128). Of the 36 cases for whom reciprocal breakpoints failed to amplify, adequate DNA was available for 30. These were tested by MLPA and 23 gave informative results. Eleven of these (48%) showed deletions of one or more of the probes, which broke down into 1/2 p190 cases, 5/9 p210 ALL cases and 5/12 p210 CML cases (see Table 3-11).

The remaining 12 LR-PCR negative cases tested by MPLA showed no deletions of the probes and therefore no large deletions. This could potentially be a result of an LR-PCR false negative, however all primers were checked with a corresponding normal control primer on normal DNA prior to use and each LR-PCR done in duplicate so this is unlikely. A large insertion could make the LR-PCR product too large to amplify, but as all primers were designed to be approximately 2-4 kb either side of the breakpoint where possible, an insertion would need to be over 4-6 kb to give a negative result. The sensitivity of the MLPA technique was found to be 60% (128) therefore it is possible that some genuinely positive cases were missed if the proportion of fusion positive cells was below this level. The final possibility is that these remaining patients had relatively small deletions that took out one or both primer sites but were too small for detection by MLPA. Potentially this could have been investigated by designing a more targeted MLPA probe set, or by high density comparative genomic hybridisation or SNP arrays.

Identification of the reciprocal translocations by LR-PCR revealed that the number of perfectly balanced translocations was; 5/25 (20%) for p190s, 2/32 (6%) for p210 ALLs and 0/32 (0%) for p210 CMLs. By comparison, in a series of AML cases with the t(15;17), only 2/29 (7%) cases showed perfectly balanced translocations (224). For the remaining LR-PCR positive cases, both the median number of nucleotides deleted at the breakpoints and the range were smaller for LR-PCR positive p190s (3 bp; n=12; range 1-27) compared to both p210 ALLs (37 bp; n=6; range 3-7883) and CMLs (33 bp; n=13; range 1-985). Only five *BCR-ABL* reciprocal breakpoints have been published to date, and all showed small deletions, duplications and insertions (214, 232, 272), although the bias to detect

small insertion/deletion events may be due to the techniques used to identify the CML cases.

Short stretches of micro-homologies (2-3 bp) were also found in more of the p190s with 4/20 (20%) compared to 1/10 (10%) of the p210 ALLs and 1/16 (6%) of the p210 CMLs. The CML results were in contrast with a recent study where they claimed 21/27 (78%) of CML breakpoints had micro-homologies, however they only used forward breakpoints to generate the data and, in addition, were not stringent about the length or position of the micro-homologies (225). Micro-homologies at translocation breakpoints although not a feature of normal V(D)J have been seen in RAG mediated recombination cell assays in the absence of TdT (273). Duplications of sequences from either or both genes were found to be slightly lower with the ALL breakpoints than CML; 5/20 (25%) of p190s, 2/10 (20%) of p210 ALLs and 6/16 (37%) of p210 CMLs. All of the duplications were between 2-7 bp except for 3 of the CMLs which were 173-534 bp. Such large duplications have been found previously in CML (232, 274).

The absence of any large tracts of homology and presence of small deletions, insertions and/or duplications, are features concordant with both *BCR-ABL*, other primary myeloid and some lymphoid malignancies and is suggestive of NHEJ following a double stranded break (214, 216, 223-225, 232, 247, 272). From the analysis of the reciprocal breakpoints it is clear that the p190 breakpoints consist mainly of this type of small event, which is confirmed by results from other studies (271, 275, 276) and that p210 CML and p210 ALL are more prone to larger deletions and insertions (124, 271, 274, 276). The proportions of types of events in p190 reciprocal breakpoints are largely consistent with other breakpoints involving illegitimate V(D)J recombination, e.g. Jager *et al.*, (270) sequenced 40 reciprocal breakpoints with the t(14;18) in follicular lymphoma and found 17.5% were perfectly balanced, 52% had deletions of between 2-22 bp and 25% had duplications of between 1-56 bp (which are substantially larger than those found in p190s which had duplications of between 2-5 bp). These duplications may be consistent with a transposition model whereby RAG1 and RAG2 can insert signal ends into new DNA sites and the subsequent joining of the ends can result in 3-5 bp duplication (277, 278). The two hallmarks of V(D)J recombination – the addition of non-templated (N) nucleotides at the breakpoint junction

and palindromic (P) sequences formed by resolution of DNA hairpins - were not observed, however this does not preclude the possibility that one or both breaks in ALL may be accomplished by aberrant RAG activity.

Identifying and trying to find an association of breakpoints with RSSs is impossible by eye and since an estimated $>10^9$ sequences can theoretically act as RSSs for RAG (263). A search for degenerate RSS motifs was therefore performed along with statistical analysis to determine if there was any evidence for clustering of the forward breakpoints.

3.4.3 Statistical analysis of forward breakpoints

The detailed statistical analyses performed on all the forward breakpoints revealed that there were no significant differences between the positions in *ABL* between CML and ALL. There was however some evidence for clustering of p210 and p190 breakpoints (see Figures 3-29 to 3-31).

The only significant association between any of the disease subsets and sequence motifs tested was for the p210 CML and p210 ALL *BCR* breakpoints which were found to have a deficit of repeat elements, which is in stark contrast to most other studies (214, 268). This deficit of repeat regions may be accounted for by the high number of recombination hotspots in the M-bcr compared to the rest of the gene (see table 3-13). A 7-mer motif has been reported to be associated with non-repeat mediated recombination and both this motif and the related 13-mer (260, 261) are over-represented in the M-bcr. Furthermore, visual inspection indicates that the positions of these motifs correlates approximately to the larger cluster found in p210 ALL *BCR* breakpoints (see Figure 3-56). In addition to searching for known sequence motifs near breakpoint junctions, all breakpoints were searched for any unknown motifs. One novel motif, CTGNNTTG, was found to be associated with p210 ALL, although the significance of this finding is unclear.

To date, many *BCR-ABL* breakpoints that have been reported to be near repeat regions, specifically *Alu* repeats (213, 279) however these observations are based on small numbers of cases. A recent paper looking for associations in 27 CML breakpoints (15 from the

literature and 12 identified in-house) also showed an association of breakpoints with repeat regions (225). The data derived in this thesis was from a larger dataset of 75 breakpoints (25 ALL and 50 CML) that were all amplified by LR-PCR from the majority of cases in the study group. Any amplification bias towards certain areas of the genes should have been minimised by the use of primers designed at regular intervals across the *BCR* and *ABL* breakpoint regions (see Figure 3-1). In contrast, other studies have involved digestion of the genomic DNA and ligation of an adapter prior to nested PCR of short fragments (225, 232). These techniques rely on evenly distributed restriction sites, which may often not be the case, and thus may lead to biases in amplification of those areas in which restriction sites occur frequently.

There was evidence of some clustering of breakpoints in p190 at the 3' end of *ABL*, although when this cluster was compared to known translocation-associated sequence motifs (see Supplementary information for full statistical analysis) no association was found. As previously reported, all p190 *BCR* breakpoints were located in the 3' half of intron 1 (the m-bcr), however there was no other clustering as has been previously suggested by studies that used Southern blotting or association with *Alu* repeats (214, 280). Importantly for this work, searches for cryptic RSSs within 50 bp either side of the breakpoints revealed no statistical enrichment for these sequences. However, these searches used a stringent RSS criteria based on the a position-specific weight matrix (PWM) from Lewis *et al.*, (264) which will find consensus or functional RSSs as defined in that study. It is conceivable that very low efficiency RSSs are involved in the generation of *BCR-ABL* in ALL that would not be detected by such a stringent search (207, 263) or, alternatively, that there are several mechanisms that lead to the formation of *BCR-ABL* with only a proportion being formed by RAG mis-recognition. Thus to definitively ascertain whether illegitimate RAG activity is relevant, sequences needed to be tested functionally.

3.4.4 Functional analysis of RSSs

Three attempts were made to construct plasmids to assess RAG recombination using commercial vectors as the parent vector. The first two attempts failed and the third was not completed due to the low efficiencies of blunt end ligation which makes insertions

difficult. At this point it became apparent that Professor Nadel's group were doing similar experiment, using mouse derived constructs and cell lines and were happy to for their constructs to be used as part of this work (see section 3.3.12). All recombination assays were therefore carried out with their already validated constructs and a newly optimised protocol using lipotransfection. Prior to any work, all constructs were digested and all the test constructs also sequenced to ensure the correct ones had been received. The protocol was then optimised using our NIH-3T3 cell line and tested with the positive control constructs to ensure the assay worked with a sufficient efficiency to detect low level recombinants. As both the RS39 and RS32 positive controls had consistently twice the number of colonies than in Marculescu *et al.*, (207) (see Table 3-20) and showed RAG mediated recombination upon sequencing, the assay was commenced. To determine which breakpoints would be most likely to act as cryptic RSSs, 70 bp either side of each *BCR* and *ABL* breakpoint flanking sequence for each of the 82 breakpoints was screened using previously defined criteria (see section 3.3.2.7) which were less stringent than for the statistical analysis. As V(D)J can result in small (1-5 bp) deletions and insertions in the formation of coding joints (273) and additions can also occur at the signal joint (281), the candidate breakpoint regions were refined to those which had the start of a forward or reverse heptamer up to 10 bp from the breakpoint (see table 3-16 and 3-17). From these, 10 were chosen to represent a good spread of types of breakpoints with a bias towards the p190s as these were the subtype hypothesised to be the breakpoints most likely to undergo illegitimate V(D)J.

The results (although not quite complete as 3 of the 31 constructs failed to ligate) revealed that 1/5 of the p190 ALL cases tested (Ital 6) had a 'specific' recombination event 4 bp from the forward breakpoint but at the exact location of the reciprocal breakpoint in *BCR* and at a reverse 12 RSS (see Figures 3-52 and 3-53). This was confirmed by the finding of both coding joints and signal joints which is the defining feature of V(D)J. The coding joint was found in 3/31 (10%) of the colonies and, despite the presence of 5 other cryptic reverse 12 RSSs in the flanking *BCR* sequence, no other coding joints were formed. The signal joint was formed using exactly the same reverse 12 RSS as with the coding joint and was found in 2/39 (5%) of the colonies. However there was an additional upstream cryptic site that was used in a signal joint formation but this was only found in one colony and was not found to make a coding joint with the other construct so it is likely that this cryptic site

has a much lower efficiency than the one found at the breakpoint. In addition none of the other cryptic RSSs were found to form signal joints or coding joints with the other 2 constructs despite the presence of 10 other cryptic sites.

The other ‘specific’ V(D)J recombination event was in the p210 ALL case P851 (see Figures 3-54 and 3-55). The exact location of the forward breakpoint in *ABL* had a cryptic forward 12 RSS which was used in the formation of a hybrid joint in the RS7 construct in 1/43 (2%) of the colonies. The efficiency of hybrid joint formation is lower than either the coding joint or signal joint as it requires the formation of both (282) and therefore this forward 12 RSS might have quite a high recombination efficiency. Indeed 4/21 (19%) of colonies were also found to form a signal joint with the same 12 RSS and although none of the expected coding joints were found with RS24, however only 12 colonies were analysed and therefore this event may have been missed.

There were no other recombination events involving the *ABL* breakpoint sequence in these constructs, however 3 cryptic reverse 12 RSS sites within *ABL* were identified in the RS8 construct in 13/47 (28%) of the colonies (see Figure 3-51). None of these formed the expected coding joints with the RS7 construct, and it was this construct which gave rise to the ‘specific’ hybrid joint. The finding of a ‘specific’ RSS in a p210 ALL case is not implausible, as it is likely that a proportion of the p210 ALLs may be true *de novo* ALL and therefore arise in a lymphoid progenitor. However it is conceivable that this is a fortuitous finding whereby the signal joint was formed more than once in addition to a hybrid joint.

Although the remaining 8 cases (4 p190 ALL, 2 p210 ALL and 2 p210 CML) showed no ‘specific’ recombination at cryptic sites near the breakpoints, the proportion of ‘non-specific’ V(D)J was much lower than in the assays performed using the same constructs by Marculescu *et al.*, (207). This may imply a decreased sensitivity of the assay and therefore RSSs with low recombination efficiency may not have been detected. The reason for this difference is unclear as both the constructs and protocol were received from Professor Nadel. Two minor variations were made to the protocol: first, the expression plasmids and control plasmids were extracted using the Qiagen endotoxin free maxi-prep kit as

recommended however the test constructs were extracted using the HQ plasmid purification kit (Invitrogen) which is suitable for transfection grade plasmids. Indeed, this protocol was used to grow the positive control RS39 which was found to have the expected efficiencies on transfection and so the use of the different extraction protocol is unlikely to have affected the results. A second reason might have been that the NIH-3T3 cells were supposed to be transiently transfected for approximately 40 hours, however, a 40 hour incubation was difficult to do due to the previous timings of the cell splitting and transfections, and consequently the cells were left to incubate for between 44 to 48 hours. Although the doubling time of NIH-3T3 cells is approximately 20 hours, this slightly longer incubation time might have increased the proportion of BR recombination events, as NIH-3T3 cells that are cultured in conditions of confluency for several weeks have been shown to induce neoplastic transformation possibly as a result of DSBs (283). However, there was no evidence of this in the positive controls. It might also be plausible that although the positive controls worked well for each experiment that there was intra-experimental variation with some constructs not working at the optimum level of efficiency. However to help prevent this, a master mix of the expression plasmids was made immediately prior to each experiment.

Overall, one of the p190 ALLs and one of the p210 ALLs were shown to be able to act as targets for V(D)J recombination at their breakpoints in several colonies despite other competing cryptic RSSs. Clearly these results need to be confirmed by repeat transfections and analysis of larger numbers of colonies, however if they are substantiated this data provides strong evidence that at least a subset of *BCR-ABL* fusions in ALL are formed by aberrant RAG recombination. Additionally neither of the two CML patients generated ‘specific’ RSSs although ‘non-specific’ sites were found (a minimum of 50 bp from the breakpoint) in some colonies of some constructs. Although neither consensus nor functional RSSs were found to be significantly associated with any of the breakpoints, this is not necessarily unexpected as such rare events may be mediated by cryptic RSSs that could be quite different from both consensus and known functional variants. Such sites in the vicinity of lymphoid breakpoints have been shown to be active RAG substrates *in vitro* (207, 263), even in the absence of cryptic RSSs if the DNA adopts a non-B form (254, 267).

It is also possible, especially for the two breakpoints found to be mediated by RAG, that these may be transposition events, where the RAG complex is able to insert signal joints into an unrelated DNA target. These types of events have been found both *in vitro* and *in vivo* (277, 278, 284-286). The preferred targets of transpositions are again non-B forms of DNA, specifically cruciforms (286). It may be that the reason for the p190 ALL breakpoints are all found within the second half of *BCR* intron 1 relates to the presence of an internal promoter (287). Potentially, this promoter might cause a distorted DNA structure that may be more fragile to transposition or nicking by the RAG complex. It is not obvious though why these effects might extend over the remaining ~30 kb of intron 1. Alternatively it may be associated with unmethylated DNA, which is the only form of DNA RAG can cut (252), indeed unmethylated promoters in *ABL* were found in p190 ALL but not in either of the p210 subtypes (253).

Chapter 4: Utility of genomic breakpoint detection in patients with hypereosinophilia

4.1 Introduction

The finding of the cryptic deletion that fuses the *FIP1L1* gene to the *PDGFRA* gene revolutionised the diagnosis of CEL (184, 193). Importantly the presence of the *FIP1L1-PDGFRA* fusion gene predicts a favourable response to the small molecule inhibitor imatinib and most positive patients show dramatic responses to therapy with rapid normalisation of peripheral eosinophil counts and achievement of nested or real time quantitative reverse transcription PCR negativity (184, 195, 197). Identification of patients who are positive for *FIP1L1-PDGFRA* is therefore critical to appropriate clinical management (197, 288). Exclusion of *FIP1L1-PDGFRA* is also important as some patients may benefit from alternative therapies such as the IL5 therapeutic antibody mepolizumab (198). The *FIP1L1-PDGFRA* fusion is the most common lesion found in IHES/CEL and the 7 published genomic breakpoints within *FIP1L1* (184) are very variable. Conversely the breakpoints in *PDGFRA* are all very tightly clustered within exon 12. This exon encodes the juxtamembrane domain, and the breakpoints occur specifically between the regions encoding the two tryptophan (W) residues in the WW-domain (see Figure 4-1).

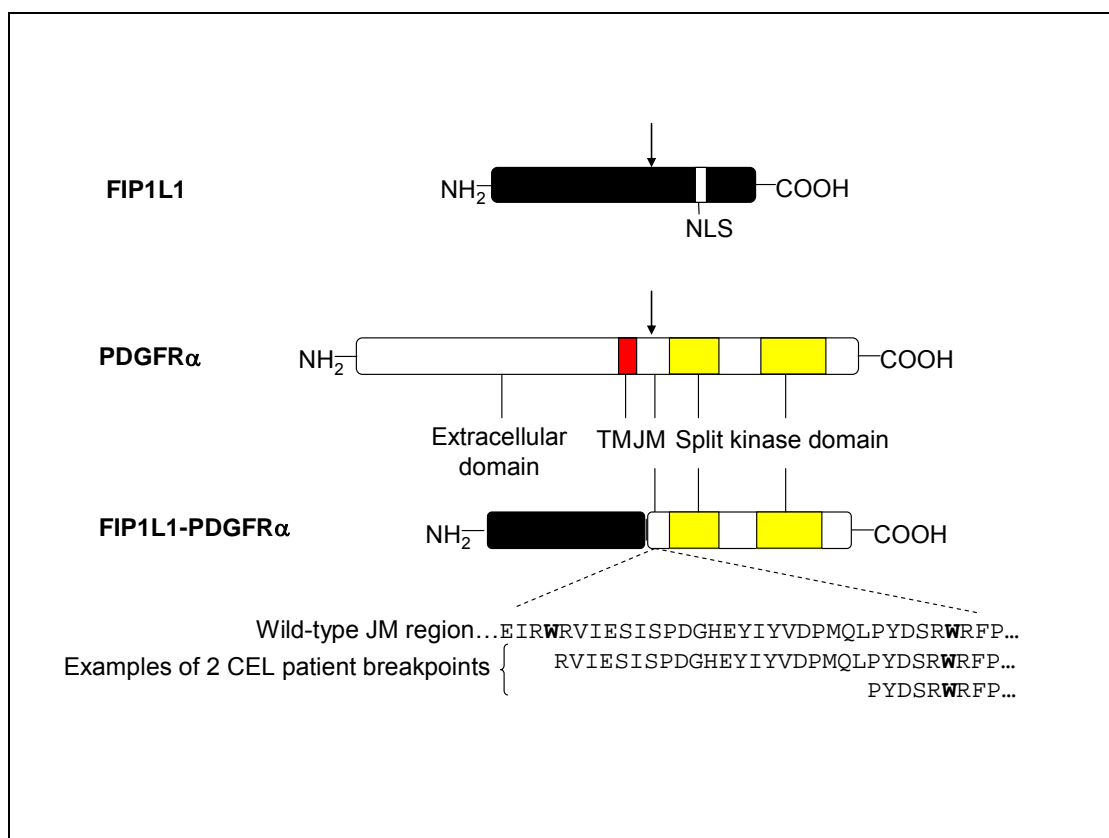


Figure 4-1: The FIP1L1, PDGFR α and FIP1L1-PDGFR α proteins. Breakpoints are highly variable in FIP1L1 but all are between the two tryptophan (W) residues in PDGFR α , the WW-domain. Two patient PDGFR α transcripts are shown to represent the two extremes of breakpoints within the WW-domain. NLS, nuclear localisation signal; JM, juxtamembrane region; TM, transmembrane region.

Unambiguous detection of *FIP1L1-PDGFR α* , however, is complicated by several factors. FISH to detect the heterozygous deletion of *CHIC2*, a surrogate marker for the fusion, can be difficult due to the variability in the proportion of cells involved in the malignant clone (289, 290), which is often low and may overlap with the intrinsic background positive rate of this technique. The other routine method for detection of this fusion is reverse transcriptase PCR (RT-PCR) but this is complicated by the considerable diversity of breakpoints in *FIP1L1*, the complex alternative splicing of *FIP1L1* and the variable use of cryptic splice sites in the fusion gene (184, 196, 291). Indeed, *FIP1L1-PDGFR α* transcripts can often only be detected by sensitive nested RT-PCR assays, even in previously untreated patients (184, 288, 292). Nested RT-PCR has a greatly enhanced risk of contamination giving rise to false positive results, potentially leading to misdiagnosis and inappropriate treatment. Although real time quantitative RT-PCR can detect *FIP1L1*-

PDGFRA in many cases (195), the diversity of mRNA fusions makes the design of workable primer/probe sets that are capable of detecting all variants very difficult. Finally, although *FIP1L1-PDGFRA* is by far the most common lesion found in IHES, it is not found in all imatinib responsive patients, (184, 293) suggesting the possible existence of other *PDGFRA* fusions or other imatinib sensitive tyrosine kinase genes. Identification of the full spectrum of imatinib-responsive abnormalities is clearly important for optimal patient management.

To help address these shortcomings the principal aims of this project were:

1. To develop LR-PCR to detect genomic DNA (gDNA) *FIP1L1-PDGFRA* breakpoints which are unique to each patient, enabling a definitive diagnosis.
2. To ascertain if the amplification of gDNA breakpoints enables more robust detection of *FIP1L1-PDGFRA* in patients at diagnosis and a more sensitive means of detecting minimal residual disease (MRD) in patients undergoing targeted therapy.
3. To use genomic DNA amplification as a means to identify new *PDGFRA* fusions.

4.2 Materials and Methods

4.2.1 Patient samples

Patient samples were referred for diagnostic analysis to test for the presence of the *FIP1L1-PDGFRA* fusion gene by nested RT-PCR.

4.2.2 Preparation of blood and bone marrow samples

See section 2.1.1.

4.2.3 RT-PCR

FIP1L1-PDGFRA RT-PCR was performed by Kathy Waghorn. The presence of the *FIP1L1-PDGFRA* fusion gene was tested by nested RT-PCR performed routinely on IHES cases using forward *FIP1L1* primers and reverse *PDGFRA* primers; FIP1E6 F1 and PDA

R1 for first step and FIP1E6 F2 and PDA R2 for the nested step, using Invitrogen Taq (Invitrogen, Paisley, UK) and standard amplification conditions with an annealing temperature of 60°C.

4.2.4 Cloning, culturing and plasmid extraction

RT-PCR products were cloned and sequenced to verify their identity (see sections 2.5 and 2.7). LR-PCR and bubble PCR products were usually sequenced directly but in some instances were cloned prior to sequencing. Culturing and extraction of plasmids was performed as described in section 2.7.

4.2.5 *PDGFRA* expression multiplex

The *PDGFRA* expression multiplex was developed and performed by Claire Hidalgo-Curtis to detect over-expression of the *PDGFRA* tyrosine kinase domain relative to the 5' part of the gene, which may be as a result of a fusion gene leading to aberrant *PDGFRA* kinase activity. The technique uses a common *PDGFRA* reverse primer (PDA Exon 15R) and two forward primers one annealing to either side of the common breakpoint region in exon 12 of *PDGFRA* (PDA-exon 10F and PDA-exon 13F) on patient cDNA (292). As *PDGFRA* is not normally expressed in peripheral blood leukocytes, patients without a *PDGFRA* rearrangement do not have either of the bands present, however those positive for a fusion gene causing over-expression of the *PDGFRA* tyrosine kinase domain, should have a band of 290 bp amplified with the PDA-exon 13F and PDA-exon 15R primers and can therefore be used to detect both *FIP1L1* and novel *PDGFRA* fusions (see Figure 4-2).

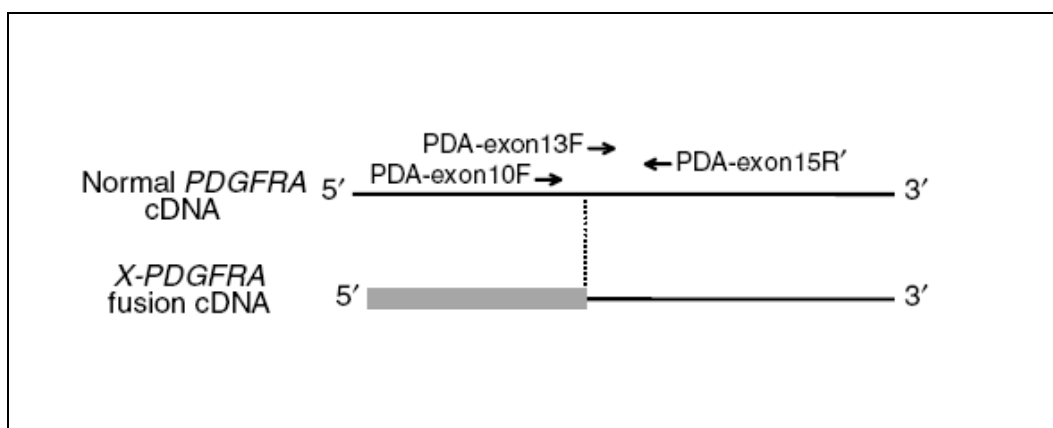


Figure 4-2: *PDGFRA* over-expression by RT-PCR. Schematic showing positions of primers relative to normal *PDGFRA* cDNA. Primers PDA-exon 10F and PDA-exon 15R amplify a 650 bp product from normal *PDGFRA*, but not the X-*PDGFRA* fusion gene (where X represents any partner gene). Primers PDA-exon 13F and PDA-exon 15R amplify a 290bp band from both normal *PDGFRA* and X-*PDGFRA*.

4.2.6 Bubble PCR (see section 2.3.3)

For patients who were positive by RT-PCR or *PDGFRA* expression multiplex, bubble PCR (first described by Zhang *et al.*, (232)) and/or LR-PCR was performed to elucidate the genomic DNA breakpoint. Bubble PCR was particularly useful for *PDGFRA* fusions since the breakpoints within this gene are so tightly clustered. Bubble PCR was performed as in section 2.3.3 using *Rsa*I and *Hae*III enzymes to digest the genomic DNA. Nested PCR was then performed on the purified DNA using standard Amplitaq Gold (Applied Biosystems) conditions (1.5 mM MgCl₂, 0.5 μM primers, 200 μM dNTPs) with *PDGFRA* reverse primers (PDAI12R3 and PDAI12R4), bubble specific forward primers (NVAMP1 and NVAMP2) and a hotstart PCR programme 95°C 15 mins (96°C 20 sec; 66 °C 50 sec; 72 °C 50 sec) for 30 cycles, then 72°C for 10 mins.

4.2.7 Gel purification of bubble PCR products

Bubble PCR products were run on a 2% agarose gel, bands of interest were excised from the gel using a clean scalpel under UV light and each band placed into a sterile 1.5 ml Eppendorf tube. Using the QIAquick Gel Extraction kit (Qiagen) the DNA was extracted according to the manufacturer's guidelines including the optimal washing step for direct sequencing of the product. All extracted DNA was eluted in 30 μl of deionised water and

then sequenced. If the products were weak then the observed product was cloned prior to sequencing.

4.2.8 Simplex LR-PCR to detect genomic *FIP1L1-PDGFR* breakpoints

As the bubble PCR technique involves a laborious, multi-step protocol, a more straightforward LR-PCR strategy was developed to amplify *FIP1L1-PDGFR* gDNA breakpoints. The initial approach, referred to as simplex LR-PCR, was performed on patients who were found to be single or nested RT-PCR positive by routine screening (see section 4.2.3). Once the cDNA breakpoint was sequenced the location of the genomic DNA breakpoint could be approximated and amplified (see section 2.3.2 for LR-PCR methods) usually with a single pair of primers.

4.2.9 Multiplex LR-PCR to detect genomic *FIP1L1-PDGFR* breakpoints

Due to the success of using simplex LR-PCR to detect *FIP1L1-PDGFR* breakpoints (see section 2.3.2.1) and multiplex LR-PCR to detect *BCR-ABL* breakpoints (see Chapter 3), a multiplex PCR was designed using 15 *FIP1L1* forward primers (see Figure 4-9) along the 72 kb region from intron 5 to the end of the gene, therefore encompassing the region within which breakpoints have been found and a common reverse primer in *PDGFR* intron 12 (see section 2.3.2 for LR-PCR methods).

4.2.10 Qualitative gDNA MRD using nested PCR

For detection of MRD in follow-up samples, patient-specific *FIP1L1* forward primers were designed 5' of each genomic breakpoint using the Primer3 website (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) to have an annealing temperature of 66°C. These primers were then tested in combination with generic nested reverse primers on the original positive samples using 50 ng to 0.0005 ng DNA in a 10 fold dilution series in normal control genomic DNA. All follow up sample DNA was quantified using the NanoDrop ND-1000 Spectrophotometer to enable addition of as close to 500 ng of follow-up DNA as possible into the first step PCR in order to be able to achieve detection of low and reproducible levels of the fusion. Nested PCRs were then performed

on the follow-up samples including neat and 0.05 ng positive controls, a normal control and a no DNA control, using AmpliTaq Gold (Roche) standard conditions with an annealing temperature of between 64°C and 66°C. PCR products were then run on an agarose gel (see section 2.4) for end point analysis.

4.2.11 TaqMan real time quantitative PCR (RQ-PCR)

In conventional PCR (see section 4.2.10) the amplified product is detected by end-point analysis by running the PCR product on an agarose gel. However, TaqMan RQ-PCR allows quantitative detection of products by utilising a sequence specific TaqMan dual-labelled probe in addition to flanking primers (see Figure 4-3).

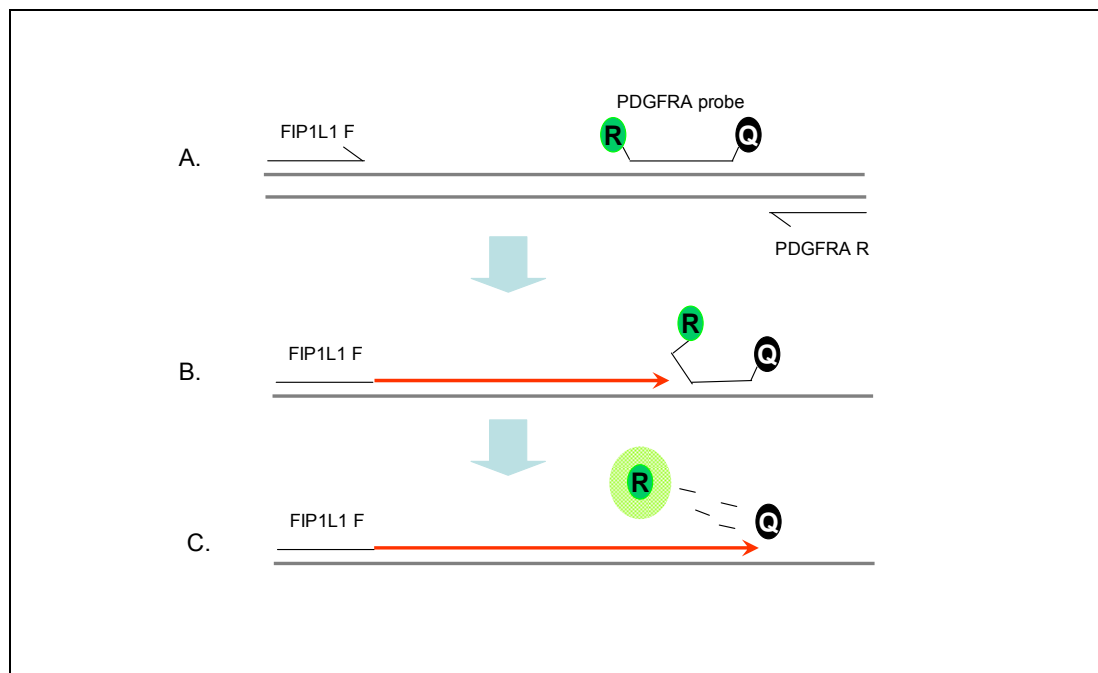


Figure 4-3: RQ-PCR using TaqMan probes. The TaqMan probe contains a 5' fluorescent reporter and a 3' quencher. When intact, the reporter and quencher are held in close proximity. During the annealing/extension step of the RQ-PCR the probe hybridises to the sequence (A), the primers then extend, and it is during extension that the probe is displaced (B) and cleaved by the exonuclease activity of the polymerase allowing the free reporter to fluoresce (C). The amount of fluorescence is proportional to the amount of PCR product and therefore the amount of fusion gene.

For MRD analysis this assay allows important information to be elucidated about the amount of fusion present before and during treatment and therefore allows determination of the success of the treatment regime.

4.2.11.1 Genomic DNA MRD detection using RQ-PCR

Genomic DNA RQ-PCR was used as a quantitative assay to measure MRD in patients with the *FIP1L1-PDGFR* fusion using unique genomic DNA breakpoints. A generic *PDGFR* probe was designed by Sigma Proligo using Molecular Beacon software to be just 3' of the breakpoint region in *PDGFR* to allow amplification of all fusions. This probe had Locked Nucleic Acids incorporated into its structure to increase its specificity. Specific *FIP1L1* forward primers were then designed (using Primer3), to the same T_m (55°C) as the generic *PDGFR* reverse primer (also designed by Sigma Proligo) and as close to the breakpoint as possible as RQ-PCR amplicons need to be short, ideally around 75-100 bp in length. The *FIP1L1* primers were tested on the appropriate patient presentation sample prior to use in MRD analysis, in normal controls and no template controls to ensure that the primers did not amplify any non-specific products.

All reactions were performed on a Corbett Research RotorGene 6000 using 2 x qPCR Mastermix (PrimerDesign, Southampton, UK) and designed to work using standard conditions. 10 µl of 2 x qPCR Mastermix, 0.2 µM forward primer, 0.2 µM reverse primer, 0.15 µM probe, 100 ng of patient gDNA where possible and made up to a final volume of 20 µl with deionised water. Cycling conditions were; 10 minutes at 95°C (15 seconds at 95°C; 60 seconds at 60°C and data was acquired through the green channel at this step) for 50 cycles.

For analysis of MRD, relative quantification using the ΔC_t method (294) was used. The ΔC_t method allows relative quantification of fusion levels by determining the difference in C_t value (ΔC_t) between the *FIP1L1-PDGFR* fusion and a control gene in the pre-treatment sample and the ΔC_t between the fusion and the control gene in the follow up samples, providing that the EAC criteria for assay performance are fulfilled (295). All samples were run in triplicate for the appropriate *FIP1L1-PDGFR* fusion and for the

albumin control gene (296) as a measure of the amount of amplifiable DNA for each sample if negative results for the fusion were obtained and were used to normalise *FIP1L1-PDGFR*A levels for use in the Δ Ct method if the patient was positive for the fusion (see Figure 4-4). A sample was designated positive for the fusion if amplification was seen in at least 2 of the 3 wells with a Ct value of 40 or less. All follow-up RQ-PCRs were set up using the Corbett Robotics Cas 1200 to ensure the highest accuracy possible.

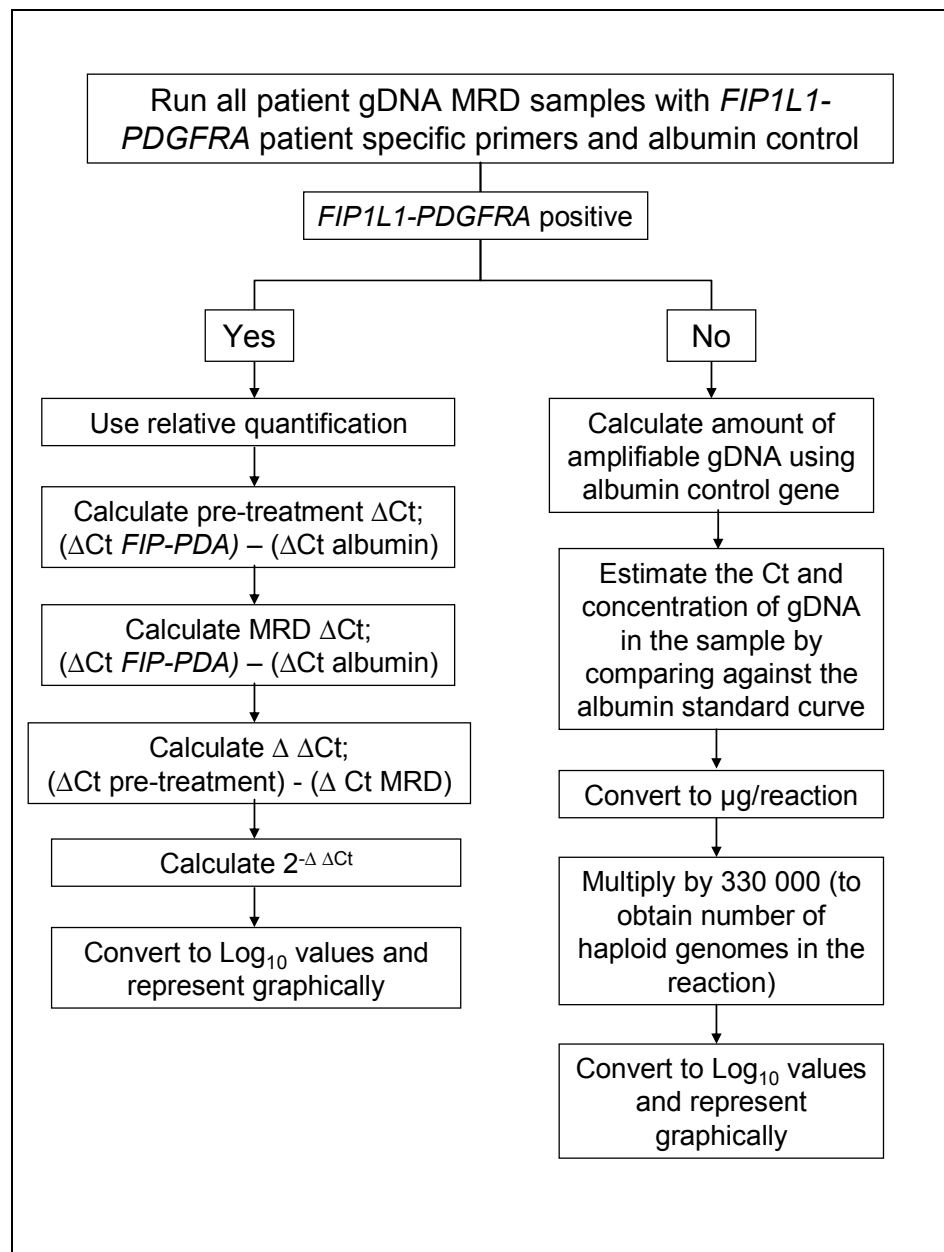


Figure 4-4: Flow diagram showing calculations for relative quantification of gDNA *FIP1L1-PDGFR*A positive MRD samples (left side) and for the amount of amplifiable gDNA for determination of assay sensitivity if the samples were *FIP1L1-PDGFR*A negative (right side).

Where follow-up samples were negative for the fusion gene, the amount of amplifiable DNA for each sample and therefore the sensitivity of the negative result, was estimated by RQ-PCR of the albumin gene. To do this, all patient samples were compared against a standard curve using serial dilutions of known amounts of human genomic DNA (Roche, Burgess Hill, UK) for the albumin gene. Each result was then converted into estimates of the actual amounts of amplifiable DNA in μg and multiplied by 3.3×10^5 to estimate the number of amplifiable haploid genomes (297, 298) (see Figure 4-4).

4.2.11.2 cDNA MRD using RQ-PCR

The cDNA RQ-PCR had been optimised and performed on the cDNA of two WRGL patients as part of a collaborative project by Jovanovic *et al.*, (195) (E370 and E359). To compare the sensitivity of fusion levels in gDNA and cDNA, further cDNA RQ-PCR was carried out to look at the expression levels of the fusion for a selection of patients. RQ-PCR was carried out as in section 4.2.11.1 except 2 μl of cDNA was used and added by hand to the robot dispensed master mixes. The probes and primers for detection of the cDNA fusions were as described by Jovanovic *et al.*, (195) except the *PDGFRA* probe was labelled with BlackHole 1 quencher instead of Tamra. In addition, two new *FIP1L1* forward primers were designed to the same T_m as the *PDGFRA* reverse primer, to test further patients and tested on the appropriate patient cDNA sample with normal control gDNA to ensure no non-specific binding.

4.2.11.3 Absolute quantification of gDNA *FIP1L1-PDGFRA* levels in presentation samples

The unique genomic DNA breakpoint junctions from each of the patients to be tested were amplified by PCR and then cloned into the TOPO pCR4 vector (Invitrogen). The breakpoint containing plasmids were then linearised by RE digestion to ensure efficient denaturation, cleaned up using the PCR purification kit (Qiagen) and quantified on the NanoDrop ND1000. The genomic breakpoint containing plasmids were then used to create standard curves which could be used to quantify the absolute levels of *FIP1L1-PDGFRA* genomic fusion in each patient presentation sample. Each *FIP1L1-PDGFRA* result was

normalised against its corresponding quantified albumin result giving a figure for the number of copies of *FIP1L1-PDGFR*A per ng DNA and then converted into number of haploid genomes (see Appendix V for sample calculations). The absolute quantification was also used to calculate the sensitivity of each of the assays.

4.2.12 FISH

See section 2.2, except: *FIP1L1-PDGFR*A FISH was performed using the Poseidon *FIP1L1-CHIC2-PDGFR*A dual labelled deletion break probe (Kreatech, Holland). The protocol was the same except that the Poseidon probes were provided labelled and required no further detection steps. *KIF5B-PDGFR*A FISH was performed with BAC clones selected from the Ensembl 37 kb clone set as follows: *PDGFR*A BACs RP11-571I18 and RP11-20410; *KIF5B* BACs RP11-16706 and RP11-460H18.

4.3 Results

4.3.1 Detection of *FIP1L1-PDGFR*A mRNA breakpoints

During routine screening for *FIP1L1-PDGFR*A by single step and nested RT-PCR, two specific problems were encountered, (i) in occasional genuinely positive cases the fusion was not reproducibly amplifiable by nested RT-PCR (e.g. only 1 of 3 replicate reactions were positive) and (ii) a period of intermittent PCR contamination was experienced. Contamination was indicated by the finding of a common fusion junction on sequencing the RT-PCR products, however different positive cases may occasionally have identical mRNA fusions and therefore it was not possible to be absolutely certain that the amplification resulted from contamination without obtaining a repeat blood sample. For these reasons we sought to develop better methods to detect the fusion.

During the study period (October 2004-April 2008), 46 patients were confirmed as *FIP1L1-PDGFR*A positive by detection of the genomic breakpoints. In addition, 20 patients were found to be false positives after giving negative results for genomic amplification and, in some cases, sequencing of the cDNA product.

4.3.2 Detection of genomic *FIP1L1-PDGFR* breakpoints

Initially, samples were analysed for genomic breakpoints by bubble PCR only, however, this was a very laborious protocol and therefore a simplex LR-PCR method was used as a primary screen based on the knowledge of the cDNA breakpoint. The simplex PCR was then replaced by the more generic multiplex long range PCR, which has primers spanning *FIP1L1* from intron 5 to the end of the gene.

4.3.2.1 Detection of *FIP1L1-PDGFR* genomic breakpoints by bubble PCR

In an initial attempt to resolve any ambiguity that arose through the results obtained by the nested RT-PCR technique, 30 RT-PCR single or nested positive patients were initially screened for patient-specific genomic DNA breakpoints by bubble PCR and unique breakpoints were identified in 12/30 patients (see Figure 4-5 and Table 4-1). The remaining 18/30 only gave normal sized bands and were therefore thought to be false positive RT-PCR results. However one of these patients (E939) was RT-PCR single step positive, indicating that bubble PCR had failed to amplify the breakpoint, possibly because the distal *RsaI* and *HaeIII* sites were too far from the *FIP1L1* breakpoint to allow amplification of a bubble product. To resolve case E939, a simplex LR-PCR was designed and performed based on the knowledge of the cDNA breakpoint (which had been previously sequenced), and was found to be single step positive by LR-PCR and sequenced to give a unique gDNA breakpoint. However, from the gDNA breakpoint obtained there was an *RsaI* site which should have given a 212bp band by bubble, so perhaps the digest did not work properly, despite the positive and normal control bands working well. As the bubble PCR is a laborious technique and clearly capable of missing some breakpoints, the simplex LR-PCR was used subsequently to confirm positive samples. Once the gDNA multiplex had been optimised, all the bubble PCR negative cases were screened using this technique and a further 3 were found to be positive, having unique genomic DNA breakpoints: cases E458, E566 and E747 (see Table 4-1).

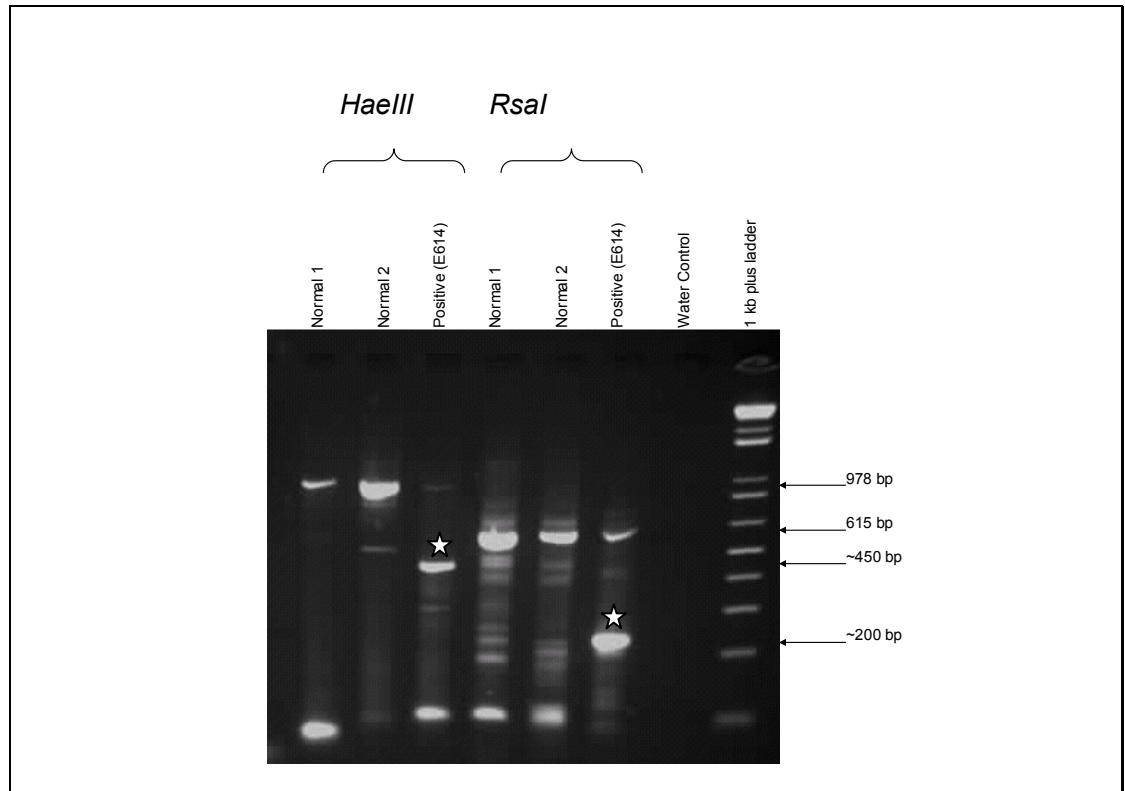


Figure 4-5: Agarose gel showing bubble PCR for *HaeIII* and *RsaI* digested *FIP1L1-PDGFRα* RT-PCR positive and normal patients. Normal bands of 978bp and 615bp for *HaeIII* and *RsaI* respectively can be seen in the normal control and positive lanes. The bands at ~ 450bp and 200bp, denoted by the stars, are products of the genomic *FIP1L1-PDGFRα* fusion.

Due to this work, suspected *FIP1L1-PDGFRα* positive cases are only reported as positive if they are amplified by single step RT-PCR and have a unique cDNA fusion or are detected by nested RT-PCR and the fusion is confirmed by identifying the unique genomic breakpoints. More rigorous precautions were also implemented to prevent cross contamination during the processes of cloning and sequencing the cDNA breakpoints and therefore reduce the chances of further false positive results.

E number	RT-PCR result	Method of gDNA detection
43	Single step positive	Bubble
176	Single step positive	Bubble
243	Single step positive	Bubble
344	Nested negative	Bubble
359	Single step positive	Bubble
370	Single step positive	Bubble
458	Nested positive	Multiplex LR-PCR
513	Single step positive	Bubble
591	Nested positive	Bubble
606	Nested positive	Bubble
614	Single step positive	Bubble
630	Single step positive	Bubble
759	Single step positive	Simplex LR-PCR
905	Single step positive	Simplex LR-PCR
939	Single step positive	Simplex LR-PCR
1025	Nested positive	Simplex LR-PCR
1080	Nested positive	Simplex LR-PCR
1131	Nested positive	Simplex LR-PCR
1279	Single step positive	Bubble
1336	Single step positive	Simplex LR-PCR
1401	Single step positive	Simplex LR-PCR
1424	Single step positive	Simplex LR-PCR
1556	Nested positive	Multiplex LR-PCR
566	Nested positive	Multiplex LR-PCR
635/636	Negative/Single step positive	Multiplex LR-PCR
747	Nested positive	Multiplex LR-PCR
1675	Single step positive	Multiplex LR-PCR
1755	Nested positive	Multiplex LR-PCR
1788	Nested positive	Multiplex LR-PCR
1803	Single step positive	Multiplex LR-PCR
1940/1962	Single step positive	Multiplex LR-PCR
1961	Nested positive	Multiplex LR-PCR
2002	Nested positive	Multiplex LR-PCR
2076	Single step positive	Multiplex LR-PCR
2245	Single step positive	Multiplex LR-PCR

2316	Nested positive	Multiplex LR-PCR
2378	Nested positive	Multiplex LR-PCR
2453	Nested positive	Multiplex LR-PCR
2465	Single step positive	Multiplex LR-PCR
2588	Single step positive	Multiplex LR-PCR
2716	Nested positive	Multiplex LR-PCR
2826	Nested negative	Multiplex LR-PCR
2831	Nested positive	Multiplex LR-PCR
2842	Nested positive	Multiplex LR-PCR
2860	Nested negative	Multiplex LR-PCR (1 mix)
3011	Nested positive	Multiplex LR-PCR (1 mix)
EOL1	Single step positive	Multiplex LR-PCR

Table 4-1: *FIP1L1-PDGfra* patients in whom genomic DNA breakpoints were found and the initial method of determining the genomic breakpoint.

4.3.2.2 Detection of *FIP1L1-PDGfra* gDNA breakpoints by simplex LR-PCR

As the breakpoints in *FIP1L1* are very heterogeneous, there are several different types of cDNA breakpoint. Some *FIP1L1-PDGfra* cDNA sequences contain short pieces of inserted intronic *FIP1L1* sequence (see Figure 4-6). These sequences provide RNA acceptor sites to allow splicing between the end of the *FIP1L1* exon and the beginning of *PDGfra* exon 12 in the absence of appropriate splice sites in the *PDGfra* sequence. Determining the origin of the inserted sequence by homology searches greatly facilitated the identification of genomic breakpoints. For example, the cDNA sequence for E1080 showed a fusion between exon 13 *FIP1L1* and exon 12 *PDGfra*, with an inserted 7 bp of *FIP1L1* intron 13 between these sequences (see Figure 4-6 C). This short inserted sequence occurred twice in intron 13, with only one proceeded with an AG (acceptor splice site) nearly 14 kb downstream of *FIP1L1* exon 13. Forward primers were therefore designed 5' to flank this inserted sequence in conjunction with *PDGfra* generic reverse primers. For this case only, a very limited amount of cDNA was available for analysis, and so nested LR-PCR was used to amplify residual genomic DNA present in the cDNA (see Figure 4-6 A and B). The resultant bands were sequenced and the genomic breakpoint was found to be as predicted from the cDNA insert (see Figure 4-6 C).

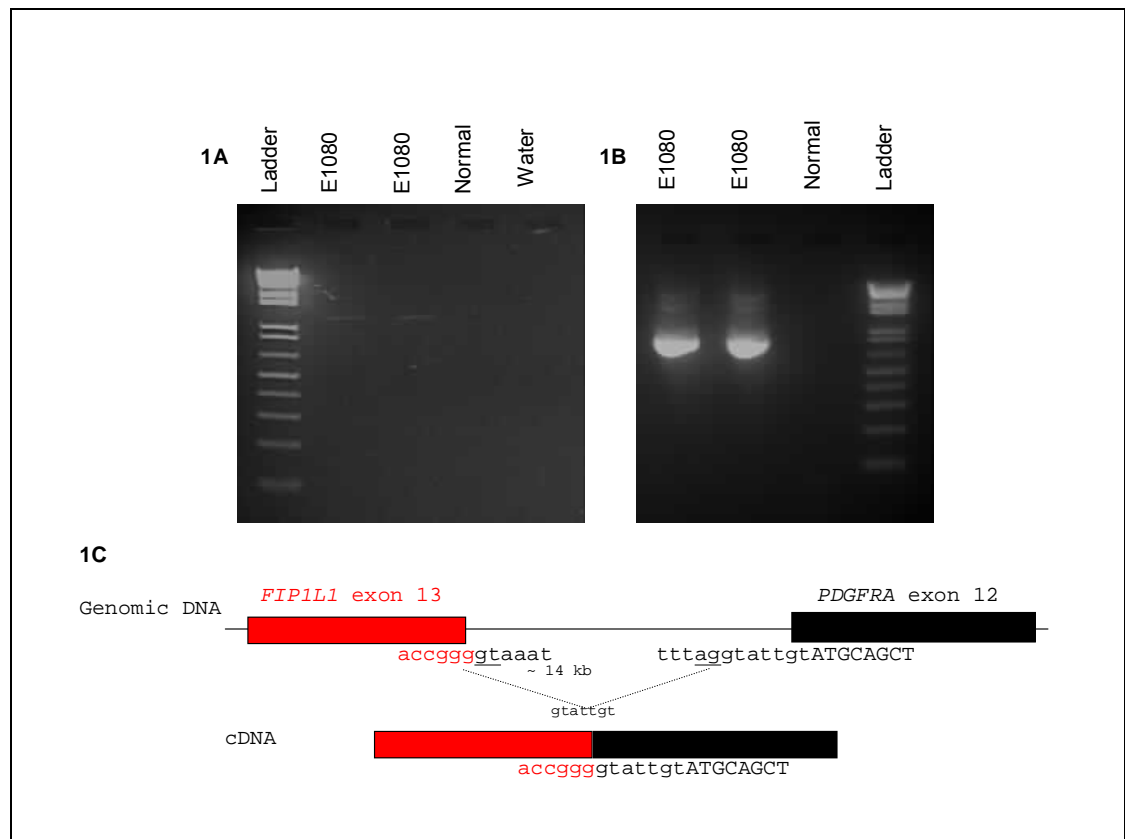


Figure 4-6: E1080 single and nested step simplex LR-PCRs (A and B respectively) amplified with *FIP1L1* intron 13 forward primers designed to anneal 5' of the 7 bp inserted sequence, as gleaned from the cDNA sequence (C) and *PDGFRA* intron 12 reverse primers to enable amplification and sequencing of the genomic breakpoint (C).

For cDNA sequences that did not contain inserted sequences, simplex LR-PCR was performed with primers designed to anneal to the approximate region where the breakpoint was expected to be. For example, E905 was found to have a cDNA breakpoint at the end of *FIP1L1* exon 12 with no inserted sequence and therefore the genomic breakpoint could be anywhere in the 2 kb of intron 12. A *FIP1L1* forward primer was therefore designed at the end of *FIP1L1* intron 11 and used in conjunction with the generic *PDGFRA* reverse primer. A 650 bp product was subsequently amplified (see Figure 4-7 A) and sequenced which revealed that the genomic sequence was the same as the cDNA sequence (see Figure 4-7 B).

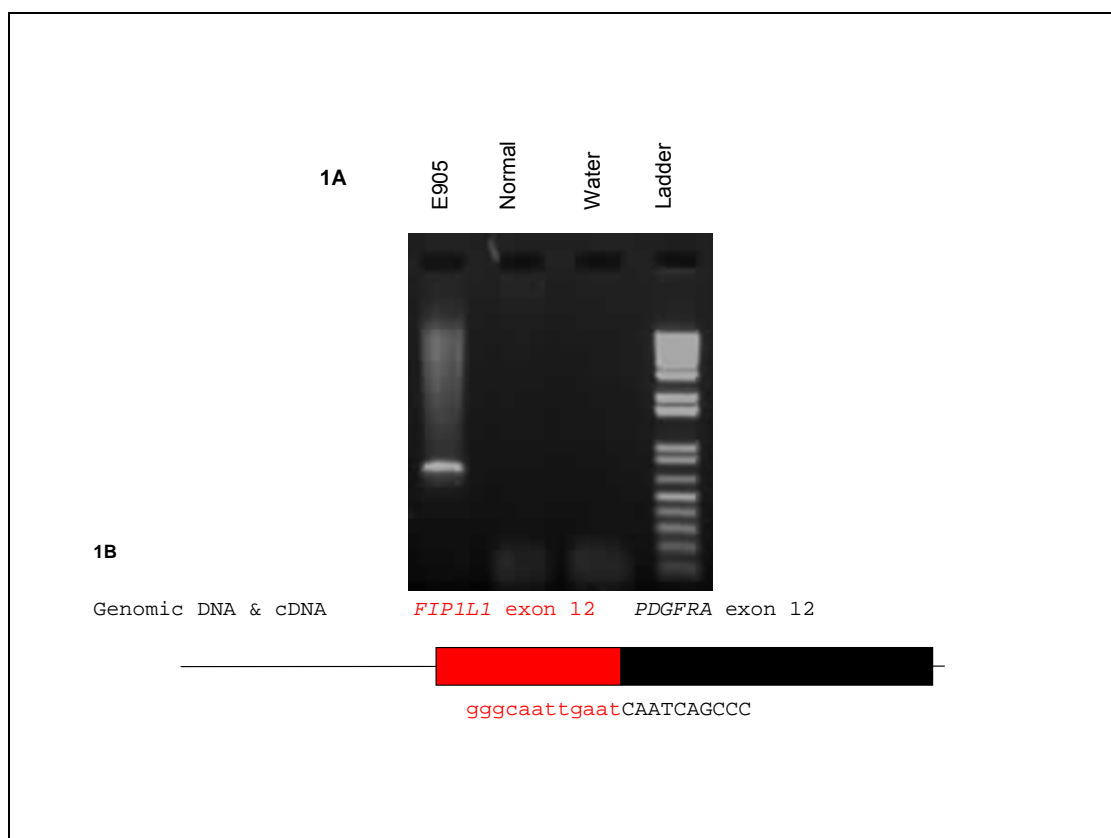


Figure 4-7: E905 genomic DNA simplex LR-PCR product (A) amplified with *FIP1L1* intron 11 forward and *PDGFRA* intron 12 reverse primers. Sequencing with the *PDGFRA* intron 12 reverse primer revealed that the genomic DNA and cDNA sequences were the same (B).

In total, 9/9 patients were found to be positive for the fusion with the simplex LR-PCR at single step when only 6/9 were single step positive by RT-PCR (see Table 4-1).

4.3.2.3 Detection of genomic *FIP1L1*-*PDGFRA* breakpoints by multiplex LR-PCR

Due to the success of the simplex LR-PCR to detect genomic breakpoints in *FIP1L1*-*PDGFRA* RT-PCR positive patients and the success of the *BCR-ABL* LR-PCR multiplex (see Chapter 3), a multiplex PCR to detect *FIP1L1*-*PDGFRA* genomic breakpoints was designed. As the breakpoints in *PDGFRA* are restricted to a very small ~ 100 bp region of exon 12, a common intron 12 *PDGFRA* reverse primer was designed and used in all the multiplex reactions. In order to comprehensively cover the region within which breakpoints occur in *FIP1L1*, 15 *FIP1L1* forward primers were designed and spaced approximately every 5 kb from intron 5 to intron 16, covering all of this ~ 72 kb region of *FIP1L1* (see Table 4-2), with the exception of a region of approximately 12 kb in intron 10

for which primers could not be designed due to the highly repetitive nature of this sequence.

Primer name	Primer position relative to start of <i>FIP1L1</i>	LR-PCR Multiplex mix primer used in
FIP Intron 5 (11403) F	11403	Mix 1
FIP Intron 9 (15351) F	15351	Mix 2
FIP Intron 9 (20848) F	20848	Mix 3
FIP Intron 10 (22850) F	22850	Mix 1
FIP Intron 10 (34060) F	34060	Mix 2
FIP Intron 11 (38554) F	38554	Mix 3
FIP Intron 11 (40225) F	40225	Mix 1
FIP Intron 11 (44469) F	44469	Mix 2
FIP Intron 12 (49317) F	49317	Mix 3
FIP Intron 13 (56273) F	56273	Mix 1
FIP Intron 13 (60580) F	60580	Mix 2
FIP Intron 14 (65181) F	65181	Mix 3
FIP Intron 15 (70370) F	70370	Mix 1
FIP Exon 16 (75275) F	75275	Mix 2
FIP Intron 16 (80314) F	80314	Mix 3

Table 4-2: Details of the primers used in the LR-PCR multiplex to detect genomic *FIP1L1-PDGFR* breakpoints.

Once the primers had been designed, it was necessary to test them in simplex format on as many positive controls as possible to ensure that the primers worked well. At the time, there were a total of 18 *FIP1L1-PDGFR* positive patients in whom genomic breakpoints had been characterised by either bubble PCR or simplex LR-PCR with enough material to carry out testing. LR-PCR is capable of amplification of up to 12 kb products and therefore some patients could have products with more than one of the primer mixes as *FIP1L1* primers were mostly designed to be 5 kb apart therefore, a total of 26 test simplex PCRs were performed on positive control patients (see Figure 4-8).

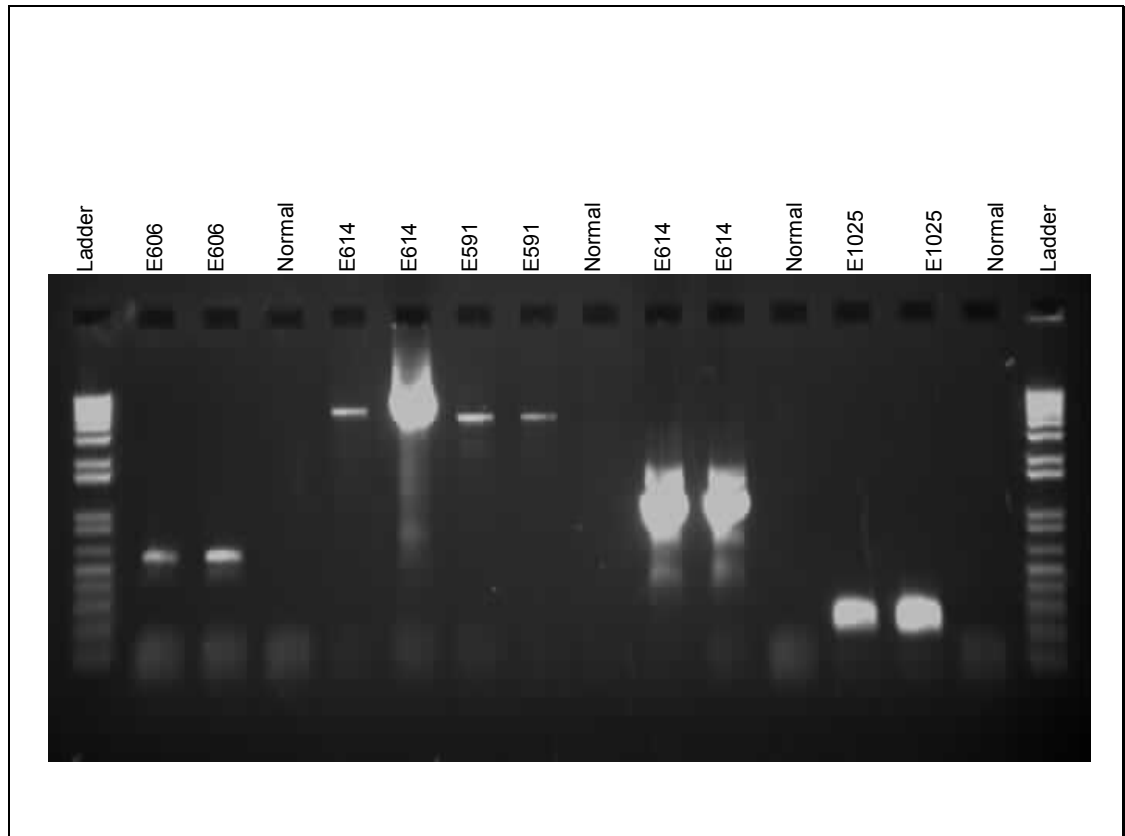


Figure 4-8: Test simplex LR-PCRs on *FIP1L1-PDGfra* positive control cases to ensure amplification of the correct sized bands.

Out of the 26 simplex test PCRs, 16 worked well. The multiplex was therefore performed after splitting the *FIP1L1* forward primers into 3 mixes, each containing 5 forward primers. They were split so that each of the 5 primers in each set were about 15 kb apart rather than with the neighbouring primers of 5 kb apart (see Figure 4-9) as this may have resulted in amplification of more than one breakpoint product with each multiplex, which may decrease the efficiency of the PCR and in turn decrease the detection of breakpoints.

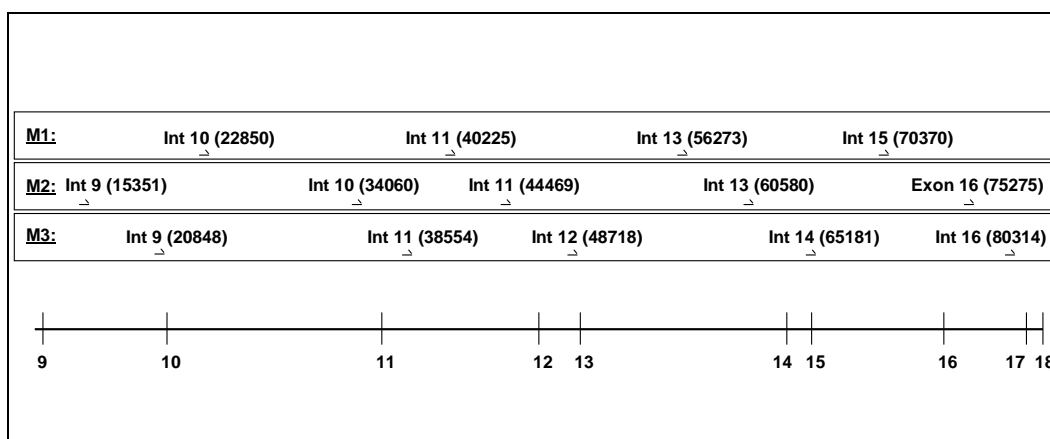


Figure 4-9: Positions of *FIP1L1* screening primers (to scale). The thick horizontal line represents the *FIP1L1* gene from exon 9 to the end of exon 18 and the black vertical lines represent each exon. The *FIP1L1* forward screening primers are represented by the horizontal arrows with their names and which of the screening mixes they are in, above, except the Int 5 (11403) forward primer which was included in Mix 1.

Initially the multiplex was run at an annealing temperature of 65°C, however, the bands obtained were quite weak and therefore the PCR was repeated at a decreased annealing temperature to 64°C in order to increase the intensity of the bands (see Figure 4-10).

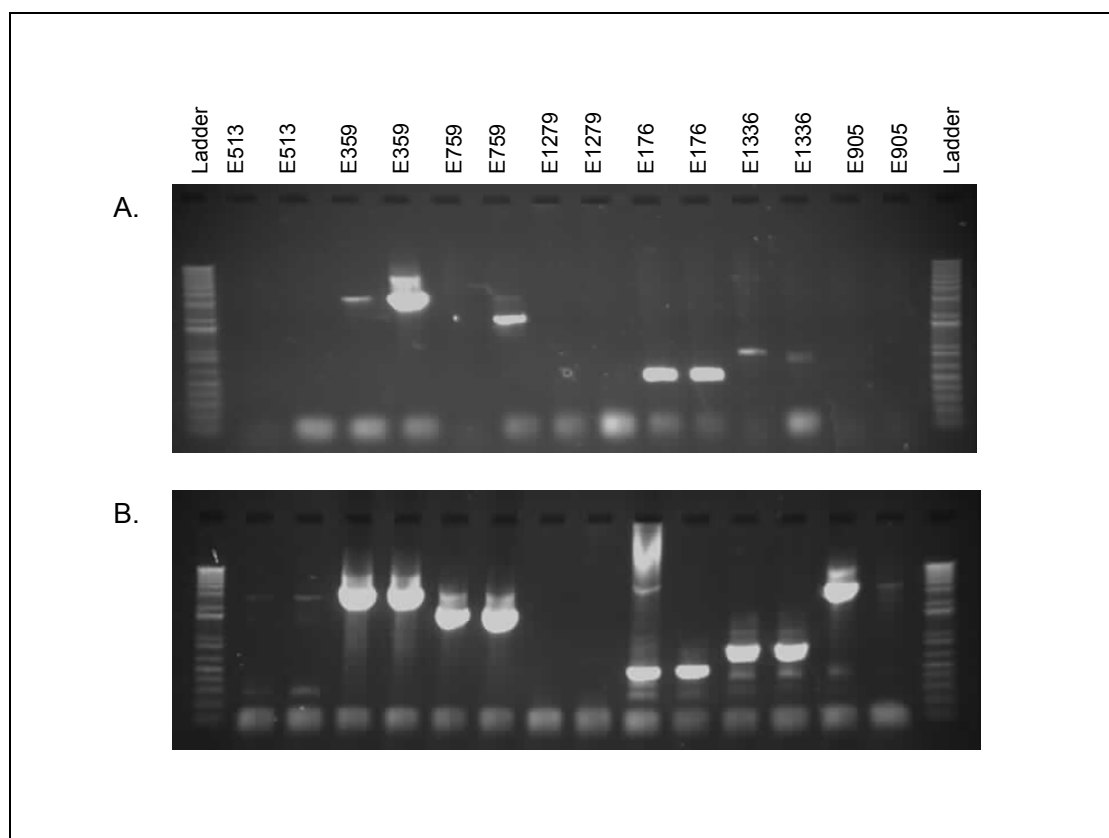


Figure 4-10: Agarose gels of multiplex LR-PCR Mix 2 on positive controls. (A) Multiplex run at an annealing temperature of 65°C. (B) Multiplex run at a decreased annealing temperature of 64°C.

Decreasing the annealing temperature increased the intensity of the bands without leading to any substantial non-specific amplification and therefore all the multiplexes from this point onwards were run at an annealing temperature of 64°C. The multiplex was tested on all the 26 positive control samples (it is important to note that some positive controls could be used in more than one mix; see Figure 4-12 A) and worked well, with 3 out of the 10 that failed to work in the simplex; samples E176 in Mix 1 and Mix 2 and E905 in Mix 2 worked when tried with the multiplex (see Figure 4-11). This may have been due decreasing the annealing temperature from the simplex at 65°C to 64°C in the multiplex.

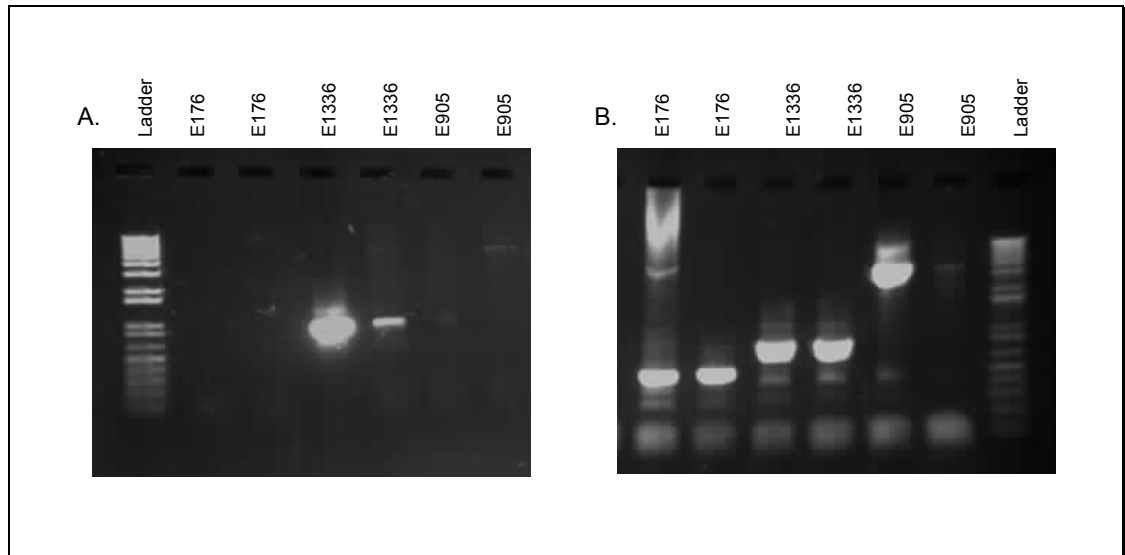


Figure 4-11: Agarose gels of *FIP1L1-PDGFRα* positive control patients. A. E176 and E905 show no bands in the simplex PCR, however, in the multiplex (B) bright bands are seen for both.

Of the remaining 7 positive controls that failed to work in either the multiplex or the simplex PCR, E1279 and E344 failed to produce bands in control LR-PCRs, indicating inadequate DNA quality. These samples represent 4 of the positive controls as they were each used as positive controls in two multiplex mixes (see Figure 4-12 A). Another 2 positive control samples E43 and E513 worked after the appropriate primers were redesigned several times, again E513 was used as a positive control in two of the mixes and therefore all 7 failures were accounted for.

Overall, therefore, all of the 23 positive controls with adequate DNA amplified with the optimised multiplex PCR at single step PCR (see Figure 4-12), suggesting that it would be a useful and rapid screening procedure for the detection of the *FIP1L1-PDGFRα* fusion. Unfortunately, though, the multiplex test cannot be used to definitively exclude the presence of *FIP1L1-PDGFRα* due to the large 12 kb *FIP1L1* intron 10 region that primers could not be designed for.

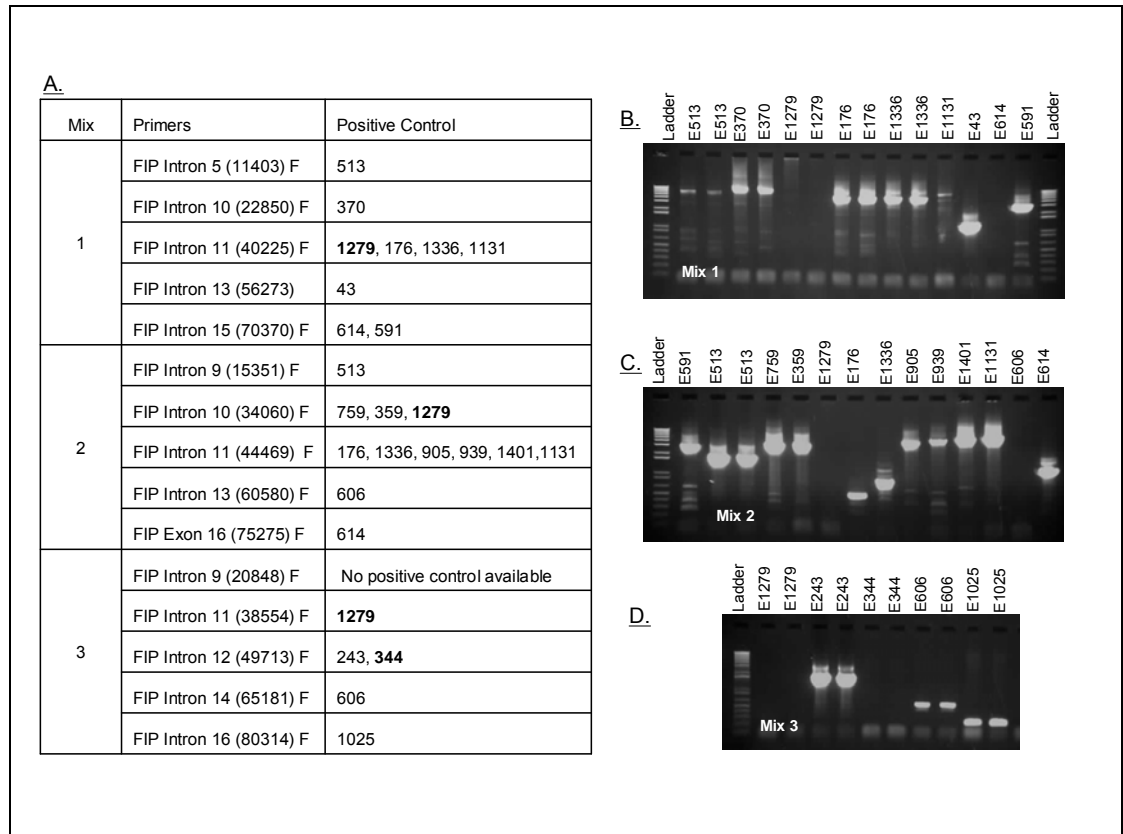


Figure 4-12: *FIP1L1-PDGFR* genomic breakpoint LR-PCR multiplex on all positive controls. (A) Table showing the forward *FIP1L1* primers used in each of the three mixes in combination with the common reverse *PDGFR* intron 12 primer R1. In bold are the positive controls that did not work. (B) to (D) Agarose gels of the *FIP1L1-PDGFR* multiplex on positive controls, Mix 1 E614 and E1131 and Mix 2 E606 failed in this run but were repeated and worked.

Finally, the multiplex was tested on 12 normal individuals (i.e. without any known form of leukaemia) to ensure that there was no non-specific binding and no false positives in normal DNA. Each multiplex mix was also tested with an appropriate positive control to ensure the PCR was working and, in addition, the *FIP1L1-PDGFR* positive cell line EOL1 was tested in order to detect and sequence its genomic breakpoint (see Figure 4-13). The normal controls were also tested with a complementary primer pair designed to amplify products of 6 kb to ensure the DNA was not degraded (ABL 5D F and R). All the normal controls produced bright control bands, and no false positive bands were seen with the *FIP1L1-PDGFR* multiplex primers.

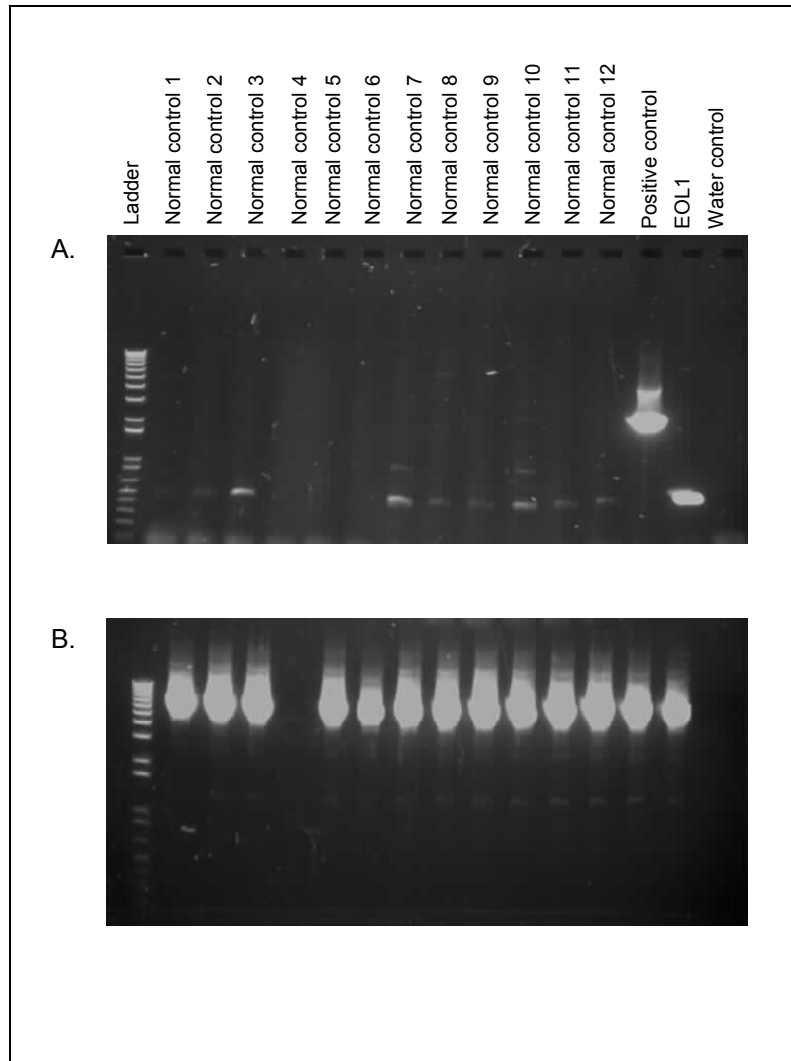


Figure 4-13: Agarose gels of *FIP1L1-PDGFRα* multiplex LR-PCR on normal controls. (A) LR-PCR multiplex mix 3. (B) LR-PCR control using *ABL* 5D F and 5D R primers, on normal controls 1 to 12, with a positive control (E243) and *EOL1* included.

The multiplex was also used to test for a genomic breakpoint in the RT-PCR nested positive patient E1556 (see Table 4-1) and gave a positive band in mix 3 (see Figure 4-14). This led to confirmation of the presence of the fusion by sequencing with the *PDGFRα* intron 12 reverse primer (see Table 4-5). In addition this band was amplified with the FIP Int 9 (20848) F, primer for which there was no previous positive control for, thus indicating that this primer also worked well.

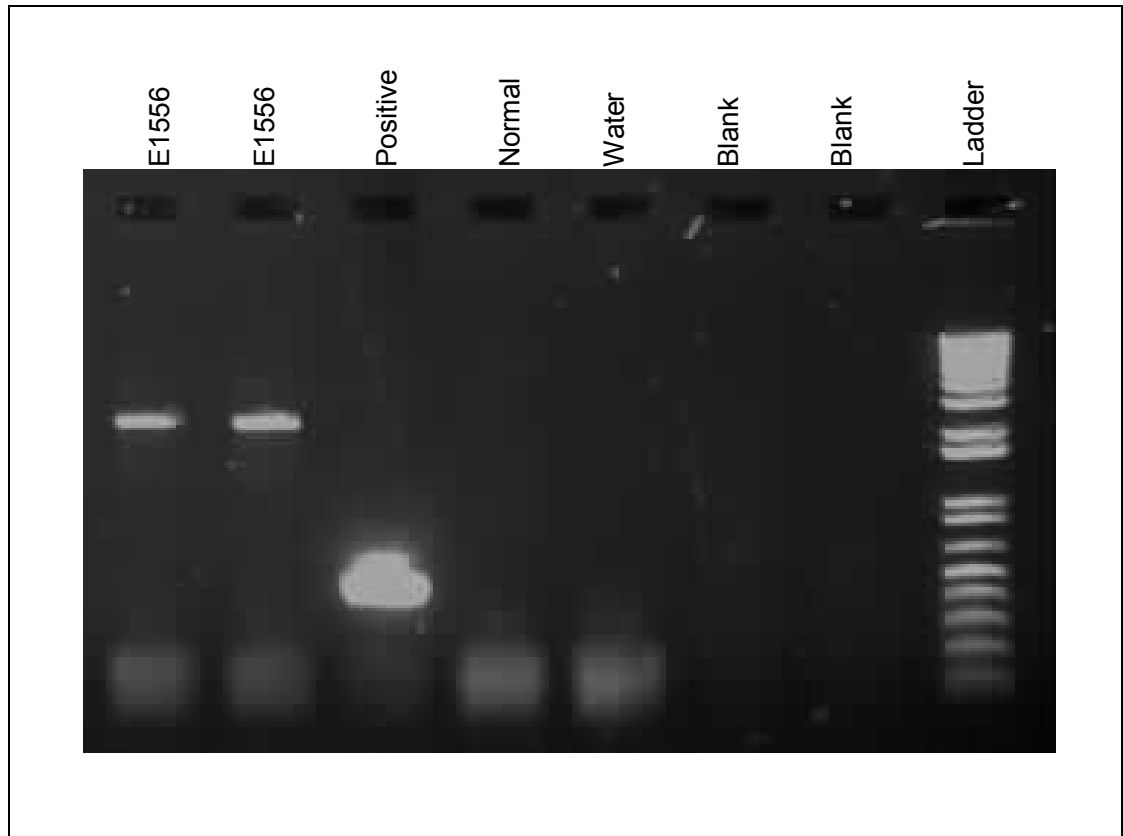


Figure 4-14: *FIP1L1-PDGFRα* gDNA multiplex LR-PCR on E1556 with a band of approximately 2 kb with mix 3.

4.3.2.4 Discrepancies between RT-PCR and multiplex LR-PCR

Unexpectedly, patient E2826 was referred for *FIP1L1-PDGFRα* screening and found to be negative by nested RT-PCR, however another laboratory found deletion of *CHIC2* by FISH. The patient was found to be positive by the LR-PCR multiplex screen with a unique breakpoint, confirming that the fusion was indeed present and had been missed by standard RT-PCR analysis. Repeat testing of the original sample by RT-PCR showed intermittent positivity, i.e. some reactions tested positive whereas others were negative.

The LR-PCR multiplex was then used to retrospectively screen all 202 samples referred for IHES or eosinophilia since the beginning of 2007 and tested negative for *FIP1L1-PDGFRα* by RT-PCR. To facilitate screening large numbers of patients, the multiplex was consolidated into 1 mix with all the 15 forward primers. However, in doing this the possibility of amplifying more than one breakpoint band for each patient was likely which may decrease the efficiency of the PCR leading to false negative results, therefore the

specificity of the reaction was decreased by increasing the concentration of magnesium (see Figure 4-15).

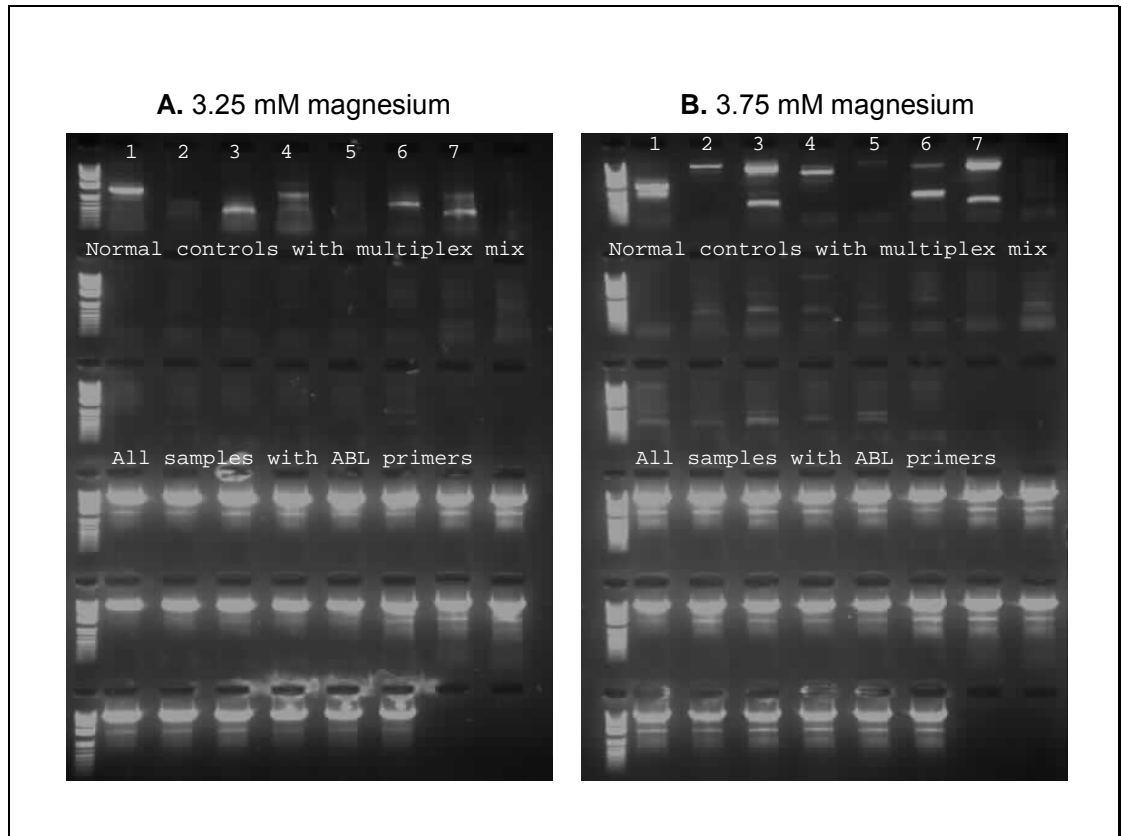


Figure 4-15: *FIP1L1-PDGFRα* gDNA multiplex LR-PCR using one multiplex mix. (A) 3.25 mM magnesium and (B) 3.75 mM magnesium. Lanes 1-7 in each gel are a selection of positive controls.

However, sample E370 (see Figure 4-15 lane 5) did not amplify well. This was because the breakpoint of this patient is at the beginning of the large intron 10 repeat region within which no primers could be designed and therefore a second PCR with this primer alone (FIP Int 10 22850 F) was also run for all patients to ensure that any breakpoints in this region would be detected.

Of the 202 patients tested, just one (E2860) was found to be positive despite being negative on the nested RT-PCR screen. Positivity in both E2826 and E2860 was confirmed by obtaining repeat specimens, both of which were positive by nested but not single step RT-PCR. Sequencing of the amplicons revealed typical mRNA fusions between *FIP1L1* exon

13 for both patients and part of *PDGFRA* exon 12. Despite the fact that both initial cDNAs had passed the quality control procedure (qualitative amplification of normal *BCR* mRNA), the discrepant results were probably due to relatively poor quality cDNA. Indeed, quantitation of normal *ABL* transcripts by RQ-PCR revealed a level of only $1-2 \times 10^3$ transcripts per 2 μ l of cDNA, compared to a level of $\geq 10^4$ (sensitivity $\geq 0.01\%$) that is generally considered as suitable quality for MRD analysis (299).

4.3.2.5 Overall comparison of RT-PCR and multiplex LR-PCR

Upon comparison of both of the diagnostic techniques, the gDNA multiplex LR-PCR gave a more sensitive detection of the *FIP1L1-PDGFRA* fusion with all the samples with adequate DNA quality were positive at single step, whereas only 50% were positive by single step amplification from cDNA (the remaining 50% required nested RT-PCR, see Table 4-3).

Analysis of patients positive by RT-PCR (n=44)	
Method of analysis	Single step positive
gDNA (LR-PCR)	44 (100%)
cDNA (RT-PCR)	22 (50%)

Table 4-3: Comparison of the numbers of *FIP1L1-PDGFRA* patients positive by single step gDNA (LR-PCR) or cDNA (RT-PCR).

4.3.3 Genomic DNA nested PCR for monitoring MRD

Next, the utility of genomic DNA analysis for detection of MRD was examined in patients undergoing treatment with imatinib. Initially nested gDNA PCR was employed: once the genomic breakpoints were identified, specific *FIP1L1* nested forward primers were designed and used with generic nested *PDGFRA* intron 12 reverse primers and to amplify short regions spanning the breakpoint. These primers were firstly tested on the presentation sample and a normal control to confirm that they worked specifically. Then the sensitivity was tested to confirm that they would pick up very small numbers of the fusion by performing a serial dilution of the patient presentation sample using 1 in 10 to 1 in 10^6

dilutions of patient DNA in *FIP1L1-PDGFR*A negative normal control DNA. From Figure 4-16 it is clear that the single step PCR for E359 is sensitive to 1 in 10^2 and nested to 1 in 10^4 using 500 ng of input DNA. Prior to testing MRD samples, each patient breakpoint-specific PCR primers were tested for sensitivity on presentation samples and all 7 tested amplified nested bands to at least 1 in 10^4 dilution. Theoretically, 1 μ g of genomic DNA is equivalent to 3.3×10^5 haploid genomes (297, 298) therefore the genomic MRD reaches a sensitivity that is very close to the maximum threshold.

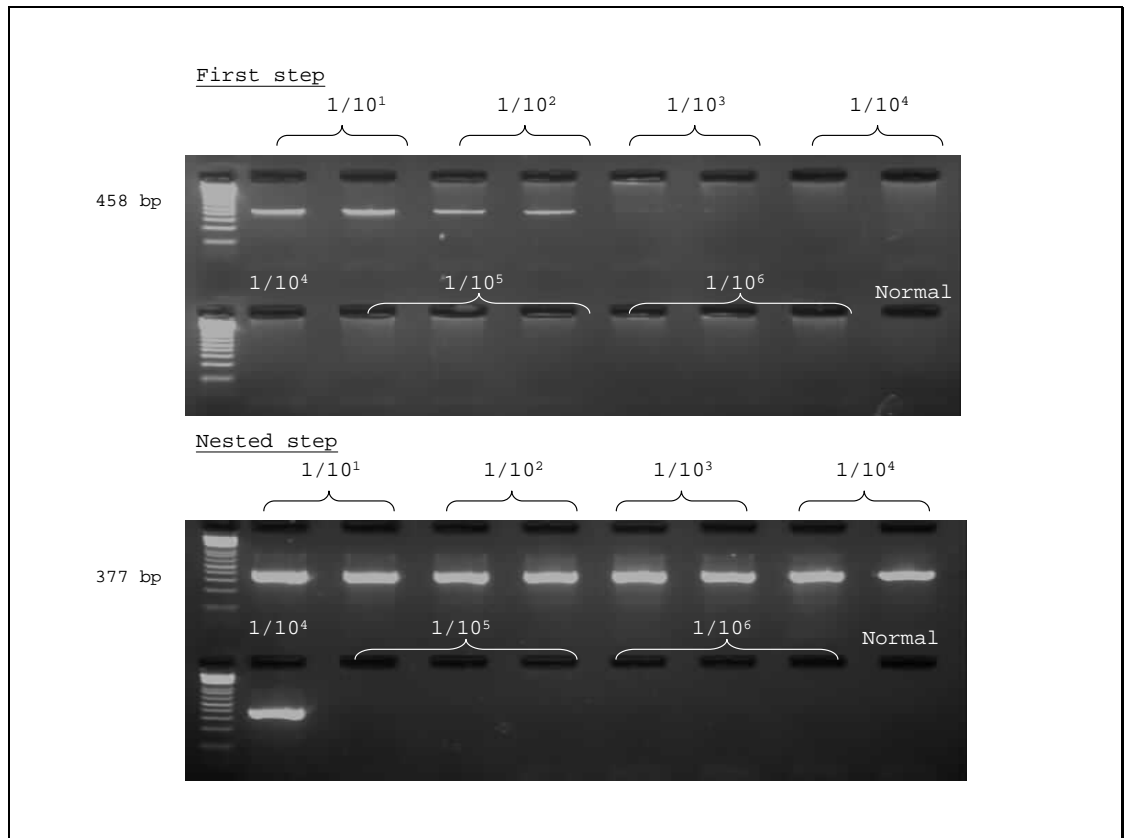


Figure 4-16: Agarose gels gDNA nested PCR for detection of MRD in E359. Serial dilution of E359 genomic DNA in normal control DNA using breakpoint specific primers FIP Exon 11 (36955) F1 and PDA I 12 R3 for the first step to yield a 458bp product and FIP Exon 11 (37004) F2 and PDA I 12 R4 for the nested step to yield a 377bp product.

All follow-up samples were tested for the presence of the *FIP1L1-PDGFR*A fusion gene using nested primers specific for each breakpoint using as close to 500ng of DNA in the first step PCR as possible in order to detect very low copies of the fusion. As shown in Figure 4-17, there are 31 data points comparing RT-PCR with gDNA MRD PCR. Of these, 20 were concordant, however 10 data points were gDNA positive, RT-PCR negative

whereas only one (E243) was gDNA negative, RT-PCR positive. This indicates that the gDNA MRD PCR is more sensitive than RT-PCR at detecting the aberrant fusion. One case that highlights the sensitivity of the genomic DNA MRD and its relevance to the clinical course of the disease is E370. This patient was non-compliant after about 12 months of treatment (when the cDNA result was found to be nested negative). He subsequently relapsed and was found, retrospectively, to be single step genomic DNA positive at this same time point showing that the genomic DNA MRD is not only more sensitive in some cases to the RT-PCR but more importantly relates to the clinical course of the disease. It is therefore clear that using genomic DNA breakpoint specific primers provides a more sensitive test for MRD in the majority of cases when compared to the conventional RT-PCR results.

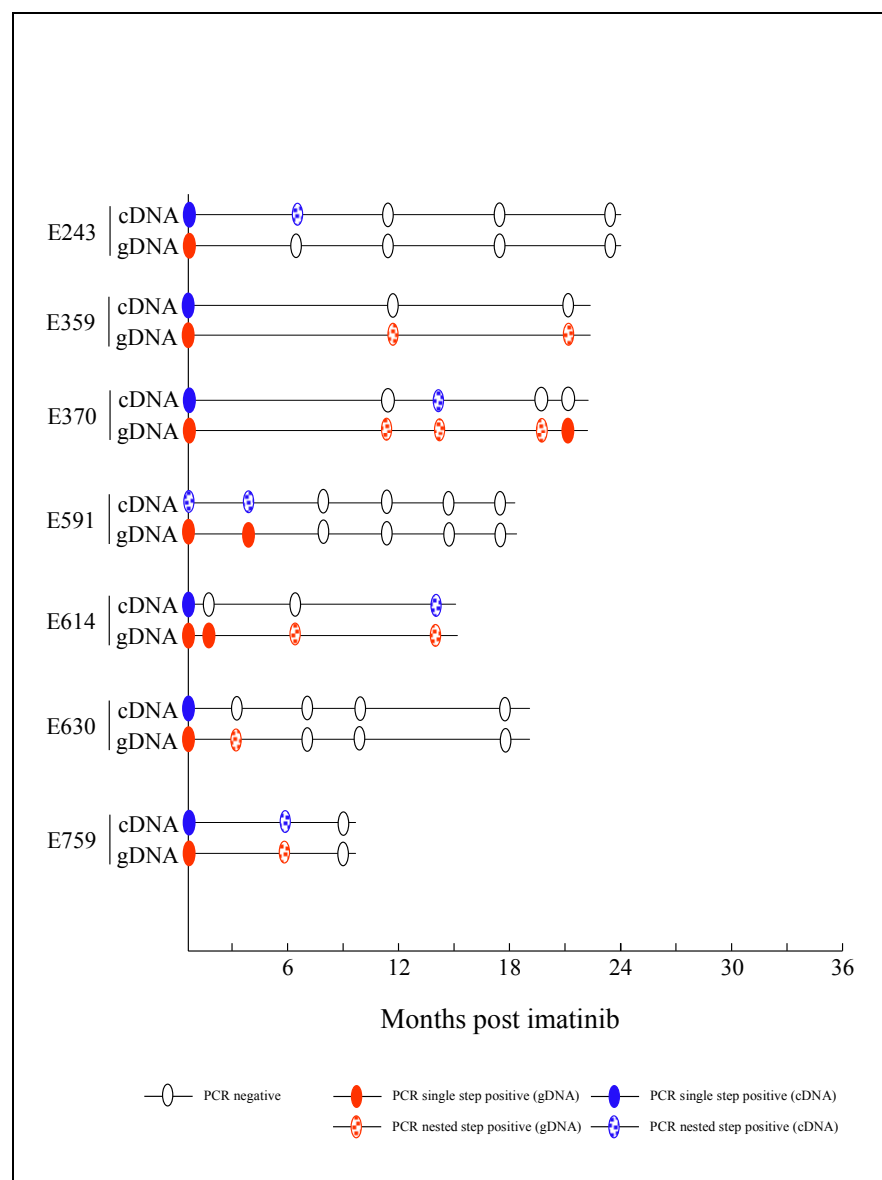


Figure 4-17: Comparison of cDNA nested RT-PCR and gDNA breakpoint specific PCR for MRD analysis of *FIP1L1-PDGFRα* positive patients.

As the gDNA nested PCR gave better sensitivity than the cDNA nested RT-PCR, a RQ-PCR method was developed to fully quantify and compare the levels of the gDNA fusion and to determine if gDNA RQ-PCR provides a better technique for monitoring MRD.

4.3.4 Genomic DNA real time quantitative PCR (RQ-PCR) for MRD detection

Due to the small region in *PDGFRα* in which a probe would be suitable i.e. close enough to the WW domain to allow efficient RQ-PCR of all breakpoints but not in the WW

domain which would exclude some patients being tested, designing of a generic *PDGFRA* probe was difficult. Initially, PrimerDesign (Southampton) designed a PerfectProbeTM, which has complementary ends to keep the reporter and quencher in close proximity and therefore the background fluorescence low in order to increase sensitivity, however this probe did not conform to EAC criteria with an R^2 value of at least 0.98 (295), give good standard curves with an efficiency as close to 1 as possible (usually between 0.95 and 1.05) or give high enough sensitivity (see Figure 4-18).

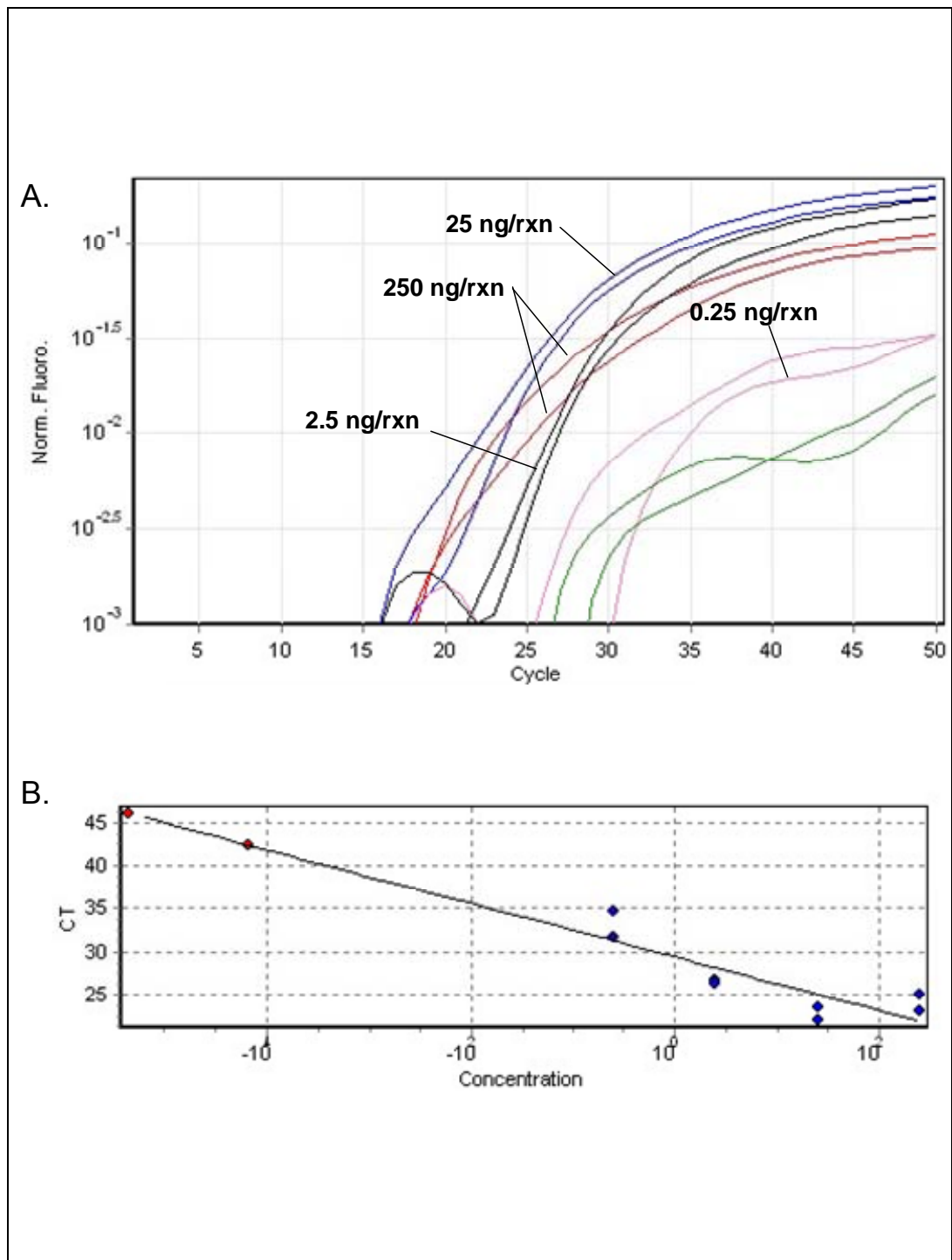


Figure 4-18: RQ-PCR using the PrimerDesign probe. (A) RQ-PCR graph showing log fluorescence against cycle number. A dilution series of normal control DNA in water was run with the *PDGFRA* gDNA PerfectProbe with the *PDGFRA* normal forward and reverse primers. DNA amounts are indicated (lines showing no amplification are the no template controls) and show a maximum sensitivity of only 0.25 ng/reaction, equivalent to 82.5 haploid genomes. (B) Standard curve showing the cycle threshold data against concentration of DNA, r^2 value of 0.7124 and amplification efficiency of 1.10. The r^2 value should be at least 0.98 and according to the EAC criteria, there should be an amplification efficiency of close to 1 (usually between 0.95 and 1.05).

PDGFRA primers and probe were then designed using Molecular Beacon software by Sigma (Gillingham, UK). To increase the specificity of the *PDGFRA* probe, Locked Nucleic Acids (LNATM) were incorporated into its structure (300) which enabled it to be designed just 3' of the WW domain and therefore suitable for use on all the *FIPIL1*-*PDGFRA* breakpoints (see figure 4-19).



Figure 4-19: Schematic showing positions of cDNA and gDNA RQ-PCR probes in *PDGFRA*. The cDNA probe lies partially over the breakpoint region in *PDGFRA* i.e. 5' of the red TGG (tryptophan residue) and therefore excludes analysis of some patients with this technique. The gDNA probe lies just within intron 12 and can be used to analyse all patients.

The probe was initially tested on serial dilutions of normal control DNA with the *PDGFRA* forward and reverse primers to ensure high sensitivity and conformation to EAC criteria (295), with a delta reaction of at least 1 in diagnostic samples and an efficiency of close to 100% (see Figure 4-20).

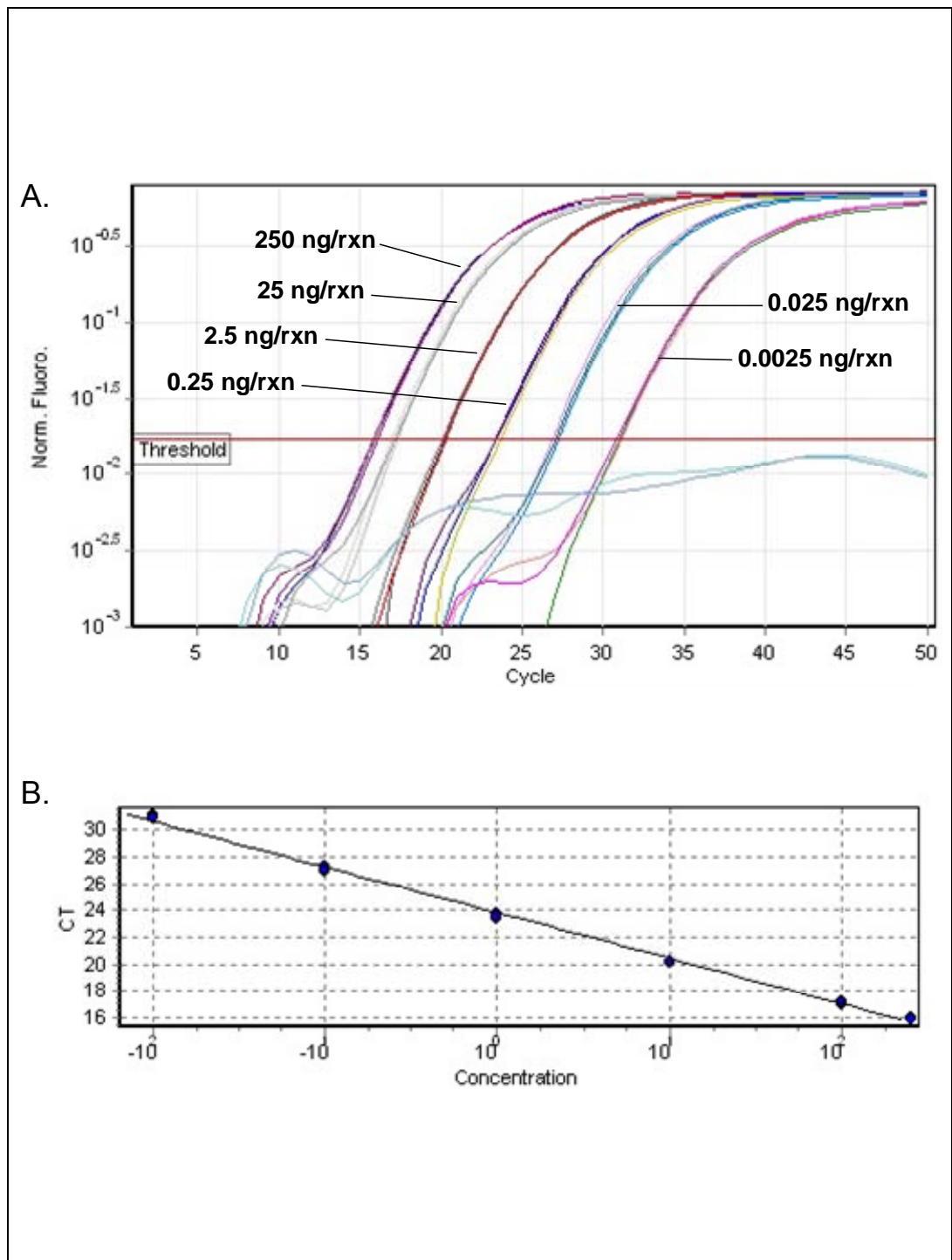


Figure 4-20: RQ-PCR using the *PDGFRA* LNA probe and primers on normal DNA serial dilutions. (A) RQ-PCR graph showing log fluorescence against cycle number. A dilution series of normal control DNA in water was run with the *PDGFRA* gDNA LNA probe with *PDGFRA* normal forward and reverse primers. DNA amounts are indicated (lines showing no amplification are the no template controls). (B) Standard curve showing the cycle threshold data against concentration of DNA, r^2 value of 0.9971 and amplification efficiency of 0.97.

The patient specific *FIP1L1* forward primers were then designed on Primer3 with the same annealing temperature as the *PDGFRA* reverse primer. The gDNA RQ-PCR was tested on *FIP1L1-PDGFRA* positive presentation samples, normal controls and no template controls to ensure patient specific amplification of breakpoints (see Figure 4-21 A). The probe was also tested on serial dilutions of 3 patient presentation samples in normal control DNA with the appropriate *FIP1L1* forward primers to ensure that changing the forward primer did not affect the assays performance (see Figure 4-21 B and C).

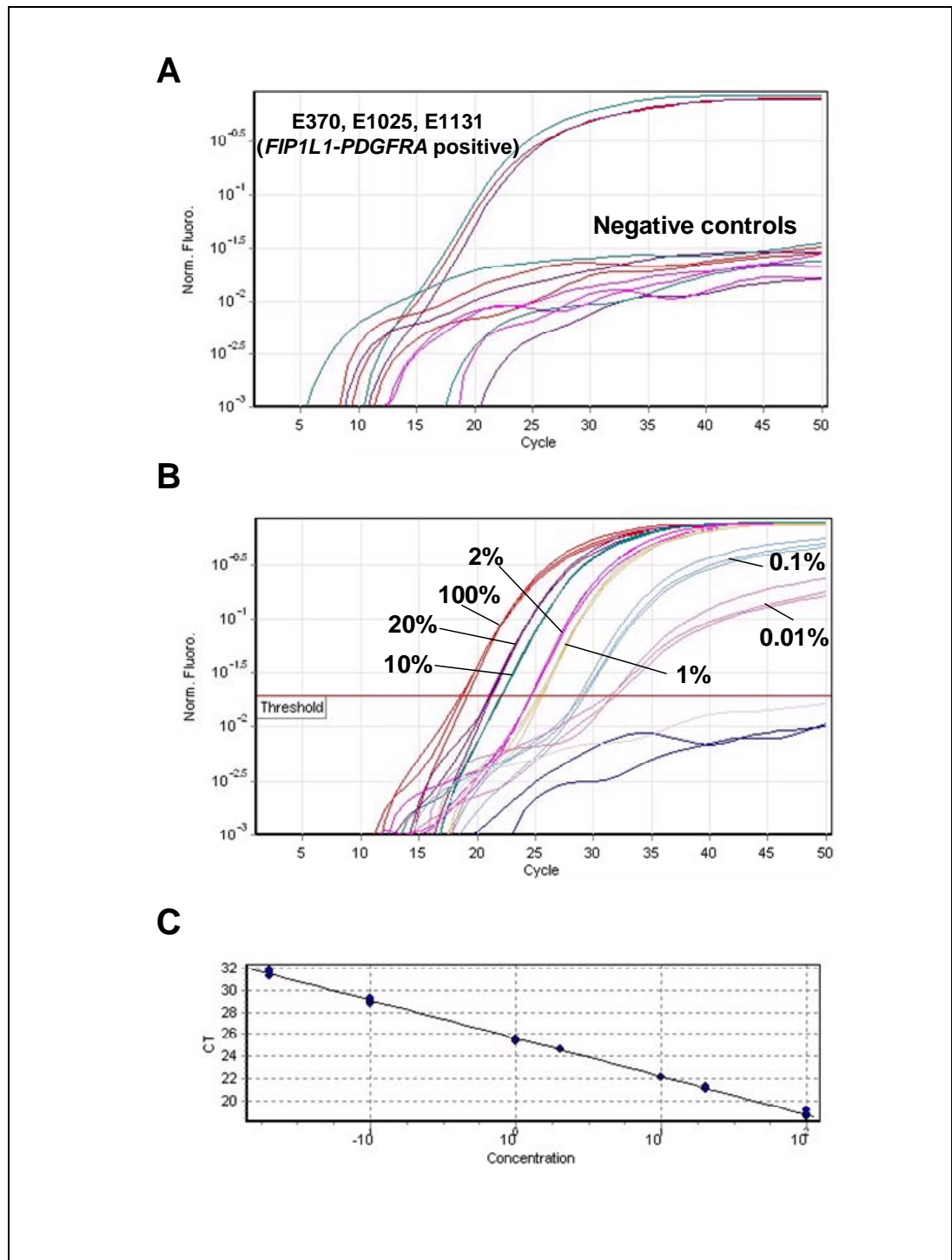


Figure 4-21: *FIP1L1-PDGfra* specific RQ-PCR using the *PDGfra* LNA probe. (A) RQ-PCR graph showing fluorescence against number of cycles. The three lines showing increasing fluorescence correspond to DNA from three *FIP1L1-PDGfra* positive cases: E1025 (green line), E1131 (red line) and E370 (purple line). No amplification was seen from the either the normal control DNAs or the no template controls (flat lines). (B) RQ-PCR graph showing a dilution series of E606 DNA in normal control DNA as indicated with the calculated cycle threshold line marked

(lines showing no amplification are the normal control DNAs and the no template control). (C) E606 standard curve showing the cycle threshold data against concentration of DNA, r^2 value of 0.998 and amplification efficiency of 95%. A similar analysis was performed for each case.

4.3.4.1 Comparison of gDNA and cDNA RQ-PCR and cDNA RT-PCR to assess the sensitivity of detecting MRD

In total, 66 samples from 13 patients who underwent imatinib treatment (median dose = 100 mg/day, median follow-up = 25 months; range 9-58) were analysed by gDNA RQ-PCR and initially compared to the results obtained by qualitative single or nested step RT-PCR. Comparison of gDNA RQ-PCR and nested RT-PCR revealed large discrepancies with 49% (31/63) samples positive by gDNA RQ-PCR but only 31% (19/61) were positive by nested RT-PCR, again highlighting the inaccuracies of the standard RT-PCR technique.

Of the 13 patients monitored for MRD using gDNA RQ-PCR, 9 were amenable to cDNA RQ-PCR analysis due to the position of the probe at the 3' end of the WW-domain which excludes approximately 25% of all breakpoints (195) and 7 of these had adequate number of sequential cDNA samples available for analysis. Of the 31 post-treatment data points available for comparison, 4 (13%) were discordant (gDNA positive and cDNA negative; see Figure 4-22 C to E). In contrast there were no instances where cDNA was positive and gDNA was negative. The gDNA RQ-PCR also showed a higher sensitivity for the negative results giving 1-2 logs better than analysis of cDNA by RQ-PCR (see Figures 4-21 and 4-22).

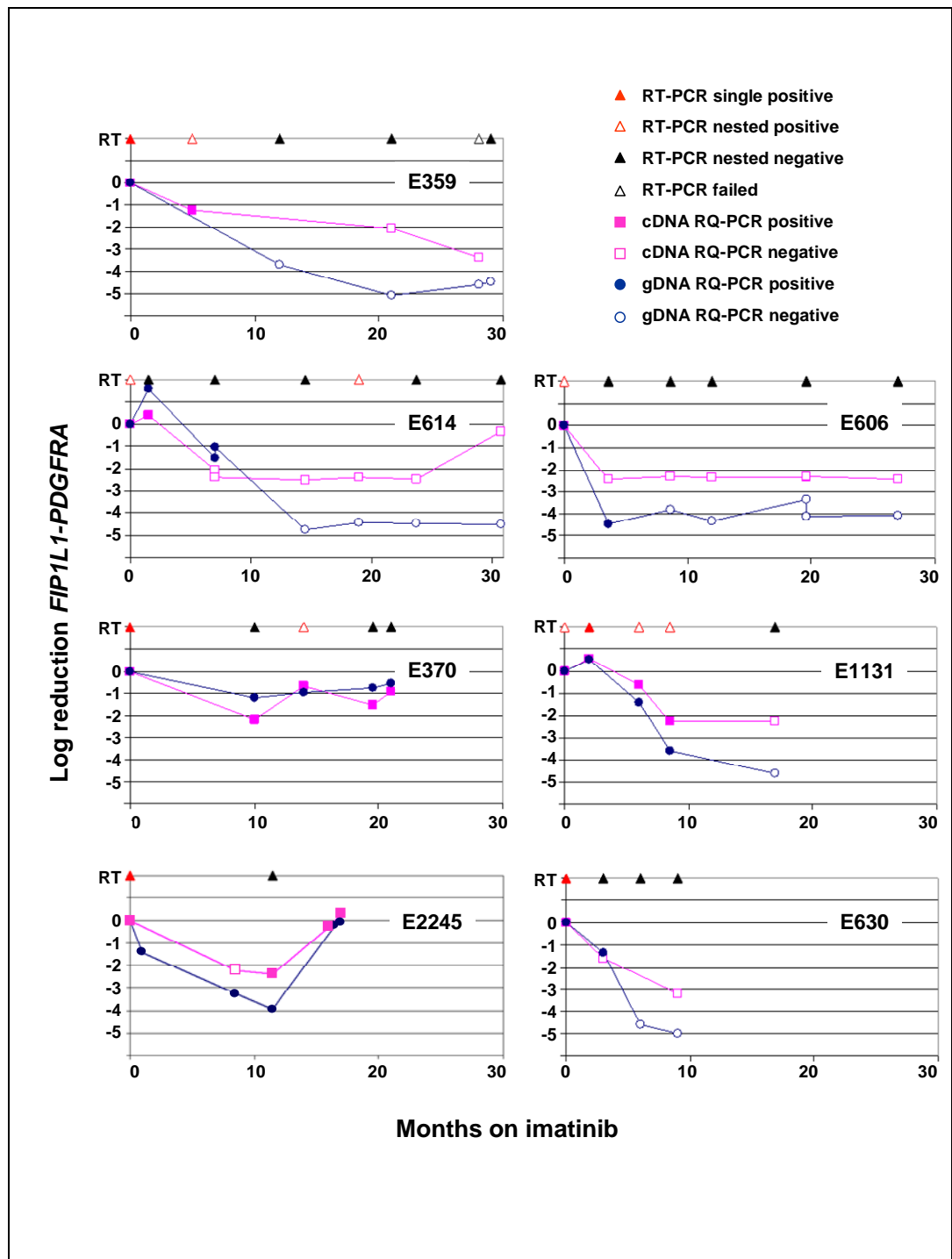


Figure 4-22: Graphs showing log reduction in *FIP1L1-PDGFRα* fusion against time of imatinib initiation in months. Three different techniques of measuring the amounts of fusions were compared; RT-PCR (triangles), cDNA RQ-PCR (pink lines and squares) and gDNA RQ-PCR (blue lines and circles) between 7 *FIP1L1-PDGFRα* positive patients and their follow-up samples.

4.3.4.2 gDNA RQ-PCR used for monitoring MRD

As the gDNA RQ-PCR showed greater sensitivity than both cDNA methods, MRD assessment using gDNA RQ-PCR showed that 11 out of 13 patients achieved a complete molecular response to imatinib within a median of 9 months (range (3-17)) of starting treatment (see Figure 4-23). As all 13 patients were tested by absolute RQ-PCR (see section 4.3.5) the sensitivity could be calculated and most could detect the fusion within a few copies of the plasmid, and as the maximum recommended concentration of 100 ng of gDNA was used for each MRD sample a sensitivity with which *FIP1L1-PDGFR*A could be excluded was approximately 1 in 30 000.

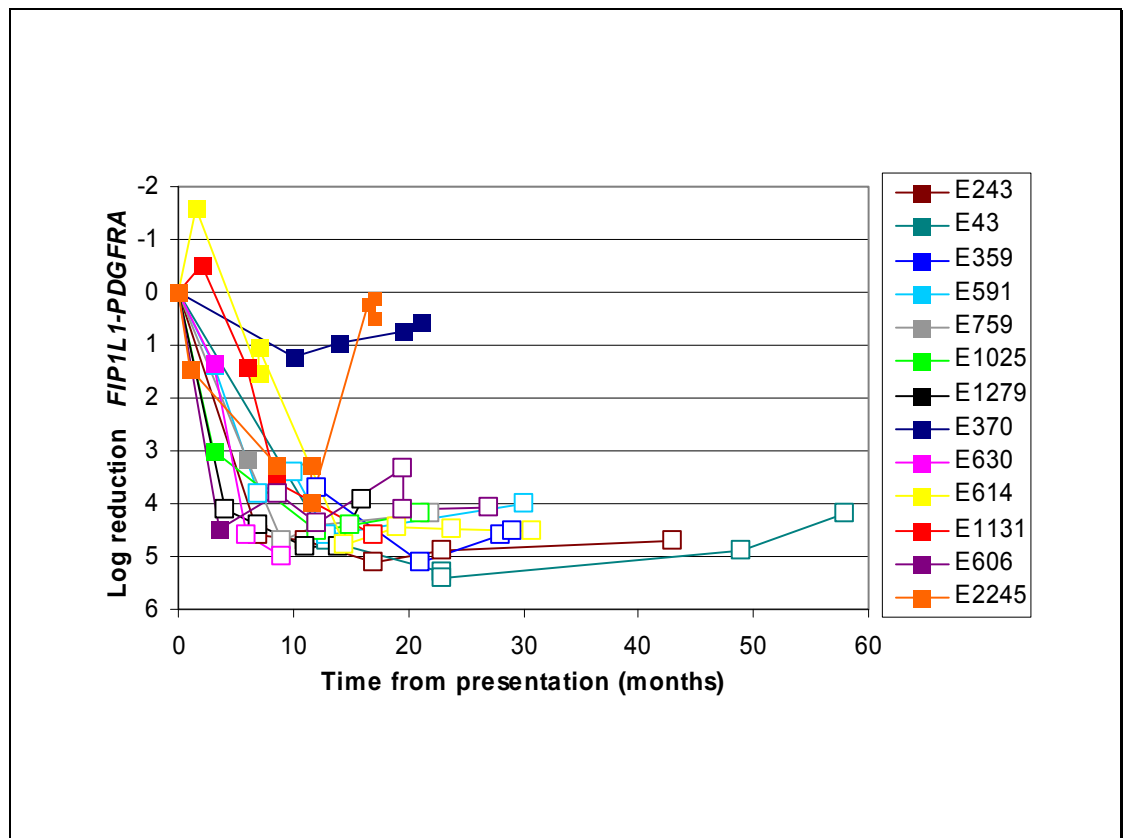


Figure 4-23: Graph showing log reduction in the levels of gDNA *FIP1L1-PDGFR*A fusion against time of imatinib initiation in months for 13 patients. The solid datapoints represent positive and empty datapoints represent negative results.

When the gDNA RQ-PCR results were compared to cDNA RQ-PCR results for Jovanovic *et al.*, (195) the median sensitivity of the gDNA was also 1 log greater than the cDNA (see Figure 4-24).

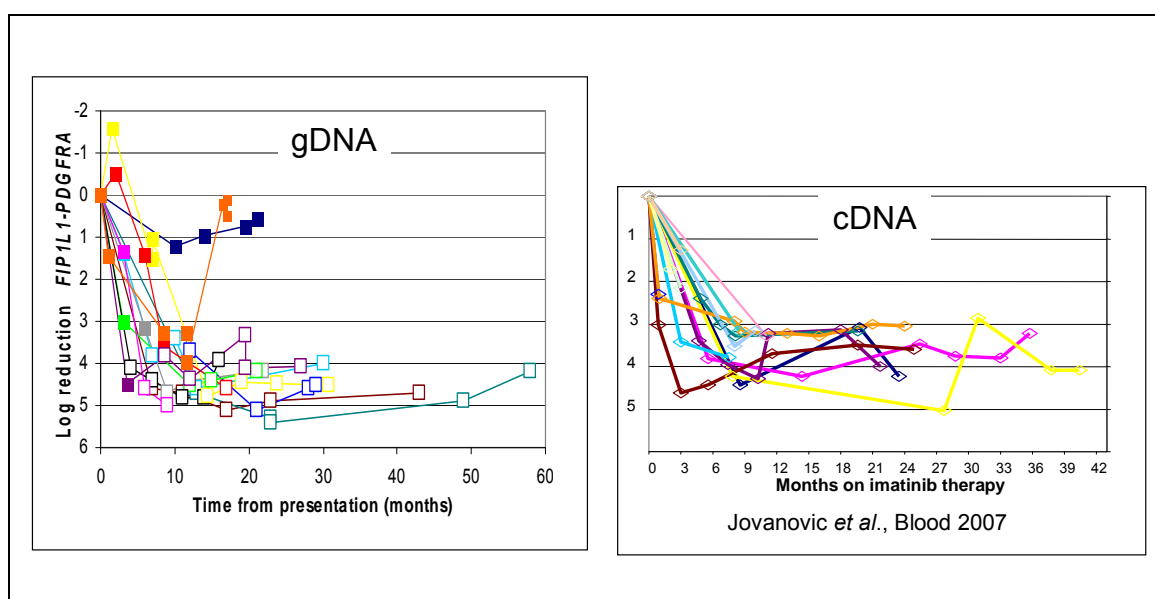


Figure 4-24: Graphs comparing the levels of gDNA *FIP1L1-PDGFRα* fusion to cDNA RQ-analysis performed by Jovanovic *et al.*, (195). The solid datapoints represent positive and empty datapoints represent negative results.

Two cases (E370 and E2245) failed to achieve a molecular response. Sequencing of the region encoding the *FIP1L1-PDGFRα* tyrosine kinase domain in the most recent samples revealed only normal sequence for E370, but E2245 had an A to C sequence change at position 57247 of *PDGFRα*, which is predicted to result in a D842V substitution. This sequence change was not detectable in the pre-imatinib sample.

4.3.5 Variability of *FIP1L1-PDGFRα* expression

As previous cDNA RQ-PCR studies have revealed considerable variation in *FIP1L1-PDGFRα* expression levels in patients prior to imatinib treatment (195), the levels of genomic DNA fusion were determined by absolute gDNA RQ-PCR to assess if the variable levels in expression may be attributed, at least in part, to variable clone size. All primer/probe combinations were specific to each patient and demonstrated linear results on analysis of dilution series of each fusion containing plasmid in normal control DNA (see Figure 4-25), with most achieving sensitivities down to at least 10 copies of the fusion.

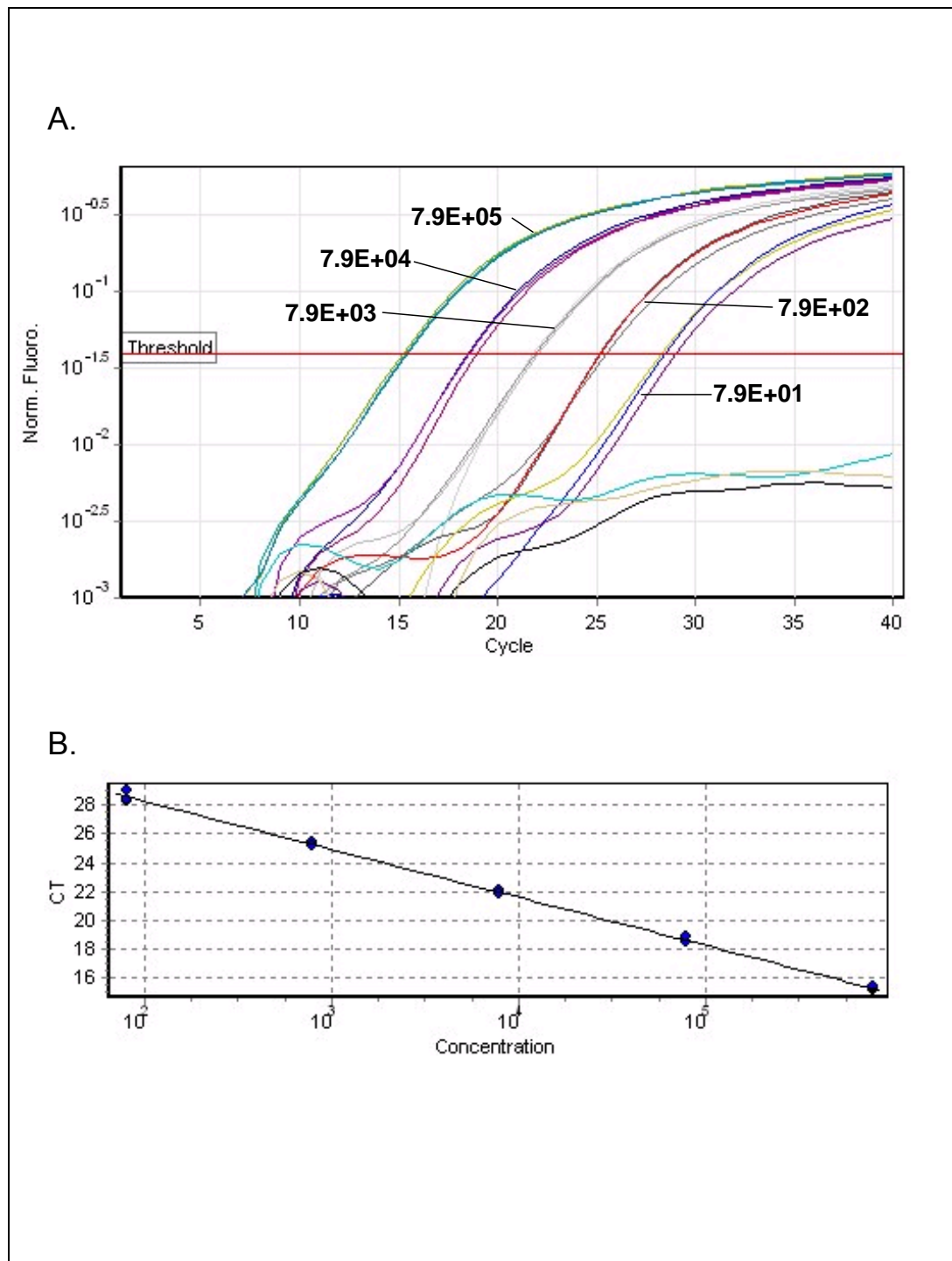


Figure 4-25: Absolute quantification using a *FIP1L1-PDGfRA* plasmid standard curve. (A) RQ-PCR graph showing log fluorescence against cycle number. A dilution series of plasmid DNA containing the gDNA breakpoint of E243 in normal control gDNA was run with the appropriate gDNA *FIP1L1* forward primer and the *PDGFRA* gDNA LNA probe and *PDGFRA* reverse primer. Copies of plasmid per reaction as indicated (lines showing no amplification are the no template controls). (B) Standard curve showing the cycle threshold data against concentration of DNA, r^2 value of 0.9983 and amplification efficiency of 1.00.

Absolute quantification of the fusion at presentation (n=13) by gDNA RQ-PCR revealed a median of 0.13 copies of *FIP1L1-PDGFR*A per haploid genome (equivalent to 13% fusion positive cells) with a 40-fold variation between patients (range 2.7%-100% fusion positive cells, see Figure 4-26).

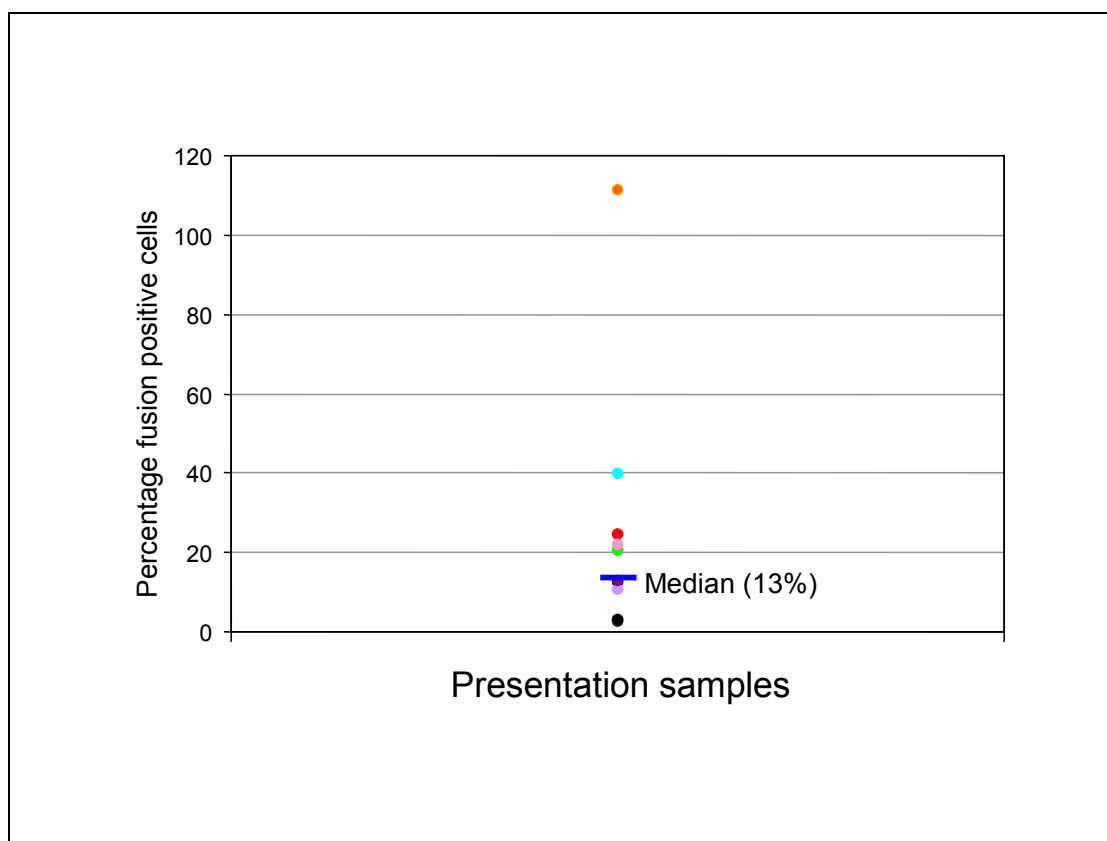


Figure 4-26: Graphical representation of percentage of cells harbouring the gDNA *FIP1L1-PDGFR*A fusion gene in cases at presentation.

4.3.6 Comparison of gDNA RQ-PCR and FISH

As the gDNA RQ-PCR results suggested that the clone size may be too small to be easily identified by FISH in some cases, *CHIC2* deletion FISH (as described in Materials and Methods section 2.2) was performed for comparison. Of the 13 cases analysed by gDNA RQ-PCR, 6 had contemporaneous peripheral blood-derived fixed cells suitable for interphase FISH. As shown in Table 4-4, the results between the two techniques were broadly concordant. The two patients with low level positive results of only 3% *FIP1L1-PDGFR*A cells estimated by gDNA RQ-PCR (cases E513 and E2002) had 6% and 4% positive cells respectively detectable by FISH compared to a background false positive rate

of 5% (289, 290). This suggests that FISH screening alone might miss some *FIP1L1-PDGFR*A positive cases and would have missed at least one and possibly two cases in this dataset.

E number	gDNA RQ-PCR (% fusion positive)	Interphase FISH (on PB) (% fusion positive)
359	Positive (20%)	Positive (41%)
513	Positive (3%)	Borderline (6%)
566	Positive (13%)	Positive (14%)
606	Positive (11%)	Positive (63%)
759	Positive (22%)	Positive (26%)
2002	Positive (3%)	Negative (4%)

Table 4-4: Comparison of the percentage of *FIP1L1-PDGFR*A positive cells estimated by gDNA RQ-PCR and FISH.

4.3.7 Structure of the *FIP1L1-PDGFR*A breakpoints

In total, 46 *FIP1L1-PDGFR*A breakpoints were identified during the course of this study. This data was contributed to a larger European LeukemiaNet study patients that aimed to classify the variability of *FIP1L1-PDGFR*A breakpoints and determine if there are any structural features associated with the breakpoints (Walz *et al*, submitted).

To facilitate the classification of the different types of *FIP1L1-PDGFR*A genomic breakpoints they were split into three groups, designated Type A, B and C depending on how the fusion was spliced. Type A breakpoints contain short pieces of inserted intronic *FIP1L1* sequence in the mRNA fusion that provides RNA acceptor sites to allow splicing between the end of the *FIP1L1* exon and the beginning of *PDGFR*A exon 12 in the absence of appropriate splice sites in the *PDGFR*A sequence. Type B breakpoints are the result of splicing of a complete *FIP1L1* exon to a cryptic splice site within *PDGFR*A exon 12 that lies downstream of the genomic *PDGFR*A breakpoint either near the beginning of *PDGFR*A or within *PDGFR*A exon 12. Type C breakpoints fuse part of an *FIP1L1* exon

directly to the truncated *PDGFRA* exon 12 creating an in-frame exonic fusion which is identical at both the cDNA and gDNA levels (see Figure 4-27).

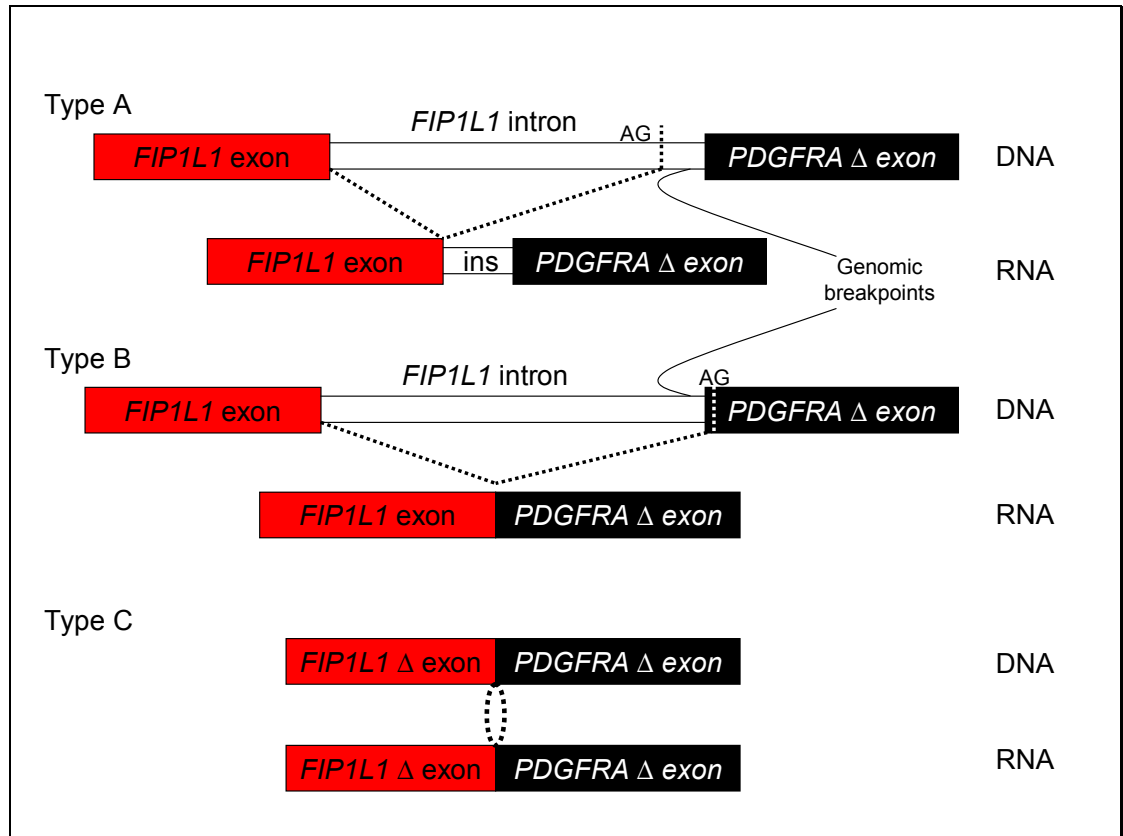


Figure 4-27: The three types of *FIP1L1*-*PDGFRA* splicing.

Of the 46 genomic DNA breakpoint sequences that were amplified, the sequence of the cDNA fusion (necessary to classify the breakpoint) was not determined in 8 cases. Of the 38 cases that could be fully classified 15 (39%) were type A, 19 (50%) were type B and 4 (11%) were type C. The positions of the breakpoints in *FIP1L1* varied considerably, occurring anywhere in the 65 kb region between intron 9 and intron 16. The genomic breakpoints in *PDGFRA* were all in exon 12 of the gene within the region encoding the WW-domain (see Table 4-5).

PDGFRA	
A.a. seq	-E--I--R--W--R--V--I--E--S--I--S--P--D--G--H--E--Y--I--Y--V--D--P--M--Q--L--P--Y--D--S--R--W--E--F--P
DNA seq	GAAATTCGCTGGAGGGTCATTGAATCAATCAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E1940/62	Intron 10
Type B	cttatttgtTGGAGGGTCATTGAATCAATCAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E1279	Intron 11
Type B	gatgttgctagGAGGGTCATTGAATCAATCAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E513	Intron 9
Type A	tctcagcatcacAGGGTCATTGAATCAATCAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E630	Intron 11
Type B	catgcatgtatttccGTCATTGAATCAATCAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E2465	Intron 11
Type A	cttttaaggtctatcGTCATTGAATCAATCAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E747	Intron 11
Type A	agagtggagacctgtcTCATTGAATCAATCAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E905	Exon 12
Type C	aggagattacctggggcaattgaatCAATCAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E2831	Intron 11
Type B	cagagtggtagtttcttagctttttAATCAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E1556	Intron 10
Type B	acottttgtgttactggtttaaatattattCAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E2588	Intron 10
Type B	attacagcattttgtatatatttggaatCAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E1803	Intron 11
Type B	tgcataatagcttttaacattattttatcAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E2316	Intron 10
Unknown	actcaggacttgctttatgattctgggtgcAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E1755	Intron 11
Type B	aaacacacacacacacacacacacacacagCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E2453	Intron 13
Unknown	gtttcatgctttttctgagtaataatagatataatagATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E939	Intron 11
Type B	TctaaaattatcatttgatgttttcatattctttaaattGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E2002	Intron 10
Type A	tgatagcacatctatatattaaatacttagtaccacagcaaGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E1424	Intron 12 (inversion)
Type B	agatccttgctgactaaagcaatcttttaacatacaatccctctttccaTATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E2378	Intron 12
Type B	tctcttgatactaagcaaaattgggatactattataattatattctgtgTATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E176	Intron 11
Type A	aaatatataaaagtatttagaacagtgttaccttgccatagtaggtgctccTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E370	Intron 10
Type A	tttctgtgggtgtcctttctgtttgttagtttcttcttaacagacaggaccctcATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E1336	Intron 11
Type A	gggaccacaagcatgtgctactatgcctagctgatattttaattattttagacATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC

PDGFRA	
A.a. seq	-E--I--R--W--R--V--I--E--S--I--S--P--D--G--H--E--Y--I--Y--V--D--P--M--Q--L--P--Y--D--S--R--W--E--F--P
DNA seq	GAAATTCGCTGGAGGGTCATTGAATCAATCAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E2716	Intron 12
Unknown	atttttaaaactgtatgtatgtgtatactcacacaggcgtatgtatacacatcggGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E614	Intron 16
Type A	ctattagtccttctcgtcaacacacttatgtgttaagttttaagttaaaggaaatTTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E1675	Exon 11
Type C	tctcagtcctcagacaagctactgcctccagaaaaagccaattcaagcgttgggaagtggcaGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E2842	Intron 13
Type A	tctataatatccagaaaatctgttattttcattttctcgtgaatcagaaaaatagtttccaGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E2076	Intron 13
Type B	ttttttctcattttactgaaacttttgatcctttgaaaatcctctccctacctcaccagACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E1131	Intron 12
Type B	cattattttattcctattttctgagtccttttaaaatgtccagagcactttattcttacctACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E2826	Intron 13
Unknown	tttacttataaagatttgattgaaataacatcttagatgaaatggagggtctcacaagaagggCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E43	Intron 13
Unknown	aatttaaaagcctgcctatcagtggttttcattaaattcaaaacagagagcaatgctaattatCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E1788	Intron 12
Type B	cacacatgtagaactattactctcccccaaaaatacatatatgtgaattaaactaccttttacCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E359	Intron 11
Type B	aatgaattatattgaaccgtgctaattggatgctcagtaatttttttttctgtctcttctgctcagGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E43	Intron 11
Unknown	ttcaattttaaaagcctgcctatcagtggttttcattaaattcaaaacagagagcaatgctaattatCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E606	Intron 14
Type A	caaggttcgtaagccctttgagactcagttttcctttttaccaaaagggttgatagaattgtgaaGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E1080	Intron 13
Type A	aaagtgtagtgtagattaactgttttttagccatattattctgtgtgattccatttaggtattgtGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
EOL1	Intron 12
Type B	ctattataattatatttctgtgtatactttttcccatctaaaattatcatttgatgttttcoatATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E344	Intron 13
Type A	gttggggagcgtacaggctatgtattttctatctatgtttccctacagtttttagtcaagtggcttgtATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E1961	Intron 13
Type A	ttaaatgtttttgaggactaaagtttgaaggggctgactagtaatttttttagtttttagcatgaatATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E635/636	Intron 10
Unknown	gataatatcctgcagagtattttgcaacttggttcattctccccgtcactttcagggtacaccaatATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E1401	Intron 12
Type B	atactaagcaaaattgggatactattataattatatttctgtgtatactttttcccatctaaaattaCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E2245	Intron 11
Type A	gtagtgtcagttgcatgtgtgtgtcttagaagggttaaggaaagtttcatottattattttgaaagtactCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E759	Intron 10
Type B	tcttttacttccatttgaacatcacaggattcttatttctatttctcttccotattccatattgcattctccttGCCTTATGACTCAAGATGGGAGTTTCC
E2860	Intron 13
Unknown	aataattacttcttcccaggagaaagctagtagtaaatgatgaaaaatagctgcttgctcctcatcagggaacaCTTATGACTCAAGATGGGAGTTTCC
E243	Exon 13
Type C	agcaaacaccctcgtttttccctccaggagctcctcccactcacttccactcctcattttcttccactcctccTTATGACTCAAGATGGGAGTTTCC

PDGFRA	
A.a. seq	-E--I--R--W--R--V--I--E--S--I--S--P--D--G--H--E--Y--I--Y--V--D--P--M--Q--L--P--Y--D--S--R--W--E--F--P
DNA seq	GAAATTCGCTGGAGGGTCATTGAATCAATCAGCCCAGATGGACATGAATATATTTATGTGGACCCGATGCAGCTGCCTTATGACTCAAGATGGGAGTTTCC
E566	Intron 13
Unknown	agattaactgttttttagccatattattctgtggtattccatttaggtattgtgaattattcagaatgcagacoccaTTATGACTCAAGATGGGAGTTTCC
E1025	Intron 16
Type B	caggtagtgccaatctgcattttgtatggatgctgtagaagttgctttgatgtctctctctttattttatactctacTGACTCAAGATGGGAGTTTCC
E591	Intron 15 Exon 16 seq of unknown origin
Type C	tttttttaaagtaccataacgtttgttttaaatcgtgtttttctttagttgcttttccccatcttctctggttctgtctctctctcAGATGGGAGTTTCC
E3011	Intron 16
Type A	gtgcaaaccacatgtacaacagtggtctcataagattacaatactgtttttactgtatcttttgttttagatacacaaaccattgtgttaaaaGAGTTTCC

Table 4-5: Genomic breakpoints of *FIP1L1-PDGFR*A patients. All breaks lie within the WW-domain of PDGFRA (W residues and codons are underlined). Lower case indicates *FIP1L1* introns in black and exons in red. Upper case in blue indicates *PDGFRA* exon 12 and splice acceptor sites are underlined where present. The pink nucleotide indicates a SNP found in all of the patients with a break in *PDGFRA* sufficiently 5'.

The European LeukaemiaNet dataset consisted of *FIP1L1-PDGFR*A junction sequences from 113 patients at the mRNA (n=113) and genomic DNA (n=85) levels. The analysis of this data was undertaken mostly by Christoph Walz (Mannheim), Andreas Reiter (Mannheim) and Joe Wiemels (San Francisco). Transcript Type A was seen in 50 (46%) of cases, Type B in 47 (43%) of cases and Type C in 12 (11%) of cases. It was shown that the location of genomic breakpoints within *PDGFRA* exon 12 and the availability of AG splice sites determine the transcript type and restrict the *FIP1L1* exons which are used for the creation of the fusion gene. In the majority of cases, short stretches of overlapping sequences were identified at the genomic junction site, suggesting that the *FIP1L1-PDGFR*A fusion is created by illegitimate non-homologous end-joining after DNA breakage. Analysis of breakpoint positions using scan statistics provided some evidence for clustering of breakpoints within *FIP1L1* that may be related to DNA or chromatin-related structural features. Significant clustering was evident for all breakpoints when considered together, and also for type B breakpoints considered in isolation.

4.3.8 Identification of a novel *PDGFRA* partner gene (292)

Because all *PDGFRA* breakpoints reported to date are so tightly clustered in *PDGFRA* exon 12, analysis of genomic DNA by bubble PCR may also provide a rapid means to

identify new *PDGFRA* fusions. Patients who were positive by the *PDGFRA* expression multiplex (see Figure 4-28) but negative for *FIP1L1-PDGFRA* were tested by bubble PCR to detect any novel fusions.

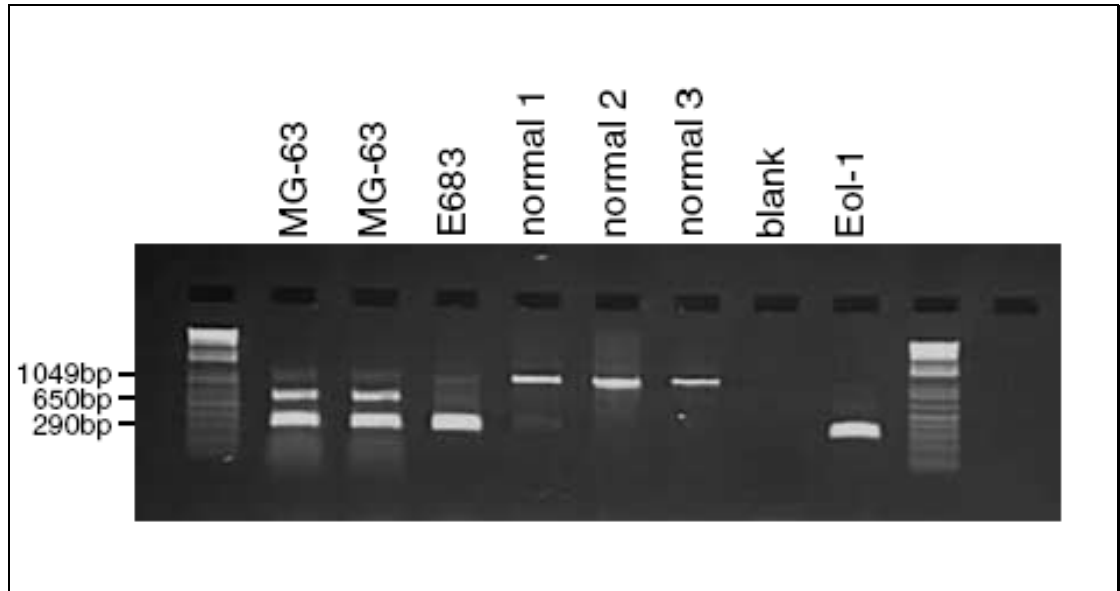


Figure 4-28: Gel showing *PDGFRA* over-expression results. MG-63 cells express normal *PDGFRA* and both the 650bp and 290bp bands are visible. *EOL1* cells express *FIP1L1-PDGFRA* but not normal *PDGFRA* and only the smaller 290bp band is visible. Normal individuals 1-3 express *PDGFRA* very weakly and the predominant amplification product (1094bp) is exons 13-15 from genomic DNA. Patient E683 shows the 290bp band but not the 650bp band suggesting the presence of a *PDGFRA* fusion gene.

The multiplex screen to detect over-expression of the *PDGFRA* kinase domain (and therefore potential *PDGFRA* fusion genes) was developed and performed by Claire Curtis (Salisbury). Samples from 229 patients referred for IHES or persistent unexplained eosinophilia were screened and all 12 *FIP1L1-PDGFRA* positive cases were also positive by the multiplex *PDGFRA* assay (292). Of the remaining 217 cases, 9 (4%) were also *PDGFRA* multiplex positive. Of these 9, DNA was available for 7 cases.

4.3.8.1 Results

4.3.8.1.1 Bubble PCR screening

All 7 cases were screened by bubble PCR. For 6 cases only normal bands were amplified, but for case E683 additional bands were amplified which upon sequencing revealed a fusion between *PDGFRA* exon 12 and exon 23 of kinesin family member 5B (*KIF5B*) at chromosome 10p11.22 (see Figure 4-29).

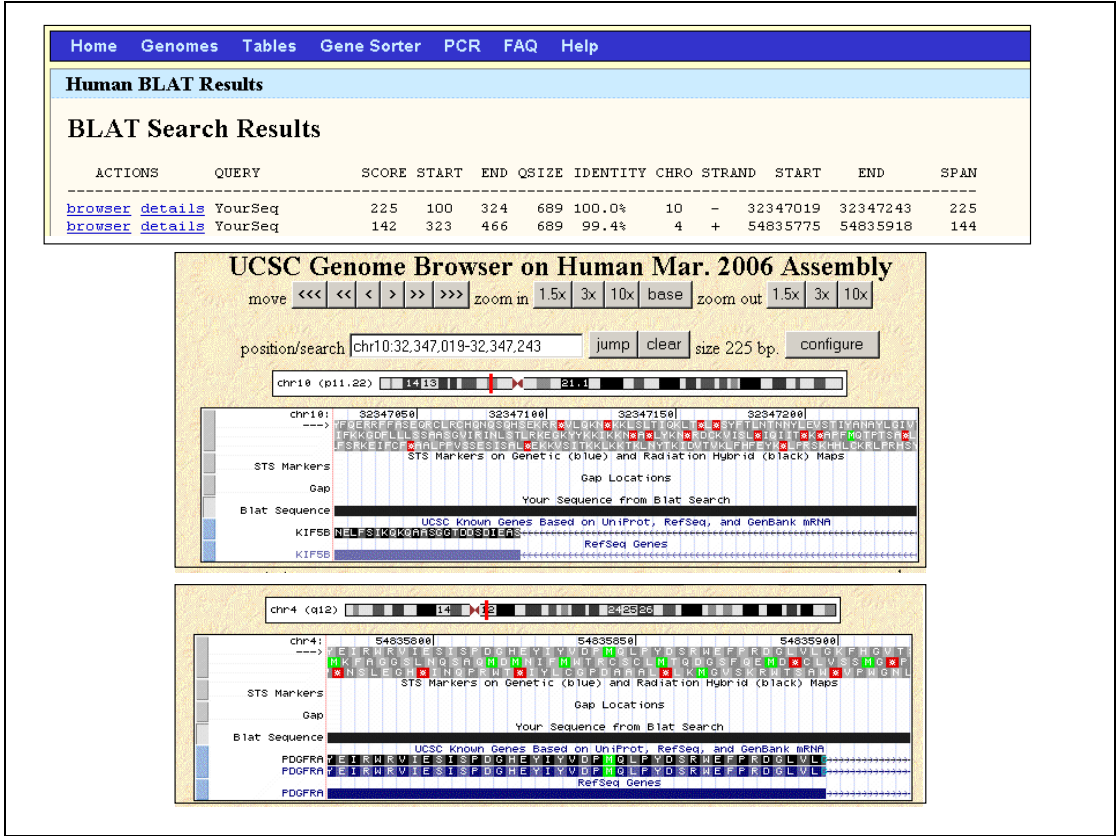


Figure 4-29: UCSC BLAT results on sequence obtained from bubble PCR from patient E683.

The KIF5B protein is composed of 3 structural domains; an N-terminal domain that hydrolyses ATP and binds microtubules, a large central α helical coiled coil domain and a C terminal domain that interacts with other proteins, vesicles and membranous organelles. The *KIF5B-PDGFRA* breakpoint is within exon 23, producing a fusion protein containing most of the first two structural domains of KIF5B, including 6 of the 7 coiled coil domains (see Figure 4-30).

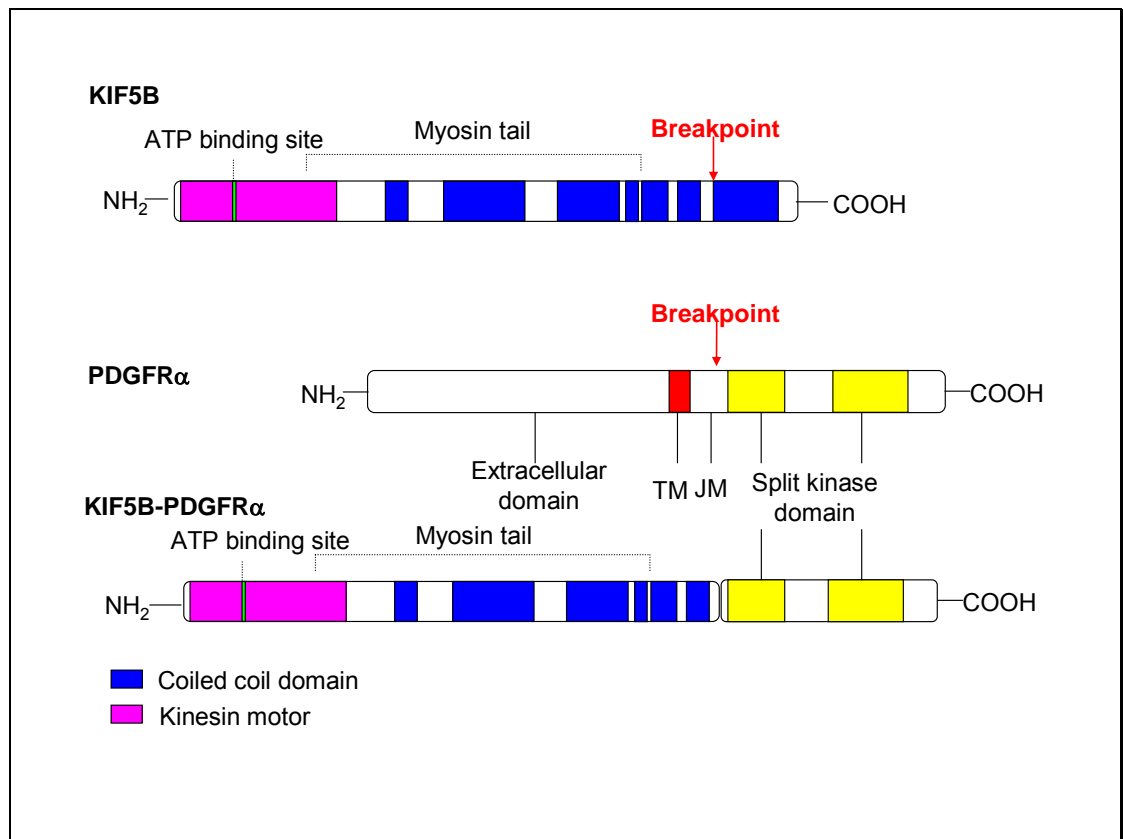


Figure 4-30: Domain structure of PDGFRα, KIF5B and KIF5B-PDGFRα.

4.3.8.1.2 Case report for E683

A 54-year old male (designated E683) presented with fatigue, night sweats, weight loss and headaches. Blood counts showed a leukocytosis of $31 \times 10^9/l$ with a marked eosinophilia of $26 \times 10^9/l$. The differential showed 6% neutrophils, 84% eosinophils, 0.5% basophils, 4% monocytes and 5.5% lymphocytes. Clinical examination showed a cardiac murmur but no thorax, abdomen or brain abnormalities. Parasitic investigations were carried out on stool samples but were found to be negative, as were blood tests for HIV 1 and 2 and TPHA. The bone marrow aspirate showed an increase of myeloid cells and a marked eosinophilia but an absence of blast cells. Immunophenotyping showed no monoclonal B or T lymphocyte populations. The patient was found to be *BCR-ABL* negative by RT-PCR, however cytogenetic analysis revealed 8 out of 16 metaphases with a complex karyotype; 46XY, del(3)(p21), add(4)(q12),-10,13q?,+der(?)(? \rightarrow cen \rightarrow ?::4q12 \rightarrow 4q28.3::10q11.2 \rightarrow 10qter). The patient was treated with prednisone then hydroxyurea but no improvement was seen. Following detection of the *KIF5B-PDGFRα* fusion, the patient was then treated with 100 mg of imatinib per day after the cessation of the other medication. Four days after the introduction of imatinib

total blood leukocytes returned to within normal parameters ($3.5 \times 10^9/l$) with the disappearance of eosinophilia ($0.105 \times 10^9/l$). A bone marrow aspirate was taken 6 months after the introduction of imatinib and all the 25 metaphases analysed had a normal karyotype.

4.3.8.1.3 Confirmation of the presence of *KIF5B-PDGFR*A

The presence of the fusion was confirmed in both genomic DNA and mRNA in both blood and bone marrow by single step PCR and RT-PCR respectively (see Figure 4-31 A). The reciprocal fusion was not detected. Upon sequencing the cDNA and genomic DNA breakpoints were found to be identical and the breakpoint within exon 23 of *KIF5B* (see Figure 4-31 B and C) creating an in-frame fusion. The breakpoint in exon 12 of *PDGFRA* is just outside the WW-domain (see Figure 4-31 B).

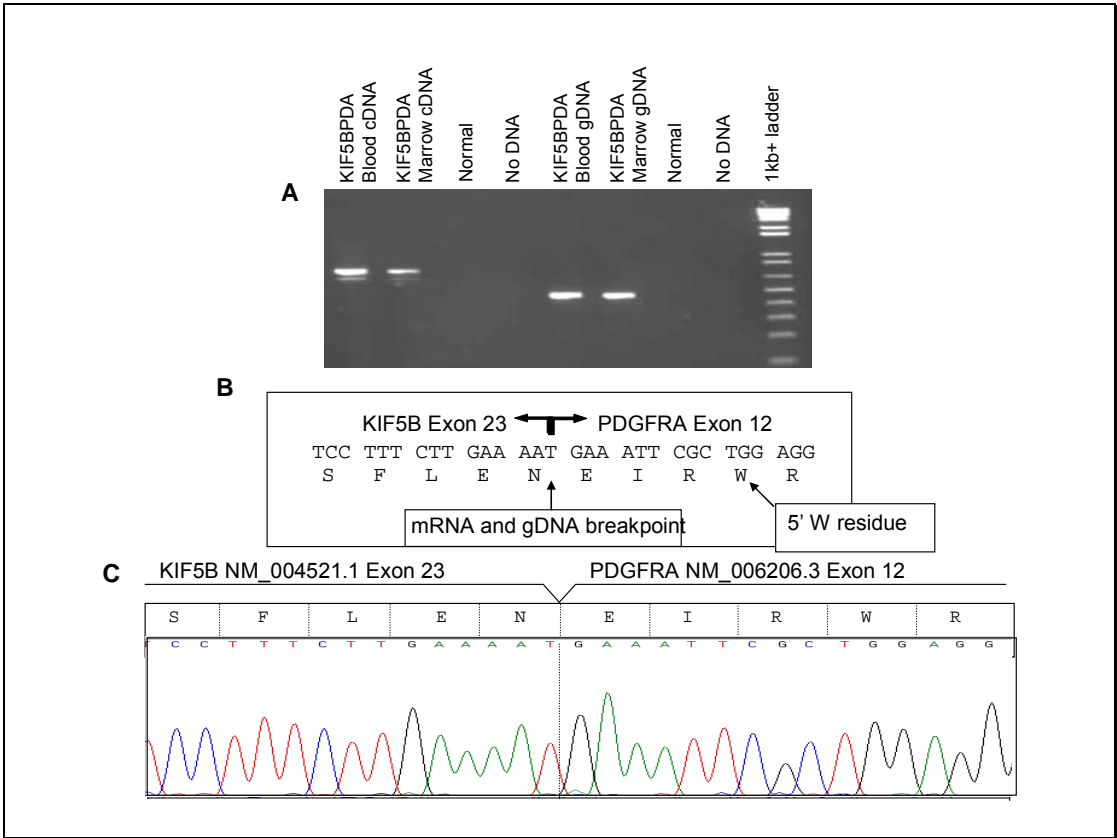


Figure 4-31: Detection of the *KIF5B-PDGFR*A fusion. (A) RT-PCR and gDNA PCR of *KIF5B-PDGFR*A in E683 and controls using primers KIF5B Exon 19 F1 + PDA Exon 14 R1 and Exon 22 F1 and PDAI12 R3 respectively. (B) Structure of the fusion mRNA/genomic DNA of the patient. The *PDGFRA* breakpoint is 5' of the tryptophan (W) residue that denotes the beginning of the WW-

domain. (C) Electropherogram from sequencing of genomic DNA *KIF5B*-*PDGFRA* breakpoint. The breakpoints in the patient are indicated.

Interphase FISH was also performed to confirm the presence of the fusion and for all results 200 interphase cells were scored. Using BACs RP11-460H18 and RP11-167O6, which are 5' and 3' respectively of *KIF5B* at 10p11.22, on normal interphase cells gave two sets of overlapping signals as expected. For patient E683, 66% of cells showed a normal pattern, but for 34% of cells one green signal was missing (see Figure 4-32) indicating a deletion downstream of *KIF5B*.

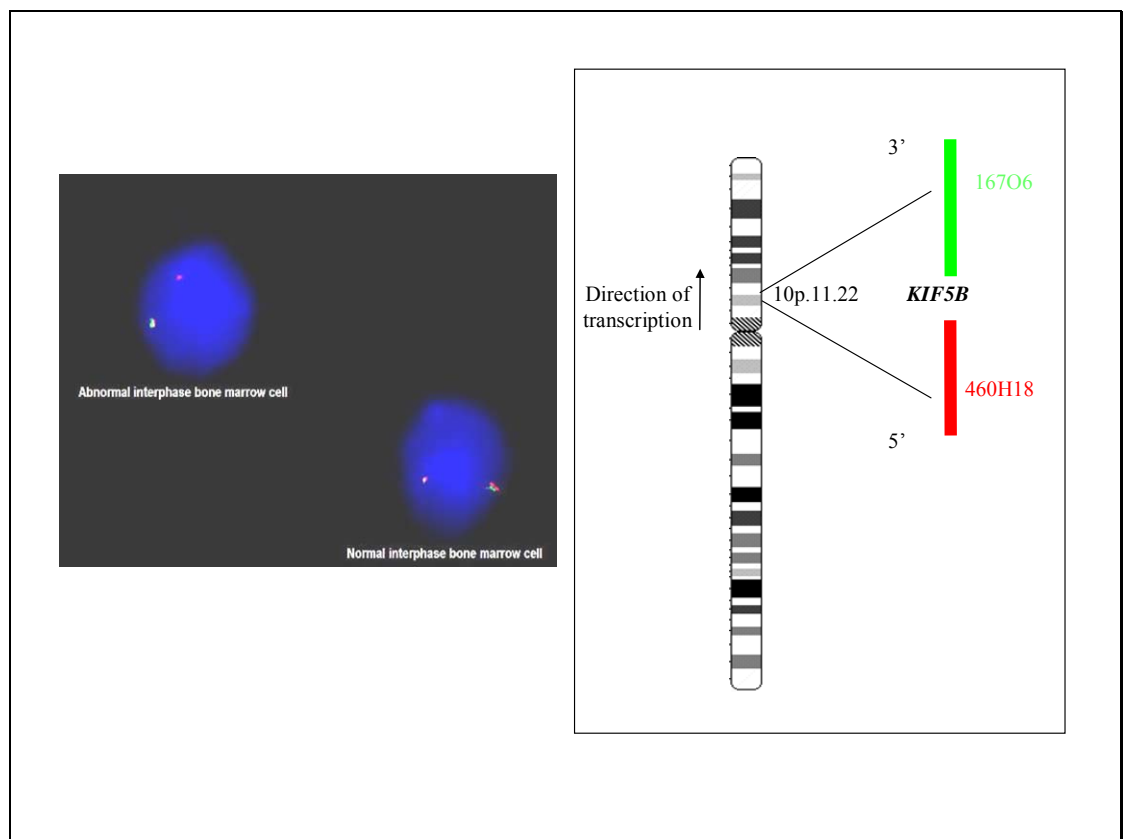


Figure 4-32: Flanking FISH analysis of E683 with differentially labeled flanking *KIF5B* BAC probes RP11-167O6 (green) and RP11-460H18 (red) demonstrating a fused signal and a lone red signal but no green signal.

Furthermore, the *KIF5B* flanking BAC RP11-460H18 in conjunction with the *PDGFRA* flanking BAC RP11-2O410 gave four separate signals in normal interphase cells, but

overlapping signals, consistent with fusion of *KIF5B* to *PDGFRA*, were seen in 41% of cells for patient E683 (see Figure 4-33).

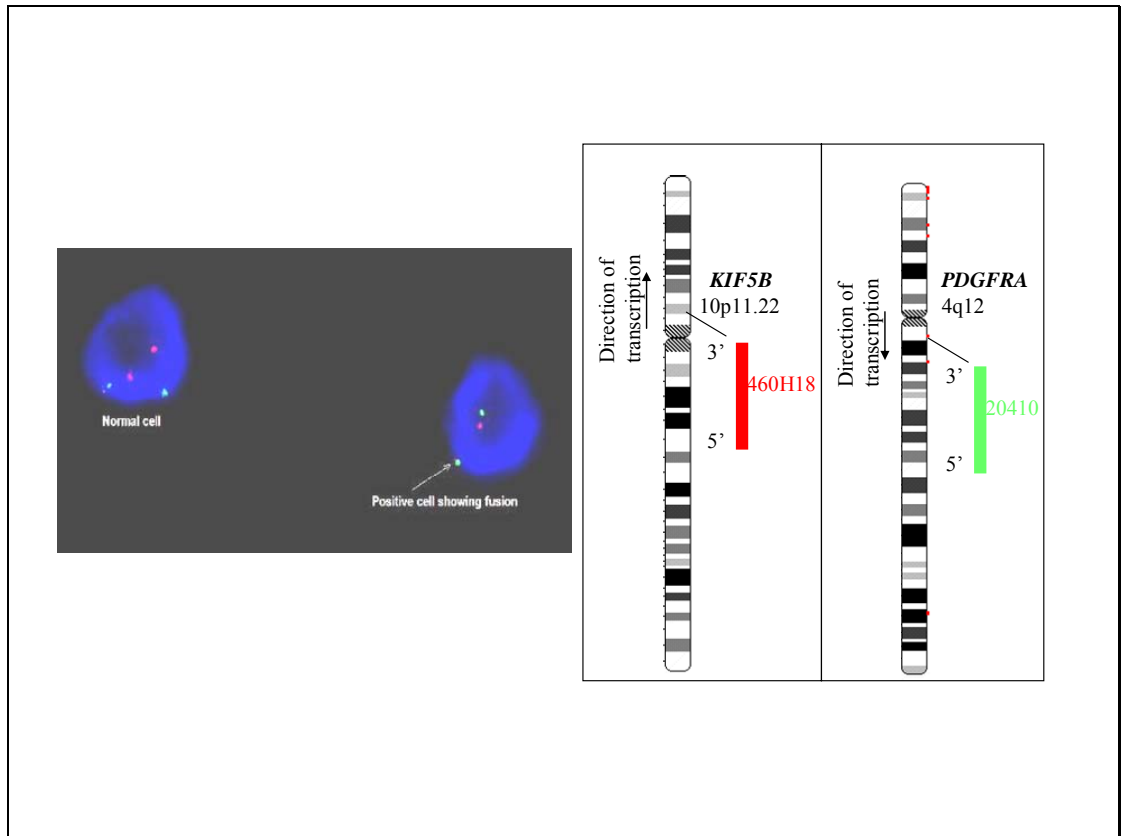


Figure 4-33: Split FISH analysis of E683, using differentially labelled BAC probes RP11-460H18, 5' of *KIF5B* and probe RP11-2O410 5' of *PDGFRA*. Abnormal cells show a co-localised red/green signal indicating fusion of these two regions flanking the *KIF5B*-*PDGFRA* breakpoint.

Unexpectedly, however, a split signal was not observed with the two BACs that flank *PDGFRA* (RP11-571I18 and RP11-2O410) suggesting that *KIF5B* has been inserted into the *PDGFRA* locus. These findings indicate a complex rearrangement, consistent with the reported karyotype (see section 4.3.8.1.2).

4.4 Discussion

4.4.1 *FIP1L1-PDGFR*A breakpoint identification

In contrast to other leukaemia-specific fusions, nested RT-PCR is routinely required to detect the *FIP1L1-PDGFR*A fusion in many pre-treatment cases (184, 196) even when good quality cDNA is used. This can lead to contamination and an increased risk of false positive results. Furthermore, breakpoint variability and low levels of expression of the fusion mean that some genuinely positive cases may be missed. Implementation of bubble PCR, simplex and multiplex LR-PCR enabled 46 *FIP1L1-PDGFR*A genomic DNA breakpoints to be identified (see Table 4-1) and confirmed as unique and patient-specific by sequencing (see Table 4-5). In addition, 20 further patients initially found to be nested RT-PCR positive were found to be false positive after screening with bubble and/or LR-PCR yielded negative results, therefore these genomic techniques have allowed resolution of ambiguities in detection of the fusion that arose from the use of nested RT-PCR.

Subsequent to the success of the multiplex LR-PCR to detect *BCR-ABL* breakpoints at single step in CML and ALL (see Chapter 3) and the success of the simplex LR-PCR to detect gDNA *FIP1L1-PDGFR*A breakpoints at single step, the gDNA multiplex was designed and optimised. The *FIP1L1-PDGFR*A gDNA LR-PCR multiplex picked up all of the positive controls (with adequate sample quality) at single-step amplification (see Figure 4-12) and was therefore implemented alongside the nested RT-PCR initially for confirming unique breakpoints in patients found to be positive by RT-PCR but is now used for screening all incoming patients with eosinophilia, alongside the nested RT-PCR.

Comparison of the nested RT-PCR and the gDNA LR-PCR revealed that all patients with adequate sample quality (46/48) were single-step positive by LR-PCR but only 50% (22/44) were single step positive by RT-PCR, the remaining 50% requiring nested amplification. Importantly, 2 patients that were nested negative by RT-PCR were single step positive by LR-PCR multiplex. The cDNA quality for these 2 cases was relatively poor as assessed by RQ-PCR for *ABL* mRNA and therefore suggests that a minimum cDNA quality criteria similar to those employed for MRD analysis (e.g. number of *ABL* transcripts $> 10^4$; (299)) should also apply to diagnostic RT-PCRs for *FIP1L1-PDGFR*A.

The gDNA multiplex LR PCR also easily detected *FIP1L1-PDGFR*A in two cases that were negative or borderline negative by FISH (see Table 4-4, E513 and E2002), however it is important to stress that the FISH analysis was performed on peripheral blood rather than bone marrow. In the six cases tested in-house, the median proportion of FISH positive cells was only 13% compared to a median of 80% in 10 cases for whom bone marrow was analysed (290). Although further comparative FISH data of blood and bone marrow are warranted, these results suggest that the absence of *CHIC2* deleted cells by interphase FISH should be interpreted with caution when peripheral blood is used for analysis. This data taken together shows that the gDNA multiplex is a more sensitive method of *FIP1L1-PDGFR*A detection when compared to both nested RT-PCR and FISH. However, currently the gDNA multiplex is used in conjunction with nested RT-PCR to screen for *FIP1L1-PDGFR*A fusions as there is a 12 kb repeat region in intron 10 within which no primers could be designed. There is a small chance therefore that the gDNA multiplex may not pick up breakpoints at the very 3' end of this region, currently however we have not detected such a case.

4.4.2 Location of breakpoints

The breakpoints in *FIP1L1* were distributed in the large 62 kb region between introns 9 to 16. Previously the most 3' break in *FIP1L1* that had been described was in exon 14 (184). Conversely, all the breakpoints in *PDGFR*A were within the 78 bp WW-domain (see Table 4-5) concurring with previous reports (184). There were found to be 3 different types of breakpoints depending on the splicing of the fusion. Type B was the most common, being seen in 50% (19/38) of cases with type A occurring in 39% (15/38) and type C the least common at 11% (4/38). There did not appear to be any particular clustering of the *FIP1L1* breakpoints in this dataset although the more detailed analysis performed by Walz *et al.*, (submitted) did indicate specific clusters, particularly for Type B breakpoints that may be related to specific chromatin features.

Within *PDGFR*A, all genomic breakpoints are tightly clustered within exon 12. The most likely reason for this clustering appears to be functional since breakpoints in this region specifically disrupt the auto-inhibitory WW-like domain between its two tryptophan residues (W559 and W586). Stover *et al.*, (301) have previously shown that this disruption of the WW-like domain is essential for the activation of the *PDGFR*A kinase by *FIP1L1*,

which, in contrast to other tyrosine kinase fusion partners, does not contribute a protein dimerisation domain. Breakpoints within the same region of exon 12 are also seen in most rare cases with variant *PDGFRA* fusions involving alternative partner genes. However, several questions remain, e.g. if a dimerisation domain is not required when the WW domain is disrupted, why does BCR disrupt the region when fused to *PDGFRα*? Also why are WW domain disrupting fusions not usually seen for *PDGFRβ* fusions given the fact that the WW domains are almost identical in *PDGFRα* and *PDGFRβ*. One possible explanation is that *PDGFRα* exon 12 is particularly susceptible to breakage, whereas the corresponding exon in *PDGFRβ* is not. Potentially this could relate to the fact that *PDGFRα* has a developmentally regulated internal promoter in intron 12 that may be associated with an open chromatin structure (302). To investigate this DNaseI hypersensitivity assays around *PDGFRα* and *PDGFRβ* exon 12 using myeloid cell lines could be performed.

4.4.3 Comparison of methods to monitor MRD

The higher sensitivity of the gDNA multiplex indicated that the use of patient-specific gDNA breakpoints might lead to increased sensitivity for detection of MRD. Initial testing by qualitative nested gDNA PCR on 7 patients revealed that this method was indeed more sensitive than nested RT-PCR in all but one datapoint (see Figure 4-17). RT-PCR is used routinely for the monitoring for MRD of other fusion genes such as *BCR-ABL* and is found to have a high sensitivity (255). The reason for the lower sensitivity in RT-PCR for the *FIP1L1-PDGFRα* fusion may be due to the fact that unlike other fusions, *FIP1L1* has many splice variants which might affect the efficiency of the nested RT-PCR but would not affect detection of the genomic DNA breakpoint.

For example, case E591 (see Table 4-5), who had an unusual clinical presentation and hence was difficult to diagnose, was found to be intermittently positive by RT-PCR on repeat pre-imatinib samples and therefore it was unclear whether the fusion was real or a false positive result. Bubble PCR was undertaken in an attempt to clarify the ambiguous results and the genomic breakpoint was found to be in exon 16 of *FIP1L1* (288). This may account for the intermittent results with the nested RT-PCR, as the RT-PCR product would

be approximately 300 bp longer than the most 3' breakpoint previously published (184). In pre-imatinib samples the fusion was amplified from gDNA by single step PCR and gDNA RQ-PCR but subsequent follow up samples have been nested gDNA and RQ-PCR negative. In addition, this case also had the *PDGFRA* breakpoint at the 3' end of the breakpoint region and this case was unsuitable for cDNA RQ-PCR.

As the gDNA nested PCR gave more sensitive results than nested RT-PCR in both detecting the fusion at presentation and for MRD, a RQ-PCR technique was designed to allow quantification and assess more accurately if the gDNA fusion was present at higher levels than the cDNA fusion. As expected, comparison of gDNA RQ-PCR with cDNA RQ-PCR and cDNA nested RT-PCR showed that the patient specific RQ-PCR (gDNA and cDNA) were much more sensitive than the nested RT-PCR (see Figure 4-22). Comparison of the gDNA and cDNA RQ-PCR showed that the gDNA RQ-PCR was more sensitive than the cDNA RQ-PCR with 13% (4/31) of the post-treatment datapoints discordant, all of which were gDNA positive and cDNA negative. There were no instances where the cDNA was positive and the gDNA was negative (see Figure 4-22). The gDNA RQ-PCR also showed a greater sensitivity of 1 log upon comparison to both the cDNA results from this study (see Figure 4-24) and also those obtained by others (195).

This lower sensitivity of the cDNA RQ-PCR compared to gDNA RQ-PCR is therefore not simply a result of decreased specificity due to the high levels of splice variants, as both RQ-PCR techniques use patient-specific primers. Other possible reasons for low and variable expression levels are discussed in section 4.4.4. It is important to note that the gDNA probe could be used for all breakpoints as it is located outside the breakpoint region of *PDGFRA* whereas the cDNA probe can only be utilised for about 75% of all breakpoints as the only feasible design was just within the breakpoint region (195). E591 is an example of a case for whom the mRNA breakpoint was not amenable for testing with the cDNA probe and therefore the gDNA RQ-PCR is the only available method to quantitatively measure MRD levels.

4.4.4 Absolute gDNA RQ-PCR of presentation samples

To ascertain if clone size was important for detection of the fusion by gDNA PCR techniques, absolute gDNA RQ-PCR was performed. This confirmed a considerable degree of variation in the proportion of *FIP1L1-PDGFR*A positive cells between patients (median 13%; range 3% to 100%). However this variation was much smaller (40-fold versus 1000-fold) than that previously described for *FIP1L1-PDGFR*A mRNA levels (195), suggesting that the difficulties in RT-PCR detection are only partly due to variable clone size. This is consistent with the observation that gDNA PCR was able to amplify *FIP1L1-PDGFR*A junction sequences in all cases, whereas only half were single step RT-PCR positive. The variability in clone size ascertained by gDNA analysis is concordant with the FISH data and with the fact that some FISH presentation samples are below the levels of detection (by FISH) but can still be amplified at single step by the gDNA multiplex or gDNA RQ-PCR. It seems likely that high levels of alternative splicing within the *FIP1L1* moiety of the fusion may lead to an apparent low level of fusion gene expression but other factors might play a role, for example polymorphic variation in the strength of the *FIP1L1* promoter or regions affecting stability of the mRNA fusion. Although the ability of gDNA analysis to detect the fusion might be affected by variable clone size, it should not be influenced by other mechanisms leading to variable expression.

4.4.5 Responses to imatinib

Absolute gDNA RQ-PCR showed that 11 of 13 cases achieved a complete gDNA molecular response to imatinib in a median of 9 months, confirming the excellent response of this disease to targeted therapy (187, 194, 303). The superior sensitivity of gDNA analysis may be particularly useful for novel approaches to disease management such as determination of optimal dosing for individual cases (304).

Resistance of *FIP1L1-PDGFR*A positive cases to imatinib is uncommon. Only two cases have been reported in the literature, both of whom developed T674I resistance mutations (184, 305). In the series studies here, two cases failed to achieve a complete molecular response, one of whom acquired a D842V mutation. This sequence variant has not been reported previously in *FIP1L1-PDGFR*A, but is the most common mutation seen in imatinib-resistant, *PDGFR*A-mutated gastrointestinal stromal tumours (306). D842V

corresponds to the D816V mutation seen in systemic mastocytosis, an activating mutation that is also refractory to inhibition with imatinib as it locks the catalytic domain into an active conformation which imatinib is unable to bind (307). The suboptimal response in the second case was believed to be a consequence of poor treatment compliance.

4.4.6 Identification of a novel *PDGFRA* fusion partner gene

Bubble PCR was used to identify a novel *PDGFRA* partner gene, *KIF5B*, in a patient with CEL. *KIF5B* is located on chromosome 10p11.22 (NM_004521.1) and encodes a member of a family of microtubule-based motor proteins involved in transport of organelles, which are highly expressed in neural cells (308). *In vivo* studies found that knockout mice (*kif5b* -/-) were embryonic lethal with severe growth retardation (309). The presence of the in-frame fusion was confirmed in mRNA and DNA (see Figure 4-31 A) and abnormal FISH patterns were also seen (See Figures 4-32 and 4-33). The fusion gene is unusual as it is identical in the cDNA and gDNA and within exon 23 of *KIF5B*. The breakpoint in *PDGFRA* is also unusual, as it is just outside of the WW-domain within which all the *FIP1L1*-*PDGFRA* breakpoints occur (see Figure 4-31 B). The KIF5B protein is composed of 3 structural domains; an N-terminal domain that hydrolyses ATP and binds microtubules, a large central α helical coiled coil domain and a C terminal domain that interacts with other proteins, vesicles and membranous organelles (see Figure 4-30). In the fusion protein, 6 out of the 7 coiled coil domains are retained, which could allow dimerisation and constitutive activation of the PDGFR α tyrosine kinase in a manner analogous to the BCR coiled coil domain in the BCR-ABL fusion (58). This was the third *PDGFRA* fusion partner gene to be reported after *FIP1L1* and *BCR* (183, 310). Furthermore this was only the second *PDGFRA* partner gene to be associated with CEL/IHES, highlighting the possibility of other fusion genes involved in the generation of this disease and indeed, now there have been 6 fusion partners for *PDGFRA* identified (168). Although the karyotype of the case was not known when the molecular analysis was performed, subsequent information revealed the presence of a complex rearrangement that included abnormalities of 4q12. This, in combination with the clinical phenotype, should have suggested the possible involvement of *PDGFRA*.

None of the other 6 cases that were found to specifically over-express the region encoding the PDGFR α kinase domain had *PDGFRA* rearrangements as determined by bubble PCR. Although it is possible that bubble might miss occasional rearrangements the ability of this technique to pick up most *FIP1L1-PDGFRA* rearrangements suggests this is unlikely. Or that the breakpoint in *PDGFRA* is not localised in exon 12. These cases may have been false positives, or possibly might result from some unexplained activation of the *PDGFRA* internal promoter (302). The function of this internal promoter is unknown, but conceivably it might be activated with pathogenic consequences in the *PDGFRA* multiplex expression positive cases.

Chapter 5: Conclusions

Genomic instability is a hallmark of malignant cells and chromosomal rearrangements are commonly associated with cancer (311). Acquired reciprocal translocations that produce oncogenic fusion genes are now known to be a common cause of haematological malignancies. The first translocation to be consistently associated with a haematological malignancy was the t(9;22), which fuses the *BCR* gene to the tyrosine kinase gene *ABL* in CML (19) and is able to transform cells *in vitro* and *in vivo* (22). Indeed, with many different acquired translocations associated with malignancy this is clearly an important mechanism for the genesis of neoplasms and as such these translocations have been intensely studied. The mechanisms by which they are formed, however, is still incompletely understood.

In its simplest form, translocations are the result of two DSBs that occur in pairs of non-homologous chromosomes (199) that must be in close proximity to one another. Indeed it has been shown that the distances between genes commonly involved in both lymphoid and myeloid translocations are a lot shorter than control genes during cell division (200, 201). The subsequent aberrant ligation represents a failed attempt for DNA repair enzymes to mend the severed DNA (203). The resultant gene fusions must confer a selective advantage to enable biological selection and clonal expansion.

DSBs can be caused by a number of exogenous factors such as ROS, ionising radiation, genotoxic chemicals and DNA replication stress (202) or by the cell itself, as ROS are by-products of normal respiration. Due to the danger that DSBs cause to the integrity of the genome, pathways for efficient and accurate repair of DSBs are critical (312, 313) and there are 2 main mechanisms by which they can be repaired: (1) NHEJ rejoins free DNA ends without the use of sequence homology and as such the DNA may undergo modifications during the ligation process and is potentially mutagenic and (2) HR which utilises long stretches of homologous DNA as a template to repair breaks and is therefore more likely to maintain genomic integrity. Single stranded annealing (SSA) is another type of HR but it uses single stranded tails to recombine at repeat regions which results in

deletions and is therefore mutagenic. In addition, lymphoid cells have the unique capability to generate (and repair) DSBs in a controlled manner, in order to generate the large repertoire of antibodies required to mount an efficient immune response (258). However these mechanisms can also be used to aberrantly produce and repair DNA breaks and therefore give rise to translocations.

By cloning translocation breakpoints and determining if there are any breakpoint clusters or sequence homologies, the mechanisms by which some translocations occur have been elucidated in some subtypes of disease. For example, the mechanisms behind many chromosomal translocations in lymphoid malignancies are believed to be the result of illegitimate recombination mechanisms such as V(D)J (204, 207, 208, 263, 270) or CSR (205, 206). This highlights the importance of determining what cell-type and stage of development the translocation may occur as this could affect what mechanisms are active and therefore likely to be responsible.

The mechanisms behind chromosomal translocations in myeloid and the remainder of lymphoid malignancies are poorly understood. Sequence analysis of breakpoints has revealed that the majority probably result from NHEJ, less frequently from SSA and very rarely from HR (223, 224, 314), however, the reason why some fusions are seen very frequently compared to other known or putative fusions is still unclear. Two main hypotheses have been proposed.

The first is the ‘selection hypothesis’ which poses that breaking and re-ligation events between chromosomes occurs relatively frequently, however leukaemia is only seen in those which provide a selective advantage for the cell (226, 229). The second is the ‘hot-spot hypothesis’ which poses that the breaks are non-random and due to particular fragile sites in the genome that are particularly susceptible to breaks such as non-B forms of DNA or S/MARs (218, 220-222, 315). However, it is likely that both of the above hypotheses have their part to play in the genesis of translocations and elucidation of these mechanisms provides vital information which could potentially lead identification of risk factors and possibly preventative measures.

In addition to the elucidation of the mechanisms that give rise to translocations, it is also vitally important to be able to quickly and accurately identify the oncogenic lesions in leukaemia patients so that the necessary and correct treatment regimes can be administered. Even within the same group of disease, such as ALL, there are many different disease causing translocations, which not only have differing prognostic outcomes, but different optimal treatments, such as *BCR-ABL* positive ALL which has a poorer outcome and requires more aggressive treatment than, for example *TEL-AML1* (316).

Once a particular fusion gene has been identified in a patient and the corresponding treatment regime undertaken, it is also important to be able to monitor the effectiveness of the regime. As the hallmark of a particular leukaemia caused by a gene fusion is the presence of this marker in the patients blood or bone marrow, it can be used to determine if treatment has been successful by monitoring for minimal residual disease (317). Indeed the depth of response gauged by such monitoring can in itself be used as a prognostic indicator (318). If the patient has a complete molecular response, regular monitoring can be used to catch any relapse at an early stage and to change therapy.

In the past 50 years since the identification of the first translocation associated with a malignancy (19), many efforts have been made to identify further genetic lesions that give rise to leukaemia, their mode of action and the mechanisms by which these arise. This research has led to many seminal discoveries, culminating in the first rationally designed drug, imatinib, that can specifically target a number of tyrosine kinases (61) including *BCR-ABL*. There is now a plethora of data on many different types of leukaemia and treatment regimes can now be tailored depending on the acquired genetic abnormality (although for most of these the therapy is not yet truly targeted).

The main unifying aim of this project was to identify genomic breakpoints, but for very different purposes. Toward this end several different techniques were optimised and used to varying degrees of success. The most successful technique which was employed for both parts of this project was the LR-PCR, which proved to be a robust and quick method of amplifying breakpoints after it had been optimised.

In the first part of this project, genomic breakpoints were amplified from p190 ALL, p210 ALL and p210 CML to try and establish if there were lymphoid-specific mechanisms responsible for the over representation of p190 breakpoints in ALL compared to CML. Although 3 other techniques were trialled, the only method by which breakpoints were successfully amplified was LR-PCR. In total 82 *BCR-ABL* breakpoints were amplified and sequenced which broke down into to 25/43 p190 ALL (i.e. amplification was attempted from 43 cases and 25 were successful), 25/32 p210 ALL, and 32/32 p210 CML. However there are relatively few breakpoints published in the literature, the paucity of breakpoint data in the literature highlights the difficulties of isolating these sequences, especially the p190 breakpoints which not only have the large *ABL* breakpoint region but also a considerably larger region in *BCR* where breakpoints can occur. The percentages of successfully characterised breakpoints were higher for p210 (89%) than p190 breakpoints (58%), and the most likely reason for this is that follow up simplex PCR was not performed for p190 cases due to the large numbers of PCRs and large amount of patient DNA this would involve.

Amplification of the reciprocal breakpoints revealed that, in concordance with previously published data generated by FISH and RT-PCR (271, 276), the p190s had much lower frequency of derivative 9 deletions compared to the established figure of around 15% in CML (125). Fine breakpoint mapping revealed that in the 20 reciprocal breakpoints amplified (of 25 tested) the largest deletion/insertion or duplication event was 27 bp, which was in stark contrast to the p210s. Statistical analysis did not reveal any differences between the locations of forward breakpoints between the subtypes of leukaemia, however, two clusters were found for p210 ALL in both *BCR* and *ABL* in addition, when all breakpoints were combined (p210 and p190) there were also two clusters observed in *ABL*. The p190 breakpoints were not found to have any clustering within the m-bcr, contrary to previous reports that suggested sub-clustering of these breakpoints, a possible clustering of breakpoints was found in the 3' end of the *ABL* breakpoint cluster region (see section 3.3.9).

Analysis of sequence motifs found one significant association, a deficit of repeat regions in p210 *BCR* breakpoints which might possibly be related to the non-repeat recombination motif (CCTCCCT) and its 13-mer family member, which has been reported to be located close to many translocation breakpoints (260). Despite the lack of association of any of the breakpoints with consensus or functional RSSs, it was still possible that a subset of p190 ALL breakpoints were caused by illegitimate RAG recombination utilising cryptic sites and therefore all breakpoints were screened for cryptic RSSs and a representative subset selected for functional testing.

The functional testing, although limited to 10 cases, revealed 1/5 p190 cases had a 'specific' RSS next to the breakpoint which was targeted for both CJ and SJ formation, the presence of both are required for the definitive V(D)J recombination. In addition 1/3 p210 ALL patients was found to have a 'specific' RSS at the breakpoint in *ABL* which showed a SJ and a hybrid joint but did not show coding joint formation. In addition there were 3 other RSSs in the flanking sequence that were also targeted by RAG, which means the origins of this breakpoint are less certain. The remaining cases tested did not show any 'specific' recombination, with the caveat that the number of 'non-specific' recombination events was low.

If there was more time there are several experiments which could be done. First, it would be desirable to identify additional forward and reciprocal breakpoints. More p190 forward breakpoints could be screened for by running the LR-PCR simplex PCRs on the multiplex negative patients, or at least those for whom sufficient DNA is available, or the multiplex LR-PCR could be repeated with an increased magnesium concentration as with the *FIP1L1-PDGFR* LR-PCR multiplex. If this was unsuccessful, more *BCR* forward primers could be designed. In addition the remaining unidentified reciprocal breakpoints could also be run on a reciprocal breakpoint multiplex LR-PCR to see if intermediate sized breaks could be identified. It would be interesting (but expensive) to perform high density array CGH on these cases to examine the extent of the deletions and provide sufficient resolution to enable the reciprocal breakpoints to be amplified by LR-PCR.

For the functional analysis, more transfections will be run on those cases with constructs already made and, in addition, some of the remaining breakpoints will be tested for 'specific' V(D)J recombination. As the two cases that showed 'specific' RSSs did not have cryptic RSSs near the breakpoint in the translocation partner, it is possible that this DNA was a target for RAG mediated transposition whereby a signal joint is inserted into an unrelated DNA target (277, 278, 284-286). This is believed to be facilitated by the fact that signal joints are degraded slowly thereby providing a prolonged opportunity for irregular events to occur (282). The structure and nucleotide sequence of the target DNA is of importance for this mode of recombination and non-B forms, especially cruciforms, are preferred targets (286). This may be relevant for *BCR* which has an internal promoter in intron 1 upstream of the m-bcr (287). Potentially this could result in a distorted DNA structure, although this would be highly unusual if it extended over such a large region. Non-B form DNA has also been found to be a target of RAG cutting which is able to generate double stranded breaks even in the absence of cryptic RSSs, as has been described in follicular lymphomas (254, 267). This may be another way of generating double stranded breaks in ALL patients with subsequent RAG directed or other mechanisms of joining facilitated by the proximity of chromosomes 9 and 22 during mitosis (200). The fact that *BCR* is located near to the *IgL* locus on chromosome 22 may also be relevant. Indeed, the finding of RAG mediated deletions of exons 3-6 in the transcription factor *Ikaros* in the majority of *BCR-ABL* positive ALL and CML in lymphoid transformation but not chronic phase CML clearly substantiates the notion that illegitimate RAG activity is important for the genesis of lymphoid malignancies (119). In addition, specific mutations in *RAG1* and *RAG2* can alter the targeting and processing of RSSs (319, 320) and a gain of function mutation has been identified *in vitro* that increases the frequency and infidelity of targets (321). It is possible therefore that mutations in *RAG* genes might be a feature of lymphoid disorders by promoting genetic instability. Alternatively, the breaks could be the result of another lymphoid specific enzyme such as AID, and although this is unlikely as this enzyme is usually active in later B-lymphoid cells, it is not impossible.

A successfully employed method for amplifying tightly clustered breakpoints was bubble PCR, which has the added advantage of allowing amplification from a known gene into an unknown partner. This resulted in the amplification of a novel fusion partner to *PDGFRA* which was only the second fusion to be associated with CEL. This highlighted the

possibility of further fusion partners for *PDGFRA* and to date a total of six have been identified (168). In addition bubble PCR was used to amplify some *FIP1L1-PDGFRA* genomic breakpoints, but this was later changed to a multiplex LR-PCR strategy which was found to be more efficient than the RT-PCR amplifying all the positive controls at single step compared to the RT-PCR technique for which many breakpoints could only be amplified by nested amplification. As a result, the LR-PCR has now been adopted in a diagnostic setting to run alongside the RT-PCR. In addition the genomic breakpoints were used to design patient-specific gDNA RQ-PCRs which were more sensitive than analysis of cDNA for detection of MRD. In addition, the results could be directly translated into the proportion of fusion positive cells, which is not possible for cDNA-based techniques due to variability in fusion gene transcript numbers per cell. Furthermore, levels of the gDNA fusion at presentation, and thus the clone size, were shown to be widely variable and in some cases very low which may account for why some patients are negative by FISH.

Finally, the question remains why the breakpoints within *PDGFRα* are so tightly clustered within exon 12. One reason for this is likely to be functional, and support for this comes from a recent analysis that showed *FIP1L1* is completely dispensable for transformation by *FIP1L1-PDGFRα*, and that the breaks between the two tryptophan residues in the WW domain is sufficient to activate the fusion in the absence of a partner gene encoded dimerisation domain (301). However, several questions remain, e.g. if a dimerisation domain is not required when the WW domain is disrupted, why do *BCR* and *KIF5B* (which are known to harbour dimerisation domains) also disrupt this region when fused to *PDGFRα*. Also why are WW domain disrupting fusions not seen for *PDGFRβ* fusions (the WW domains are almost identical in *PDGFRα* and *PDGFRβ*). One possible explanation is that *PDGFRα* exon 12 is particularly susceptible to breakage, whereas the corresponding exon in *PDGFRβ* is not (302). Potentially this could relate to the fact that *PDGFRα* has a developmentally regulated internal promoter in intron 12 that may be associated with an open chromatin structure. To investigate this DNaseI hypersensitivity assays could be performed around *PDGFRα* and *PDGFRβ* exon 12 using myeloid cell lines.

The work in this thesis has demonstrated the value of translocation breakpoint characterisation to help understand the genesis of specific diseases as well as providing the

means for improved diagnosis and follow up of affected patients after treatment. Although the current techniques are laborious and cumbersome, breakpoint characterisation is likely to become easier and cheaper in the future with the development of new analysis techniques such as ‘next-generation’ sequencing.

Appendix I

Solutions

Guanidine Thiocyanate (GTC): 4M Guanidine thiocyanate, 5mM EDTA, 25mM citrate pH 7.0, 0.5% sarcosyl. 7 µl β-mercaptoethanol per 1 ml of GTC was added fresh prior to use.

Red cell lysis buffer: 155mM KHCO₃, 0.1mM EDTA, pH 7.4 at 4°C.

RSB: 4.38g NaCl, 48 ml EDTA made up to 1 litre with sterile water.

10% SDS: Made with equal volumes of 20% SDS and sterile water.

P1: 5mM glucose, 25mM Tris-Cl, 10mM EDTA, pH 8.0

P2: 0.2M NaOH, 1% SDS.

P3: 5M potassium acetate, 3M glacial acetic acid.

Biotin (Bio) Labelling mix: 0.5M Tris-Cl pH 7.8, 0.05M MgCl₂, 0.1M B-mercaptoethanol, 20 µM dATP, 20 µM dGTP, 20 µl dCTP, 10 µM dTTP, 0.5 µg BSA and 25 µg of d UTP conjugated with Biotin.

Digoxigenin (Dig) Labelling mix: As Biotin mix except 0.5 µg BSA and 25 µg digoxigenin added fresh when labelling reactions are set up.

SSCT: 4 x SSC, 0.1% Triton-X pH 7.0.

Hybridisation Buffer (HB): 50% formamide, 2 x SSC, 20% dextran sulphate, pH 7.0.

SSCTB: 2 x SSC, 0.1% Triton-X, 0.03 g/ml BSA.

SSCT: 4 x SSC, 0.1% Triton-X, pH 7.0.

Phosphate buffered saline (PBS): PBS was prepared as instructed by the manufacturer (Oxoid, Basingstoke, UK). One tablet was dissolved in 100 ml of distilled water.

6 x loading buffer with Orange-G loading dye: 40% sucrose, 0.25% Orange-G, 1 μ M EDTA.

LB Broth: 25 g LB broth per litre dH₂O. Autoclaved before use.

LB Agar: 3 g agar per 200 ml LB broth. Autoclaved before use.

Appendix II

PCR programs

High Fidelity PCR program

Step	Temp (°C)	Time
1	95	2 min
2	94	20 sec
3	66/64	4 min
4	68	14 min
5	10 times to 2	
6	94	20 sec
7	66/64	40 sec
8	68	14 min
9	22 times to 6	
10	72	9 min
11	15	Forever
12	End	

Long Range PCR program

Step	Temp (°C)	Time
1	95	2 min
2	94	20 sec
3	66/64	30 sec
4	68	8 min
5	10 times to 2	
6	94	20 sec
7	66/64	40 sec
8	68	8 min
9	25 times to 6	
10	68	9 min
11	15	Forever

12	End	
----	-----	--

Sequencing program

Step	Temp (°C)	Time
1	96	1 min
2	96	10 sec
3	50	30 sec
4	60	2 min
5	25 times to 2	
6	4	Forever

Pfu polymerase program

Step	Temp (°C)	Time
1	96	1 min
2	96	30 sec
3	64	30 sec
4	72	1 min
5	30 times to 2	
6	4	Forever

Primers

M13 primers for TOPO TA clone sequencing

M13 F 5' GTA AAA CGA CGG CCA G 3'

M13 R 5' CAG GAA ACA GCT ATG AC 3'

BCR primers used in PCR screening of p190s

BCR forward primers sets A and B:

Forward Primer A Primer Sequence (5' to 3')

BCR 1A Bt 1 (374) F GGA GTC ATC TCT AGC GCC CAC AT

BCR 2A Bt 1 (13651) F TGC AGT CAC CAT GAT CCC TTC TC

BCR 3A Bt 2 (11553) F GCA GGA CTC CCC TAC CCC ATA AC

BCR 4A Bt 2 (23655) F CAG GGA CTG ACT CCA CAA CAG GA

BCR 5A Bt 3 (15209) F CTC CTC AGT GCT GGT GTC TGA GC

Forward Primer 'B'	Primer Sequences (5' to 3')
BCR 1B Bt 1 (9246) F	CAA AGC ATT TTC CCC TGC GTA TC
BCR 2B Bt 2 1F	TAT GAG TTT CTG CCC ATG ATG GG
BCR 3B Bt 2 (17001) F	CTT GCT GAC CTT GAT GGA GTG GA
BCR 4B Bt 3 (6494) F	CTA GAG AAC CCA TCC CAC GCA GT
BCR 5B Bt 3 (22728) F	CTC CAA AAG ACG TGT GCT GCT GT

BCR reverse primers used to test forward primers work:

Reverse Primer A	Primer Sequences (5'-3')
BCR 1A Bt 1 (9246) R	GAT ACG CAG GGG AAA ATG CTT TG
BCR 2A Bt 2 (2801) R	AGG GTC ATC AAG GAA TCG GGA AT
BCR 3A Bt 2 (16990) R	AAG GTC AGC AAG GCG AAG AGA GA
BCR 4A Bt 3 (6370) R	TCT TTA TGC ACC AGT GGC CTG AA
BCR 5A Bt 3 (22870) R	TTT TCA CAG AGG CGG ACA TCA AA

Reverse Primer B	Primer Sequences (5'-3')
BCR 1B Bt 1 (17176) R	TCA ACC TTC CCA GAC TTC CTT GC
BCR 2B Bt 2 1R	CAA GGT TCT CAC AAA TGC CAG AG
BCR 3B Bt 2 (23641) R	CAG TCC CTG TCT CTG CCA TTT CA
BCR 4B Bt 3 (14649) R	GAA GGT GCC TAT GAA CCC TGC TG
BCR 5B Bt 3 (29314) R	CAA CAG GCA TCT CAC AAC GAC CT

BCR C primer set:

Forward Primer	Primer Sequences (5' to 3')	Reverse Primer
BCR 1C bt 2 (18817) F	TTCCAGTGAGTCGGTGAAAGCTG	Use 3B (23641) R
BCR 2C bt2 (21316) F	CCCACCTTTTCTTTGTGGGACAGC	Use 3B (23641) R
BCR 3C bt3 (972) F	GGGCTCATTTTCACTGGATGGAC	Use 4A (6370) R
BCR 4C bt 3 (3305) F	TCCTGTACATCAGACCCTTTGC	Use 4A (6370) R
BCR 5C bt 3 (8176) F	CCACCCTTTACCTGACCAGCATC	Use 4B (14649) R
BCR 6C bt 3 (10651) F	GTGAGGGTGAAAACCAGGGTCAG	Use 4B (14649) R
BCR 7C bt 3 (12440) F	CAAGGAGCAGGAATCAGCCTTGT	Use 4B (14649) R
BCR 8C bt 3 (17350) F	GCCTGCGTTTTTCTACCAACACC	Use 5A (22870) R
BCR 9C bt 3 (20584) F	TGGAGACAGGGGATGGTCCTTTA	Use 5A (22870) R
BCR 10C bt 3 (25406) F	GGGGCAACAAGTATGGATTGGAG	Use 5B (29314) R

BCR 11C bt 3 (27463) F GGAGTGGAGAAGGGATCCTCACA Use 5B (29314) R

BCR primers for p210 screening

BCR p210 forward primers used for screening p210s:

Forward Primer	Primer Sequences (5'-3')
BCR B2 F (exon 13)	ACT CGT GTG TGA AAC TCC AGA CT
BCR B3 F (exon 14)	TGG GTT TCT GAA TGT CAT CGT CC

ABL primers used in PCR screening p210s and p190s

Reverse Primer A	Primer Sequences (5'-3')
ABL 1A (26693) R	AAC ACG GAG AAG TGG CAA ACC TC
ABL 2A (26877) R	TAA GGC AGT TAC CAG GAA GCA TTT
ABL 3A (8944) R	GGA AGG AGG AGG AAA TGA CAG CA
ABL4A (16040) R	ACT GTT CAC TAA GTG GCA CTG TG
ABL 5A (5357) R	CAT GAT GTG CTT TGC AGG GTA GC

Reverse Primer B	Primer Sequences (5'-3')
ABL1B (8470) R	GAT CTG AAG CAC AAG CAC GGT TC
ABL2B (15120) R	GGA ACA GGA ATC CTA ATG GCC AAC
ABL3B (20180) R	ACA TGG GGC ACA GTC TCT TGA TG
ABL4B (7790) R	AGG AAT GGG TAT GCT GGG GTT G
ABL5B (16551) R	ATG GAA TGA CTC CCA CCT GAA AG

Reverse Primer C	Primer Sequences (5'-3')
ABL 1C (937) R	AAG CCA CTG GCA CAC TTC ATA CG
ABL 2C (4301) R	TGG ACC AGG CTT TAG CCC TAT CA
ABL 3C (2114) R	CCC GCA GTA TCC CTC AAA ATC AG
ABL 4C (3) (28389) R	ACA TGA GGT TTG CAG AAG CAC CA

Reverse Primer D	Primer Sequences (5'-3')
ABL 1D (17887) R	AGC CAT AAC CAT TCT CCC AAG CA
ABL 2D (21000) R	ACC AAA GCC TCC CCT TGT ACC TC
ABL 3D (14776) R	TCA TGG AGA AAG GGG GAG AAC TG
ABL 4D (24077) R	CCA GAA GAA CAC CCC AAG AAG GA

ABL 5D (9338) R	AAA CAT TCT GCC GCA TCT GGA TT
-----------------	--------------------------------

ABL primers used to test reverse primers work:

Forward Primer A	Primer Sequence (5' to 3')
ABL 1A (18236) F	CAA AGC ATT TTC CCC TGC GTA TC
ABL 2A (19817) F	TAT GAG TTT CTG CCC ATG ATG GG
ABL 3A (920) F	CTT GCT GAC CTT GAT GGA GTG GA
ABL 4A (11147)	CTA GAG AAC CCA TCC CAC GCA GT
ABL 5A (1735) F	CTC CAA AAG ACG TGT GCT GCT GT

Forward Primer B	Primer Sequences (5' to 3')
ABL1B 1F	TCA ACC TTC CCA GAC TTC CTT GC
ABL2B (6300) F	CAA GGT TCT CAC AAA TGC CAG AG
ABL3B (14450) F	CAG TCC CTG TCT CTG CCA TTT CA
ABL4B (1280) F	GAA GGT GCC TAT GAA CCC TGC TG
ABL5B (8750) F	CAA CAG GCA TCT CAC AAC GAC CT

Forward Primer C	Primer Sequence (5' to 3')
ABL 1C (92) F	GCG ACA GTT CCT TCC AAT TCC AC
ABL 2C (241) F	AGA GCG CCT GCT GTT TGA TTT TC
ABL 3C (2) (27017) F	ACA TCG GAA CAC TGG TCT GGT CA
ABL 4C (3) (22069) F	TGG ACG CTA CCT TGA CAG AGT GTG

Forward Primer D	Primer Sequence (5' to 3')
ABL 1D (10441) F	TTT TAG CCT TGG CAC ACC AGT CA
ABL 2D (14516) F	GGA GAC CAT GTC TCA GTG GTG GA
ABL 3D (7683) F	GCT GTG TGT TCC TGT GGA GCT GT
ABL 4D (17150) F	TTT GTT GTT TCT TCG CCA TCA GC
ABL 5D (3948) F	TTC TGG GGA AAT TGC CTG TCA TT

ABL upstream primers for p190s and p210s

Forward primer	Primer sequence (5' to 3')
ABL US 2 F	CCTCCCAATGCCTAGACTCTCCA

ABL US 3 F	TGACATGGTATGCCAACATGGAA
ABL US 4 F	GCAGATCAGTGCAGGCTTCAGAG
ABL US 5 F	ACCCAGAGGCACACATCTCACAG
Reverse primer	
ABL US 2 R	TCCATGTTGGCATAACCATGTCAG
ABL US 3 R	CTCTGAAGCCTGGCACTGATCTGC
ABL US 4 R	GTGCAAATCCATGTCTGGTCCTG
ABL US 5 R	CACAATTCAAAGGGGCTTCATCA

New BCR and ABL primers used in new screen

Forward pirmer	Primer sequence (5' to 3')
BCR 4A(b) bt2 (23655) F	CAGGGACTGACTCCACAACAGGA
BCR 5A (b) bt3 (15152) F	TCTTCCCCTCTGTCTGCCACTCT
BCR 3B (b) bt2 (16990) F	TCTCTCTTCGCCTTGCTGACCTT
BCR 4B (b) bt3 (6508) F	CCACGCAGTTGTTTTGGACACAT
BCR 5B (b) bt3 (22728) F	CTCCAAAAGACGTGTGCTGCTGT
Reverse primer	
ABL 3A (b) (8929) R	TGACAGCAGAGGAGCTTCATCCA
ABL 4B (b) (10202) R	ACAGGGGCCCAGAGTGGTTAGTT
ABL 1C (b) (937) R	AAGCCACTGGCACACTTCATACG
ABL 2D (b) (21151) R	TGTCTGCGGTCATGGATCTCTTC
ABL 3D (b) (14858) R	CCTCCCCCAATTCCGTACTIONTCTT

p210 ALL forward breakpoint primers

Primers used to amplify breakpoint band with LR-PCR:

Patient ID	BCR forward primer	ABL reverse primer
P682	BCR B3 1F	ABL 2B R
P851	BCR B3 F	ABL 5B R
22251	BCR B3 1F	ABL 3B R
26231	BCR B3 F	ABL 4B R
24375	BCR B3 F	ABL 4C R
26219	BCR B3 F	ABL 1A R

19715	BCR B2 F	ABL 3C R
6875	BCR B3 F	ABL 5D R
F22666	BCR B2F	All ABLs
10887	BCR B2FR	ABL 5B R
10258	BCR B3 F	ABL 1D R
11129	BCR B2 F	ABL 3C R
11128	BCR B2 F	ABL 4A R
F17685	BCR B3 F	ABL 5B(b) R
F17730	BCR B2 F	ABL 4D R
F18544	BCR B2 F	ABL 4B R
F17237	BCR B3 F	ABL 2D R
F16604	BCRB2 F	ABL 1A R
F15010	BCRB2 F	ABL 4C R
F22757	BCR B3 F	ABL 4C R
F14472	BCRR B2 F	ABL 5A R
F11329	BCR B2 F	ABL 4B R
F11751	BCR B3 F	ABL 2C R
F24559	BCR B2 F	ABL 2C R
F20099	BCR B3 F	ABL 2C R

Primers used to sequence over breakpoints:

Patient ID	Primer name	Primer Sequence (5'-3')
P682	BCR bp (P851) 2F	TGC TCT GTC GAG CTG GAT GGA T
P851	BCR bp (P851) 2F	TGC TCT GTC GAG CTG GAT GGA T
22251	BCR B3 (1850) F	TGC TGT TTG CGC TCA CAT TTA CA
26231	BCR B3 (2340)F	GGC CTC CTC CCT GGT CTT TGT AG
24375	BCR bp (P851) 2F	TGC TCT GTC GAG CTG GAT GGA T
26219	BCR B3 F	TGG GTT TCT GAA TGT CAT CGT CC
19715	BCR B2 (368) F	CTC CTC AGA TGC TCT GTG CCT TG
6875	BCR b3 (1850) F	TGC TGT TTG CGC TCA CAT TTA CA
F22666	Cloned	M13 R
10887	BCR B2 (368) F	CTC CTC AGA TGC TCT GTG CCT TG

10258	BCR bp (P851) 2F	TGC TCT GTC GAG CTG GAT GGA T
11129	BCR B2 (368) F	CTC CTC AGA TGC TCT GTG CCT TG
11128	BCR B2 (368) F	CTC CTC AGA TGC TCT GTG CCT TG
F17685	BCR B3 (1762) F	GAATGGGGTTGGGAGAGAGGACT
F17730	BCR B2 F	ACT CGT GTG TGA AAC TCC AGA CT
F18544	BCR B2 F	ACT CGT GTG TGA AAC TCC AGA CT
F17237	BCR B3 (2340) F	GGC CTC CTC CCT GGT CTT TGT AG
F16604	BCR B2 (368) F	CTC CTC AGA TGC TCT GTG CCT TG
F15010	BCR B2 (368) F	CTC CTC AGA TGC TCT GTG CCT TG
F22757	BCR bp (P851) 2F	TGC TCT GTC GAG CTG GAT GGA T
F14472	BCR B2 (368) F	CTC CTC AGA TGC TCT GTG CCT TG
F11329	BCR B2 (368) F	CTC CTC AGA TGC TCT GTG CCT TG
F11751	BCR B3 F	TGG GTT TCT GAA TGT CAT CGT CC
F24559	BCR B2 (368) F	CTC CTC AGA TGC TCT GTG CCT TG
F20099	BCR B3 (2340) F	GGC CTC CTC CCT GGT CTT TGT AG

p210 CML forward breakpoint primers

Primers used to amplify breakpoint band by LR-PCR:

Patient ID	BCR forward primer	ABL reverse primer
NP241	BCR B2 F	ABL 4B R
E656	BCR B3 F	ABL 5B R
1716	BCR B3 F	ABL 4A R
1349	BCR B3 F	ABL 1A R
1578	BCR B2 F	ABL 4A R
1103	BCR B3 F	ABL 4A R
1810	BCR B2 F	ABL 2D R
1569	BCR B2 F	ABL 2B R
1338	BCR B3 F	ABL 2D R
1460	BCR B3 F	ABL 2C R
230	BCR B2 F	ABL 5A R
248	BCR B2 F	ABL 1B R
266	BCR B2 F	ABL 2A R
316	BCR B2 F	ABL 4B R

403	BCR B2 F	ABL 5D R
440	BCR B2 F	ABL 2B R
446	BCR B2 F	ABL 1A R
483	BCR B2 F	ABL 4A R
495	BCR B2 F	ABL 1D R
524	BCR B2 F	ABL 4D R
545	BCR B2 F	ABL 1B R
282	BCR B3 F	ABL 3C R
286	BCR B3 F	ABL 1C R
287	BCR B3 F	ABL 3D R
291	BCR B3 F	ABL 4C R
300	BCR B3 F	ABL 4B R
317	BCR B2 F	ABL 5A R
375	BCR B3 F	ABL 4C R
386	BCR B3 F	ABL 1D R
436	BCR B3 F	ABL 2B R
389	BCR B3 F	ABL 3D R
470	BCR B3 F	ABL 2A R

Primers used to sequence over breakpoint:

Patient ID	Primer name	Primer sequence (5'-3')
NP241	BCR B2 F	ACT CGT GTG TGA AAC TCC AGA CT
E656	BCR B3 F	TGG GTT TCT GAA TGT CAT CGT CC
1716	BCR b3 (1850) F	TGC TGT TTG CGC TCA CAT TTA CA
1349	BCR b3 (2340)F	GGC CTC CTC CCT GGT CTT TGT AG
1578	BCR B2 (368) F	CTC CTC AGA TGC TCT GTG CCT TG
1103	BCR B3 (2737) F	TAGGGGCTTTAGCTGGGGTTTGT
1810	BCR B2 (1) F	GTTTCAGAAGCTTCTCCCTGACA
1569	BCR P851 F	TGC TCT GTC GAG CTG GAT GGA T
1538	BCR B2 (368) F	CTC CTC AGA TGC TCT GTG CCT TG
1460	BCR P851 F	TGC TCT GTC GAG CTG GAT GGA T
230	BCR B3 F	TGG GTT TCT GAA TGT CAT CGT CC
248	BCR B2 F	ACT CGT GTG TGA AAC TCC AGA CT

266	BCR B2 F	ACT CGT GTG TGA AAC TCC AGA CT
316	BCR B2 F	ACT CGT GTG TGA AAC TCC AGA CT
403	BCR B2 F	ACT CGT GTG TGA AAC TCC AGA CT
440	BCR B2 F	ACT CGT GTG TGA AAC TCC AGA CT
446	BCR B2 (516) F	ATGCACATGTGTCCACACACACC
483	BCR B2 F	ACT CGT GTG TGA AAC TCC AGA CT
495	BCR B2 F	ACT CGT GTG TGA AAC TCC AGA CT
524	BCR B3 1850 F	TGC TGT TTG CGC TCA CAT TTA CA
545	BCR B2 F	ACT CGT GTG TGA AAC TCC AGA CT
282	BCR B3 F	TGG GTT TCT GAA TGT CAT CGT CC
286	BCR P851 F	TGC TCT GTC GAG CTG GAT GGA T
287	BCR B3 F	TGG GTT TCT GAA TGT CAT CGT CC
291	BCR B3 F	TGG GTT TCT GAA TGT CAT CGT CC
300	BCR B3 (2737) F	TAGGGGCTTTAGCTGGGGTTTGT
317	BCR B2 F	ACT CGT GTG TGA AAC TCC AGA CT
375	BCR 1850 F	TGC TGT TTG CGC TCA CAT TTA CA
386	BCR P851 F	TGC TCT GTC GAG CTG GAT GGA T
436	BCR 1850 F	TGC TGT TTG CGC TCA CAT TTA CA
389	BCR B3 F	TGG GTT TCT GAA TGT CAT CGT CC

p190 ALL forward breakpoint primers

Primers used to amplify breakpoint band with LR-PCR:

Patient ID	BCR forward primer	ABL reverse primer
Ital 1	BCR 3B F	ABL 5A R
Ital 2	BCR 4A F	ABL 2A R
Ital 4	BCR 4A F	ABL 3 C
Ital 6	BCR 3B F)	ABL 1B R
Ital 9	BCR 5A(b) F	ABL 4A R
Ital 10	BCR 5B F	ABL 5B R
21296	BCR 5A F	ABL 1A R
23827	BCR 4B F	ABL 2D R
25169	BCR 5A	ABL 4A R
25373	BCR 5B F	ABL 4D R

25654	BCR 5B F	ABL 4A R
9463	BCR 3B F	ABL 4D R
20763	BCR 5B	ABL 2B R
44095	BCR 3B F	ABL 2B R
27523	BCR 4B F	ABL 1B R
28624	BCR 5B F	ABL 4A R
28835	BCR 5A F	ABL 2C R
28852	BCR 4B F	ABL 4A R
19635	BCR 4AF	ABL 1C R
24350	BCR 4A F	ABL 4C R

Primers used to sequence over breakpoint:

Patient ID	Primer name	Primer Sequence (5'-3')
Ital 1	BCR bt2(18020) F	CAC CCA CTA GTT GTC CCT GCT CA
Ital 2	ABL 4A (16040) R	ACT GTT CAC TAA GTG GCA CTG TG
Ital 4	BCR bt 3 (6175) F	TTT TTC CAG GCA AAG CAC CTC TC
Ital 6	BCR bt2(19326) F	ACT TGC CCA GCT CTG TCC TCT GT
Ital 9	ABL 4 (11131) R	CCGCGCCCAGCTGAAAAATACCT
Ital 10	BCR bt3(24281) F	TTT CAC TAG GCC CTC CCT CTT CC
21296	BCR bt3(17086) F	GGG ACT CCC TCC AGT CAC TTT CA
23827	BCR bt 3 (13916) F	AAT GGA AGA GTT CCT GGC CAC A
25169	BCR bt 3 5A (22870)F	CTC CTC AGT GCT GGT GTC TGA GC
25373	BCR bt 3 (25887) F	GGG CTC GGA AAG CAC TCA GAT AA
25654	ABL 4 (11230) R	GAG ACG CCA CCA ATG ACC AAA AC
9463	BCR 3B Bt2 (23641) F	CTT GCT GAC CTT GAT GGA GTG GA
20763	BCR bt 3 (23282) F	TAG GCA TAT CTG GCC TGG GAG AG
44095	BCR bt 2 (20399)F	CTG TGG GCT TCC AGG GAG ATT TA
27523	BCR Bt 3 (13468) F	GATTCCTGAAGTGGAGCCCTCA
28624	BCR Bt 3 (28590) F	TCCAGAAGACTGCCCCCTAACGTC
28835	BCR Bt 3 (17550) F	CCC ACA AAA TCA CCC TCT TCC TC
28852	BCR bt 3 (12440) F	CAA GGA GCA GGA ATC AGC CTT GT
19635	BCR Bt 3 (4496) F	GGT TCA CTG CAT TTC TCG GGT CT

24350	BCR Bt 3 (6038) F	GTGGTGGCCCATACTGATCACCT
-------	-------------------	-------------------------

p210 ALL reciprocal breakpoint primers

Primers used to amplify reciprocal breakpoint bands:

Patient ID	<i>ABL</i> forward primer	<i>BCR</i> reverse primer
P851	ABL 5D F	BCR B4 (4099) R
26231	ABL 4B F	BCR B4 (4099) R
P682	ABL 2D F	BCR B4 (4099) R
22251	ABL3B F	BCR B4 (4099) R
26219	ABL1 (20970) F	BCR B4 (4099) R
24375	ABL 4D F	BCR B4 (4099) R
19715	ABL 2 (27499) F	BCR B4 (4099) R
11129	ABL3B F	BCR B4 (4099) R
10258	ABL 1D F	BCR B4 (4099) R
11128	ABL 4A F	BCR B4 (4099) R
6875	ABL 2B F	BCR B4 (4099) R
10887	ABL 5 (11708) F	BCR B4 (4099) R
F22666	ABL 2B F	BCR B3 (3000) R
F17685	ABL 5B F	BCR B3 (3000) R
F17730	ABL 4D F	BCR B3 (3000) R
F18544	ABL 4B F	BCR B3 (3000) R
F17237	ABL 2D F	BCR B3 (3000) R
F16604	ABL 1A	BCR B3 (3000) R
F15010	ABL 3 (27172) F	BCR B3 (3000) R
F22757	ABL 4C (3) F	BCR B3 (3000) R
F14472	ABL 5A F	BCR B3 (3000) R
F11329	ABL 4B F	BCR B3 (3000) R
F11751	ABL 2C F	BCR B3 (3000) R
F24559	ABL 2C F	BCR B3 (3000) R
F20099	ABL 1 F	BCR B3 (3000) R

New primers used to amplify reciprocal breakpoints:

Primer	Primer sequence (5'-3')
--------	-------------------------

ABL1 (20970) F	GCACCTCTGCCACTCTGTCCTTT
ABL 2 (27499) F	GCCAGAGTGAATGTCCTGAGCAA
ABL 5 (11708) F	CAGTGGGAATGCGTAGTGAATGG
ABL 3 (27172) F	AGTAGGGGGACCAAGAGGGAGGT
BCR B4 (4099) R	CTGCAGGGATGTCCAGAGCTACA
BCR B3 (3000) R	TTTTCCACATCACCCACATGGTC

Sequences of primers used to sequence over the reciprocal breakpoint in positives:

Patient ID	Primer	Primer sequence (5'-3')
26231	ABL 4 (9361)F	CACACCAGTTGCTCCATTTGGTC
19715	ABL 2 (29501) F	TTAGCCAGGCATGGTGGCGTGCA
11129	BCR B3 (844) R	TAGAGCCCCGGAGACTCATCATC
10258	ABL 1 (17143) F	GATCTGCTGGGGAACTCCTTCCT
6875	BCR (2613) R	CTCTTTGCCCCATAGTACAGCGG
10887	BCR B3 (844) R	TAGAGCCCCGGAGACTCATCATC
F22666	BCR B3 (1909) R	ACCCTACACTTGGAATGGATGAA
F17237	BCR B3 (3000) R	TTTTCCACATCACCCACATGGTC
F14472	BCR B3 (1053) R	TGTGGAGTGTTTGTGCTGGTTGA
F11751	BCR B3 (1053) R	TGTGGAGTGTTTGTGCTGGTTGA

p210 CML reciprocal breakpoint primers

Primers used to amplify reciprocal breakpoint bands:

Patient ID	ABL forward primer	BCR reverse primer
NP241	ABL 4B F	BCR B4 (4099) R
E656	ABL5B F	BCR B4 (4099) R
1716	ABL 4 (7792) F	BCR B4 (4099) R
1349	ABL1 (20970) F	BCR B4 (4099) R
1578	ABL 4 (7792) F	BCR B4 (4099) R
1103	ABL 4 (7792) F	BCR B4 (4099) R
1810	ABL 2D F	BCR B4 (4099) R
1338	ABL 2D F	BCR B4 (4099) R
1569	ABL 2B F	BCR B4 (4099) R
1460	ABL 2C F	BCR B3 (3000) R

230	ABL 4 F	BCR B3 (3000) R
248	ABL 1 F	BCR B3 (3000) R
266	ABL 2 F	BCR B3 (3000) R
316	ABL 4B F	BCR B3 (3000) R
403	ABL 5A F	BCR B3 (3000) R
440	ABL 2B F	BCR B3 (3000) R
446	ABL 1A F	BCR B3 (3000) R
483	ABL 4A F	BCR B3 (3000) R
495	ABL 1 (10441) F	BCR B3 (3000) R
524	ABL 4D F	BCR B3 (3000) R
545	ABL 1C F	BCR B3 (3000) R
282	ABL 3C (2) F	BCR B3 (3000) R
286	ABL US (36126) F	BCR B3 (3000) R
287	ABL 3D F	BCR B3 (3000) R
291	ABL 4C (3) F	BCR B3 (3000) R
300	ABL 4B F	BCR B4 (4099) R
317	ABL 5B F	BCR B3 (3000) R
375	ABL 3B F	BCR B3 (3000) R
386	ABL 1D F	BCR B3 (3000) R
436	ABL 2B F	BCR B4 (4099) R
389	ABL 3D F	BCR B3 (3000) R
470	ABL 2A F	BCR B4 (4099) R

Sequences of new primers used to amplify reciprocal breakpoints:

Primer	Primer sequence (5'-3')
ABL 4 (7792) F	ACCCCAGCATACCCATTCCTCTT
ABL1 (20970) F	GCACCTCTGCCACTCTGTCCTTT
ABL 1 (10441) F	TTTtagccttggcacaccaggtca
ABL US (36126) F	AACACTCATCGTGCAGCTGTTC

Primers used to sequence over the reciprocal breakpoint in positives:

Patient ID	Primer	Primer sequence (5'-3')
1716	ABL 4 (9361)F	CACACCAGTTGCTCCATTTGGTC

1103	ABL 4 (8980) F	GTGCTTTTTCTTCAGGGCACCAG
1569	BCR B3 (2282) R	TGGTCTGCTCTCCCTCCGTAAA
1460	BCR B3 (1808) R	ACTGCAGATGAACCCAAGGGGGA
248	BCR B3 (485) R	CTGTTTTGCATTCACTGTTGCAC
266	BCR B3 (707) R	GACACTGGCTTACCTTGTGCCAG
316	BCR B3 (844) R	GATGATGAGTCTCCGGGGCTCTA
403	BCR B3 (844) R	GATGATGAGTCTCCGGGGCTCTA
483	BCR B3 (485) R	CTGTTTTGCATTCACTGTTGCAC
524	BCR (2613) R	CTCTTTGCCCCATAGTACAGCGG
286	BCR B3 (2282) R	TGGTCTGCTCTCCCTCCGTAAA
375	BCR (2613) R	CTCTTTGCCCCATAGTACAGCGG
386	BCR (1659) R	AGCCGAGCTTGTGCCACTGCATT
436	BCR (2613) R	CTCTTTGCCCCATAGTACAGCGG
389	BCR (1659) R	AGCCGAGCTTGTGCCACTGCATT
470	BCR B3 3000 R	TTTCCACATCACCCACATGGTC

p190 ALL reciprocal breakpoint primers

Primers used to amplify reciprocal breakpoint bands:

Patient ID	ABL forward primer	BCR reverse primer
Ital 1	ABL 5 (49) F	BCR bt2 3B R
Ital 2	ABL 4A F	BCR bt3 (5691) R
Ital 5	ABL 2D F	BCR bt3 (5691) R
Ital 6	ABL 1B F	BCR bt2 (21867) R
Ital 10	ABL 5 (11708) F	BCR bt3 (26236) R
21296	ABL 1 (20970) F	BCR bt3 5A R
23827	ABL 2D F	BCR bt3 (17384) R
25169	ABL 4A F	BCR bt3 (17384) R
25373	ABL 4D F	BCR bt3 5B R
25654	ABL 4 (7792) F	BCR bt3 (26236) R
9463	ABL 4D F	BCR bt2 (21867) R
20763	ABL 2 (9742) F	BCR bt3 (26236) R
44095	ABL 2 (9742) F	BCR bt 2 3B R
Ital 4	ABL 2 (20977) F	BCR bt 3 (8881) R

28852	ABL 4A F	BCR bt3 (17384) R
28835	ABL 2C F	BCR bt3 5A R
19635	US ABL (34129) F	BCR bt3 (5691) R
28624	ABL 4A F	BCR bt3 5B R
27523	ABL 1 (4679) F	BCR bt3 (17384) R
24350	ABL 4C (3) F	BCR bt 3 (8881) R
Italian 9	ABL 4 (7792) F	BCR bt3 5A R
P747	ABL 4A F	BCR bt3 (5691) R
P653	ABL 4C (3) F	BCR bt3 5A R
44240	ABL 1 (23775) F	BCR 5A R
11491	ABL 1 (4679) F	BCR bt 3 (8881) R

Sequences of new primers used to amplify reciprocal breakpoints:

Primer	Primer sequence (5'-3')
ABL 5 (49) F	CTT TCC CCT GGA GAC GTC AAA GA
ABL 5 (11708) F	CAG TGG GAA TGC GTA GTG AAT GG
ABL 2 (9742) F	CGTGGAGTAAGGCAGGAAAGCAT
ABL 2 (20977) F	CACATGGCAGCTCACTCAGGAAT
US ABL (34129) F	CAGCTGACCTCTGCTGTGACTCC
ABL 1 (4679) F	GGAGCTTTAGGTGGAATCGGAGAA
ABL 1 (23775) F	ACATTGCACAAGGGGGAAAGAAA
BCR bt3 (5691) R	TCT ACA CTA ACC AAG CGC CAC CA
BCR bt2 (21867) R	GTG CCT AAT AAG GGG CCC AAG AC
BCR bt3 (26236) R	CAT GAC CCT CCA GAC TTG GCT TT
BCR bt3 (17384) R	CCAGGTAGGCCTGGTGTGTTGGTAG
BCR bt 3 (8881) R	CTCAGGAAGCTCGCTCCAAGAAC

Primers used to sequence over the reciprocal breakpoint in positives:

Patient ID	Primer	Primer sequence (5'-3')
Ital 1	ABL 5 (1285) F	CATTCACCGAGGCTCACTGTACC
Ital 2	BCR bt3 (2516) R	AGGGCCATTCCCAAATCTCTGAT
Ital 6	BCR bt2 (19958) R	AGACACAGAGGCCATTCCCTGAC
Ital 10	ABL 5 (13568) F	CTGCCAGGCTAGAGTGCAGAGAC

23827	ABL 2 (18967) F	CAGAGAGGCCAGGACACACACAT
25169	ABL 4 (12151) F	GGGTTTCACCATGTAGCCAGGA
25373	BCR bt3 (26294) R	ATCCCCTTGGCCTCCTTTAACAA
25654	BCR bt3 (23077) R	TCAAGAGAAGAAAGGGGCACTGG
9463	ABL 4 (19877) F	CTCGGGAGGCTGAGACAGGATAA
20763	BCR bt3 (23706) R	CGCCCAGCCTCTAAAAGAACTGA
44095	ABL 4 (11296) F	TTCAGCTATTTTTTCGGCGTTGCT
Ital 4	BCR bt 3 (6545) R	CCCTGAAGGAACGCTATGTGTCC
28852	BCR bt3 (12913) R	TATCACCATGCCACTGCAGACAC
28835	BCR bt3 (18274) R	AGAATGCGACCAACAGAGCCTTC
19635	BCR bt3 (4983) R	CAACCCCAAGATGACCCAAATGT
28624	ABL 4 (14732) F	GGAGCAATTCAAACAAGCGAGATG
Italian 9	Cloned	M13 forward and reverse
P747	ABL 4 (11609) F	GCAGGTTGGAAAGCTGCCATTAC
44240	ABL 1 (25851) F	ATGGGGTTGGAAGTCCCTTGGAT
11491	BCR bt3 (8881) R	CTCAGGAAGCTCGCTCCAAGAAC

Construct Primers

pGlow forward oligo

5'gggtttgtccagctctgtagcactgtgggtacccccctgttgacaattaatcatcgaactagttaactagaattccacagtggttagtactcca
ctgtctggctgtacaaaaaccgaattcaggactgcaattcatgaaa-3'

RAG oop

5'gggtttgtccagctctgtagcactgtggaattcatcagcaggtggaagagggactggattccaaagttctcaatgctgcttctgttctt
gaatggggggctgttgacgacgacatggctcgattggcgcgacaagttgctgcgattctaccaataaaaaacgcccggcgga
accgagcgttctgaacaaatccagatggagttctgaggtcattactggtttaaccacagtggttagtactccactgtctggctgtacaaa
aacccc 3'

Primers for Professor Nadel constructs

Insert PCR and sequencing primers (5' to 3'):

RAG forward: CAACTTCTGGTCCGGTAACGTGCTG

RAG reverse: CGATGCCATTGGGATATATCAACGGTGG

Primers for amplifying RS2 and RS3 inserts:

RS2 MluI F: 5'-*ACGCGTGGAAAGCAGGGAGGGAAGCTGCT*-3'

RS2 SalII R: 5'-*GTCGACCCGCCACAAACCCCAAGGCAG*-3'

RS3 MluI F: 5'-*ACGCGTCACAAACCCCAAGGCAGATCTAG*-3'

RS3 SalII R: 5'-*GTCGACCCCAAAGCAGGGAGGGAAGCT*-3'

MluI SalI tagged BCR primer (italics = tagged restriction sites):

p190	Primer name	Primer Seq (5' to 3')
Italian 1	BCR int 2 (18020) F MluI	<i>acgcgt</i> CACCCACTAGTTGTCCCTGCTCA
	BCR int 2 (18373) R SalI	<i>gtcgac</i> TTACAGGAGTCCCCTCTGCCTTG
Italian 2	Na	na
Italian 6	BCR bt2 (19533) F MluI	<i>acgcgt</i> TCCTGGAGCATTGAGTCATCCTG
	BCR bt2 (19781) R SalI	<i>gtcgac</i> GTAGATCCTGTTGGTCCCCATGC
25169	BCR bt3 (15358) F MluI	<i>acgcgt</i> gttggtgctgggttgatccttat
	BCR bt3 (15687) R SalI	<i>gtcgac</i> caggccaggagttcgagactagc
25654	BCR bt3 (22728) F MluI	<i>acgcgt</i> CTCCAAAAGACGTGTGCTGCTGT
	BCR bt3 (23077) R SalI	<i>gtcgac</i> ACAGCAGCACACGTCTTTTGGAG
p210 ALL		
P851	Na	na
F17730	BCR b2/b3 (368) F MluI	<i>acgcgt</i> ctcctcagatgctctgtgccttg
	BCR b2/b3 (744) R SalI	<i>gtcgac</i> aaatcaaccatccggtggacact
F24559	BCR b2/b3 (368) F MluI	<i>acgcgt</i> ctcctcagatgctctgtgccttg
	BCR b2/b3 (744) R SalI	<i>gtcgac</i> aaatcaaccatccggtggacact
p210 CML		
266	na	na
386	BCR b2/b3 (1053) F MluI	<i>acgcgt</i> atgccttctgggtgtggaattgt
	BCR b2/b3 (1766) R SalI	<i>gtcgac</i> cattccccaacaatccctggttaa

SpeI SacII tagged BCR primers ((italics = tagged restriction enzyme sites):

p190	Primer name	Primer seq (5' to 3')
Italian 1	BCR int2 (18020) F SpeI	CACCCACTAGTTGTCCCTGCTCA
	BCR int2 (18373) R SacII	<i>ccgcgg</i> TTACAGGAGTCCCCTCTGCCTTG
Italian 2	Na	na
Italian 6	BCR bt2 (19533) F SpeI	<i>actagt</i> TCCTGGAGCATTGAGTCATCCTG
	BCR bt2 (19781) R SacII	<i>ccgcgg</i> GTAGATCCTGTTGGTCCCCATGC
25169	BCR bt3 (15358) F SpeI	<i>actagt</i> gttggtgctgggttgatccttat

	BCR bt3 (15687) R SacII	ccgcggcaggccaggagttcgagactagc
25654	BCR bt3 (22728) F SpeI	<i>actagt</i> CTCCAAAAGACGTGTGCTGCTGT
	BCR bt3 (23077) R SacII	ccgcggACAGCAGCACACGTCTTTTGGAG
<u>p210 ALL</u>		
P851	Na	na
F17730	BCR b2/b3 (368) F SpeI	<i>actagt</i> ctcctcagatgctctgtgccttg
	BCR b2/b3 (744) R SacII	ccgcggaaatcaaccatccggtggacact
F24559	BCR b2/b3 (368) F SpeI	<i>actagt</i> ctcctcagatgctctgtgccttg
	BCR b2/b3 (744) R SacII	ccgcggaaatcaaccatccggtggacact
<u>p210 CML</u>		
266	Na	Na
386	BCR b2/b3 (1053) F SpeI	<i>actagt</i> atgccttctgggtgtggaattgt
	BCR b2/b3 (1766) R SacII	ccgcggcattccccaacaatccctggtaa

MluI SalI tagged ABL primer (italics = tagged restriction sites):

<u>p190</u>	Primer name	Primer seq (5' to 3')
Italian 1	na	na
Italian 2	ABL 4 (10777) F MluI	<i>acgcgt</i> CCAGCCAAATGAGACCAGGAAGT
	ABL 4 (11325) R SalI	<i>gtcgac</i> TGCAGTGACTGAAGTCCAACAAGC
Italian 6	na	na
	na	na
25169	na	na
25654	na	na
<u>p210 ALL</u>		
P851	ABL 5 (4091) F MluI	acgcgtGTCTGTTGCAAATGGCCTGTGAC
	ABL 5 (4520) R SalI	gtcgacGCTGGGTGAACTCAGATGTGGAA
F17730	na	na
F24559	na	na
<u>p210 CML</u>		
266	ABL 2 (24179) F MluI	<i>acgcgt</i> taggctggagtgcaatgggtctat
	ABL 2 (24680) R SalI	<i>gtcgaccaca</i> agggaattctaggggggtga
386	na	na

SpeI SacII tagged ABL primer (italics = tagged restriction sites):

<u>p190</u>	Primer name	Primer seq (5' to 3')
Italian 1	Na	na

Italian 2	ABL 4 (10777) F SpeI	<i>actagt</i> CCAGCCAAATGAGACCAGGAAGT
	ABL 4 (11325) R SacII	<i>ccgcgg</i> TGCAGTGAAGTCCAACAAGC
Italian 6	Na	na
25169	Na	na
25654	Na	na
<u>p210 ALL</u>		
P851	ABL 5 (4091) F SpeI	<i>actagt</i> GTCTGTTGCAAATGGCCTGTGAC
	ABL 5 (4520) R SacII	<i>ccgcgg</i> GCTGGGTGAAGTCAAGATGTGGAA
F17730	Na	na
F24559	Na	na
<u>p210 CML</u>		
266	ABL 2 (24179) F SpeI	<i>actagtaggctggagtgcaatggtgctat</i>
	ABL 2 (24680) R SacII	<i>ccgcggcacaaggaattctagggggtga</i>
386	Na	na

Bubble primers

Primer Name	Primer Sequence (5' to 3')
BUB-T	AAGGATCCTAGTCTAGCTGTCTGTCTGCGAAGGTAAGGAACGGACGAGCACTGAG
BUB-B	CTCAGTGCTCGTAGTAATCGTTCGCACGAGAATCGCAAGATCTAGGATCCTT
NVAMP1	TGC TCG TAG TAA TCG TTC GCA C
NVAMP2	GTT CGC ACG AGA ATC GCA AGA T
PDAl12R3	AAG TTG TGT GCA AGG GAA AAG GG
PDAl12R4	AGG TTA CCC CAT GGA ACT TAC CA

Simplex LR PCR primers for gDNA breakpoint detection in FIP1L1-PDGFR

Patient ID	Primer name	Primer Sequence (5'-3')
E1080	FIP Intron 13 (63966) F1	ACA GCT GGT CAG GTG GCT GTT C
	FIP Intron 13 (64195) F2	TGC TTG CAA GGG TGG GTC TAA AT
E905	FIP Intron 11 (47665) F1	CAG AAA TGG CAA ACT GGT TCC TT
E759	FIP Intron 11 (47665) F1	CAG AAA TGG CAA ACT GGT TCC TT
E939	FIP Intron 11 (47665) F1	CAG AAA TGG CAA ACT GGT TCC TT
E1025	FIP Intron 15 (75072) F	CCA CTG CAC CTG GCA TGA TAG AC
E1131	FIP Intron 11 (47665) F1	CAG AAA TGG CAA ACT GGT TCC TT
E1336	FIP Intron 11 (36955) F	CCT CCA GAA AAG CCA ATT CAA GC
E1401	FIP Intron 11 (47665) F1	CAG AAA TGG CAA ACT GGT TCC TT
E1424	FIP Intron 11 (47665) F1	CAG AAA TGG CAA ACT GGT TCC TT

Multiplex LR PCR primers for gDNA breakpoint detection in FIP1L1-PDGfra

Mix	Primer names	Sequence (5'-3')
1	FIP Intron 5 (11403) F	TCTTGGTCATTAACCACCCACCA
	FIP Intron 10 (22850) F	CTCGGCCAGATCACTTACCTTTTG
	FIP Intron 11 (40225) F	AAATGTGGAGTTTGAGGCATTGCT
	FIP Intron 13 (56273)	GTGGCGAGTTGCTCATATCTAGGG
	FIP Intron 15 (70370) F	TTGGAACATGCCTTTCTCCACCT
2	FIP Intron 9 (15351) F	AAATAGGGCCATGATGAGCAGGA
	FIP Intron 10 (34060) F	ATGCTGGGTTTCCAACGTTCTGT
	FIP Intron 11 (44469) F	GGG TTG TGA AAC AAT TTG GTG GTC
	FIP Intron 13 (60580) F	AGGGGAAAGAAGAGGGACGTTTG
	FIP Exon 16 (75275) F	GACACCAGCAAGCAGTGGGACTA
3	FIP Intron 9 (20848) F	CTTTCCTCATGGGAAAGGAGAG
	FIP Intron 11 (38554) F	TGAACCTAAAAGGGCCAGACAACA
	FIP Intron 12 (49317) F	GTG GGC CTG AGT TTT GAT TGT GA
	FIP Intron 14 (65181) F	CCCGATATCAAAACAGGGCAAAG
	FIP Intron 16 (80314) F	CCCCTGCCCACACAGATGTAAT
Reverse	PDA intron 12 R1	TGTGCAAGGGAAAAGGGAGTCTT

Breakpoint specific primers for MRD in gDNA

Patient ID	PCR	Primer name	Primer sequence (5'-3')
E243	First	FIP Intron 13 (49895) F	GAA TGT CTG ACC TGC TGC CAT GT
	Nested	FIP Intron 13 (49974) F	TCA ACT TTG GTT GGG TGA CAA AGA
E359	First	FIP Exon 11 (36955) F1	CCT CCA GAA AAG CCA ATT CAA GC
	Nested	FIP Exon 11 (37004) F2	GAG GGC CGA ATC ACC TGA TCT AA
E370	First	FIP Intron 10 (31242) F1	TGA ATT TCC TCC CGT AGC TTG GA
	Nested	FIP Intron 10 (31324) F2	TTT GTT CCG TTG CTT GTG AGG AA
E591	First	FIP Intron 15 F1	TGG ATC TTC CCG ATA CTG TGG GTA
	Nested	FIP Intron 15 F2	TGA CAC ATT TTT GGT GGG CAT TT
E614	First	FIP Intron 16 (75917) F1	TCA GTG CCT GGC TTC AGA GTT TG
	Nested	FIP Intron 16 (76002) F2	TCA AGT GAT CCT CCT GGC CTA GC
E630	First	FIP Intron 11 (40568) F1	TGG AGC CCA ACC TGG TCT CTA AA
	Nested	FIP Intron 11 (40845) F2	GAA GCA GTT TCA GTG AGT GCA GGA
E759	First	FIP Intron 10 (35023) F1	GAT GGT GTC CTT TGG GTT TCT GC
	Nested	FIP Intron 10 (35259) F2	TCC CAC TCC GCC ACT TAT CAC TT

FIP1L1-PDGfra RQ-PCR primers

All sequences 5' to 3'.

cDNA RQ-PCR control gene probe and primers:

ABL EAC primers and probe.

cDNA RQ-PCR probe and primers from Jovanovic *et al.*, 2007:

PDGfra R: CAAGACCCGACCAAGCACTAGT

PDGfra probe: FAM-CTGCCTTATGACTCAAGATGGGAGTTTCCAA-Blackhole1

FIP1L1 Exon 12 F: GGGCAAATGAGAACAGCAACA

cDNA RQ-PCR primers designed by Joannah Score:

FIP1L1 Exon 14 F: GCTCCACCTCCATCTCTTAT

FIP1L1 Exon 16 F: CATCTTCCTGGTTCTGCTC

gDNA RQ-PCR control gene probe and primers:

Alb F: TGAAACATACGTTCCCAAAGAGTTT

Alb R: CTCTCCTTCTCAGAAAGTGTGCATAT

Alb probe: FAM-TGCTGAAACATTACCTTCCATGCAGA-Blackhole 1

gDNA RQ-PCR LNA PDGfra probe and primers:

PDGfra LNA F: ACTAGTGCTTGGTAAGTTCC

PDGfra LNA R: CCTATAAATTGTAAAGTTGTGTGC

PDGfra LNA probe: FAM-cct+Ccc+Aag+Act+Ccct-Blackhole1

gDNA RQ-PCR FIP1L1 forward patient specific primers:

E43 F: TTAAAAAGCCTGCCTATCAG

E243 F: CAATTTTAGCAAACACCTC

E359 F: AATTATATTGAACCGTGCTAATG

E370 F: TGTTAGTTTTTCCTTCTAACAGACA

E591 F: CATCTTCCTGGTTCTGCTC

E606 F: AGGGTTGATAGAATTGTGAAGAT

E614 F: GGTTTTCCTATTAGTCTTTCCTG

E630 F: CAATGTTAAATCACGAATACATTT

E759 F: CACAGGATTCTTATTCCTATTTCT
E1025 F: ATTTTGTATGGATGCTGTAGAAG
E1279 F: TTGATCAAGTGCTCTGAATG
E975/1131 F: ATTCCTATTTTCTGAGTCCTTTT
E1961 F: TTTGAGGACTAAAGTTTGAAGG
E2002 F: AAGAAAGTGTTGATAGCACATCTA
E2378 F: CACCAGTGATTTTACCACCT

PCR primers for making plasmids for absolute RQ-PCR:

E43 F: TTGGCAGAAGCTTGGCCTGTATT
E243 F: GAATGTCTGACCTGCTGCCATGT
E359 F: CCTCCAGAAAAGCCAATTCAAGC
E370 F: TGAATTTCTCCCGTAGCTTGGA
E591 F: TGGATCTTCCCGATACTGTGGGTA
E606 F: TCATGGGTGGCAGTATGGTGTTT
E614 F: TCAGTGCCTGGCTTCAGAGTTTG
E630 F: TGGAGCCCAACCTGGTCTCTAAA
E1961 F: TTGCCATGCCAACTTTCTGATGT
E2002 F: CAGAGTGTTGCTGTAACTGGTTGC
E2378 F: GTGGGCCTGAGTTTTGATTGTGA

PDGFRA (1330) R: GGTATCTTTGCCAGCCCTCTTCA

KIF5B-PDGFRA primers

	Primer name	Primer Sequence (5'-3')
cDNA	KIF5B Exon 19 F1	TGA ACA GCA GAT CCA GAG CCA TA
	PDA Exon 14 R1	TGG CCC CAG GTG AGT CAT TAT CT
gDNA	KIF5B Exon 22 F2	CTG CGC AAA CTC TTT GTT CAG GA
	PDAI12R3	AAG TTG TGT GCA AGG GAA AAG GG

Appendix III

Position of forward breakpoints

Reference sequences used:

NM_004327 BCR from ATG

NM_007313 ABL from ATG

p210 ALL

Patient ID	BCR bpt	ABL bpt
P682	110061	45934
P851	110042	128608
22251	110640	80502
26231	111313	104250
24375	110146	117901
26219	109487	11224
19715	109368	60782
6875	110950	39021
10887	109253	138274
10258	110172	17477
11129	109211	88554
11128	109251	108853
F22666	110271	37683
F17685	110397	133833
F17730	109149	113725
F18544	108792	99563
F17237	111291	46309
F16604	109033	23724
F15010	109203	91522
F22757	110138	87352
F14472	109329	127029
F11329	109120	98551
F11751	109477	32662

F24559	109149	34587
F20099	111031	28149

p210 CML

Patient ID	BCR bpt	ABL bpt
NP241	108923	100389
E656	109473	137729
1716	110672	104268
1349	111413	25318
1578	109263	105366
1103	111574	103666
1810	108688	48293
1569	110255	40196
1338	109540	47245
1460	110253	31164
230	109832	121996
248	108785	6941
266	109131	55259
316	109063	96194
403	108984	126382
440	109092	38909
446	109294	22454
483	108807	107887
495	108754	17597
524	110884	114565
545	108966	1292
282	109918	60312
286	110614	-3162 (upstream ABL)
287	109483	76292
291	109766	88366
300	111520	97218

317	108803	135075
375	110977	83626
386	109999	12188
436	110767	39959
389	109758	75580
470	111252	51756

From Hammersmith		
S1	109711	82412
S2	108922	36293
S9	108995	-6872 (upstream ABL)
S10	109059	91004
S11	109880	74312
S12	109195	56535
S13	109741	91213
S14	111042	122256
S15	109345	85519
S16	109736	130755
S18	109698	137794
S19	109374	106169
S20	110116	108325
S24	109002	129000
S29	109268	48489
S30	109201	42796
S31	108783	45247
S33	109428	17084

p190 ALL

Patient ID	BCR bpt	ABL bpt
------------	---------	---------

Italian 1	37687	125796
Italian 2	45801	105587
Italian 4	49688	63186
Italian 5 & 8	48056	46024
Italian 6	39135	3202
Italian 9	62710	105351
Italian 10	67879	138187
P653	64986	89570
747	43865	106383
21296	60916	22505
23827	57573	50072
25169	58970	107002
25373	69389	115708
25654	66293	105216
9463	36737	114421
20763	66845	42708
44095	41572	42364
44240	61908	26106
27523	57016	7982
28624	72229	109358
28835	61397	32232
28852	56004	107680
11941	52114	8511
19635	48131	-3752 (upstream ABL)
24350	49755	88277

Forward and reciprocal breakpoint sequences

p190 ALL

Ital 1	
BCR	<u>TCAGCTGTGGGACTCGTGTGGGGACAAGCTGAACAGTGTGCT. GATCTCTGATCACCACCTCAAGCCACAAGTGACTGGACA</u>
BCR ABL	<u>TCAGCTGTGGGACTCGTGTGGGGACAAGCTGAACAGTGTGTTAGCAGCCTAAAATGGTGGCTTTACCAGTTTCTGTAATG</u>
ABL	<u>CCTAGCAGAAAAGTAAGCTGTTGTTACTTTAAGCAAATATTTAGCAGCCTAAAATGGTGGCTTTACCAGTTTCTGTAATG</u>
ABL BCR	<u>CCTAGCAGAAAAGTAAGCTGTTGTTACTTTAAGCAAATATTTAGATCTCTGATCACCACCTCAAGCCACAAGTGACTGGAC</u>
Ital 2	
BCR	<u>ATGCCATGTGGCTTGCAAAGCTAAGATATTCATTAGCTGGCTCTTCCAGGAAAAGTTTGACCCCTGCTTTAGGATAAAA</u>
BCR ABL	<u>ATGCCATGTGGCTTGCAAAGTGAAGATATTCATTAGCTGGTTTCTGTTTCTCCAGGCATTCTGGTATGGACATTTATAC</u>
ABL	<u>TACAGCAGGAGGTGTTTCTGCTTCTACTTTGCTGTGA. . .CTTCTGTTTCTCCAGGCATTCTGGTATGGACATTTATAC</u>
ABL BCR	<u>TACAGCAGGAGGTGTTTCTGCTTCTACTTTGCTGTGACTTCTCTTCCAGGAAAAGTTTGACCCCTGCTTTAGGATAAAA</u>
Ital 5	
BCR	<u>CACCTGGGACATGAAATTAGACATCAACATCTTTTGATGTTGTTCCATAGATGCTTGAGGTTTTATCCATTTTCT</u>
BCR ABL	<u>CACCTGGGACATGAAATTAGACATCAACATCTTTTGATGTTTGGGTTTGTAATAGAAGCTACTGTCCCCCTGATCA</u>
ABL	<u>AAGAGCTGGGACACTGTTGGCCATTAGGATTCCTGTTCTTTTGGGTTTGTAATAGAAGCTACTGTCCCCCTGATCA</u>
ABL BCR	MLPA failed
Ital 6	
BCR	<u>GAGGGGCTGTTCTGCCACCTGCCCTTTGAGGCTGTGTGGGCACGGTTCCTCTAGAGCTCCATGTCTAGCAGAGCATGCG</u>
BCR ABL	<u>GAGGGGCTGTTCTGCCACCTGCCCTTTGAGGCTGTGTGGGAGTGAGATCTTACCTCTACAGAAATTTAAAAAATTAGCCA</u>
ABL	<u>ATCACCTGAGGTCAGGAGTTCGAACCCAGCTGTGCAATCTAGTGAGATCTTACCTCTACAGAAATTTAAAAAATTAGCCA</u>
ABL BCR	<u>ATCACCTGAGGTCAGGAGTTCGAACCCA. . . TGGGCACGGTTCCTCTAGAGCTCCATGTCTAGCAGAGCATGCG</u>
Ital 9	
BCR	<u>TGTGAGCTGTTCTAGCAATTTTTTTTTTTTTTTTAAGA. . TGGAGCCTCGCTCTGTTGCCAGGCTGGAGTGCACT</u>
BCR ABL	<u>TGTGAGCTGTTCTAGCAATTTTTTTTTTTTTTTTAAGACTGGGACTACGGGCACACACCACCATGCCCGGCTACTTTT</u>
ABL	<u>CCAGGTTCAAGCAATTCATTGCCTCAGCCTCCCCAGTAGCTGGGACTACGGGCACACACCACCATGCCCGGCTACTTTT</u>
ABL BCR	<u>CCAGGTTCAAGCAATTCATTGCCTCAGCCTCCCCAGTAGCTGGTGGAGCCTCGCTCATGTTGCCAGGCATGGAGTGCA</u>
Ital 10	
BCR	<u>GGATAAAGCTTTGCAGGAAGGATGTGGCTGATGAGACCATCAACCTCAGTCCTGCCAAGGGTGAGGCGGCCAGAGAGGT</u>
BCR ABL	<u>GGATAAAGCTTTGCAGGAAGGATGTGGCTGATGAGACCATGCCAGTTAGTGATTTTTAAATGTAGTTTCCTTAAATGAG</u>
ABL	<u>CCCAATGTTCTAGGGTTACAGGCATGAGCTACTGTGCCTGGCCAGTTAGTGATTTTTAAATGTAGTTTCCTTAAATGAG</u>
ABL BCR	<u>CCCAATGTTCTAGGGTTAC. CAGTCCTGCCAAGGGTGAGGCGGCCAGAGAGGT</u>
25169	
BCR	<u>TCACTGTAGCCTCAGCCTCCTGGGCTCAAGTGATCCTCCACCTCAGCCTCCCAAGTAGCTGAAAACCTACACTTGACCAC</u>
BCR ABL	<u>TCACTGTAGCCTCAGCCTCCTGGGCTCAAGTGATCCTCCACCATTCATATTGCTTTTTCTATAGATTGTCTTTTCCG</u>
ABL	<u>CCCTTTCAGCAGTCTTACCCCACTTCCAGCCCCAGCGACCATTC. . TTTGCTTTTTCTATAGATTGTCTTTTCCG</u>
ABL BCR	<u>CCCTTTCAGCAGTCTTACCCCACTTCCAGC. AAGTAGCTGAAAACCTACACTTGACCAC</u>
9463	
BCR	<u>CTTCAGGCTCCTTTTGAAGGTCAATTGTTTTCTTAACAGATGGTTATTTGGACTCTTACAAGTTCTCTGTAAGTTTGTA</u>
BCR ABL	<u>CTTCAGGCTCCTTTTGAAGGTCAATTGTTTTCTTAACAGAGCAATCCAGGGCTTTAATATCCTTTTGTAGGAATTCCA</u>
ABL	<u>GAAACTGCGTCTCAAAAAAAGAAAAGAAATTAGAAACTCAATCCAGGGCTTTAATATCCTTTTGTAGGAATTCCA</u>

ABL BCR	GAAACTGCGTCTCAAAAAAAAAAGAAAAGAAATTAGAACT . . TTATTTGGACTCTTACAAGTTCTCTGTAAGTTTGTA
21296	
BCR	<u>CTCTAGGCTCCCCTCTTGTTCAGACCCAGACATCTCCTAAATGGAGCCCAAGTCCTTGGCCTGAAGCCTGAGGCTCCCC</u>
BCR ABL	<u>CTCTAGGCTCCCCTCTTGTTCAGACCCAGACATCTCCTACATACTTATCTTTCTGTGAACCATCTCATCAGGTTCTTTG</u>
ABL	<u>CCCGCCGGTTGAACGTTTTTAACATGTTTAGGGGAGATACATACTTATCTTTCTGTGAACCATCTCATCAGGTTCTTTG</u>
ABL BCR	MLPA = ABL 1b & EXSOC2 DELETION 07.01.26
23827	
BCR	<u>AGTTGGGGATCCTCCTTCTTAGCTGTGGAGCTCAGGGTATGTTGGTTCACGTCCAAACCTTCAACCTTCATCTGTAAAA</u>
BCR ABL	<u>AGTTGGGGATCCTCCTTCTTAGCTGTGGAGCTCAGGGTATCTGAAGTGTATTGCTGACCTTTCCCATGTTCAAAGTTCC</u>
ABL	<u>TATTGCTGGAGGAACTAGCTCTGTCCTGTTGGTGGGACTGAACTGTATTGCTGACCTTTCCCATGTTCAAAGTTCC</u>
ABL BCR	<u>TATTGCTGGAGGAACTAGCTCTGTCCTGTTGGTGGG . GTTGGTTCACGTCCAAACCTTCAACCTTCATCTGTAAAA</u>
25373	
BCR	<u>TCTTAAGATCGTCTTCAGATAAACAGCAAAACCGTTGGTTAAAAGCTGGTCTAACCCACCTTATTTTGTGTTGTTTGT</u>
BCR ABL	<u>TCTTAAGATCGTCTTCAGATAAACAGCAAAACCGTTGGTTTACAAGACTATTTTTCAGACTTTTTCAGACTTTTAGT</u>
ABL	<u>ATACAGGGGAAAGGATTGCTGTTTTTTATTACCAGATTTTACAAGACTATTTTTCAGCTTTTACAAGACTTTTAGT</u>
ABL BCR	<u>ATACAGGGGAAAGGATTGCTGTTTTTTATTACCAGA . TTAAGAGCTGGTCTAACCCACCTTATTTTGTGTTGTTTGT</u>
25654	
BCR	<u>AGCCATGACAGGGTGGGGTTTTGATGTCCGCCTCTGTGAAAACCTCCACCACAGCTGGCCGAGCCACCGGATGTTTGATC</u>
BCR ABL	<u>AGCCATGACAGGGTGGGGTTTTGATGTCCGCCTCTGTGACTTTTTTTTTTTTTTTTTTTTGGAGACGGAGCCTCACTTT</u>
ABL	<u>AACCAGCCAAATGAGACCAGGAAGTGAAGAAAATACCTTTTTTTTTTTTTTTTTTTTGGAGACGGAGCCTCACTTT</u>
ABL BCR	<u>AACCAGCCAAATGAGACCAGGAAGTGAAGAAAATA . . . TCCACCACAGCTGGCCGAGCCACCGGATGTTTGATC</u>
20763	
BCR	<u>CTATGAATCTGACTACTGTAGGTACCTCATATGAGTGAATCATAACAATATTTGTCTTTTGTGCTGGCTTATTCCAC</u>
BCR ABL	<u>CTATGAATCTGACTACTGTAGGTACCTCATATGAGTGAACCATTTGCGAGAATTCCTTCCAGCCGTCTGTGAGAAGGC</u>
ABL	<u>CCAGAGCAGGGAGCCGATGAAAGGGGGACCCAGAACATTCATTCTGCAGAATTCCTTCCAGCCGTCTGTGAGAAGGC</u>
ABL BCR	<u>CCAGAGCAGGGAGCCGATGAAAGGGGGACCCAGAACATTCATAACAATATTTGTCTTTTGTGCTGGCTTATTCCAC</u>
44095	
BCR	<u>TCAGAGCAGGGGAAGTGGTTCCCGAGGAAGGAGGCAGAGGAGTGTAGATTTTGTAGAGAAGGCTAAGAGAGGGTGGGATGTG</u>
BCR ABL	<u>TCAGAGCAGGGGAAGTGGTTCCCGAGGAAGGAGGCAGAGGATCTTCAGATATCTAAAAGAACTTTGTAACACAGCGTCAGCC</u>
ABL	<u>GAGGGGAAGGGATGTTTCTGTGTTGCTCTTGTGAGTATCTTCAGATATCTAAAAGAACTT . GTAACACAGCGTCAGCC</u>
ABL BCR	<u>GAGGGGAAGGGATGTTTCTGTGTTGCTCTTGTGAGTAGTGTAGATTTTGTAGAGAAGGCTAAGAGAGGGTGGGATGTG</u>
44240	
BCR	<u>GAAGTCCCACATCAAGGTGTCCGCAGGGCCAGGCTGTCTCTGGAGG . CTCTAGGGGAGGATCCTTCCTTTCTGTTCCCAGC</u>
BCR ABL	<u>GAAGTCCCACATCAAGGTGTCCGCAGGGCCAGGCTGTCTCCGCCCCAGACATCATCAGTCCTTGCTGTGTATGATTC</u>
ABL	<u>GAGGAGGCCTCAGAGTTCTCATGGGCAAAGATTCCAAAGGCGCCCCAGACATCATCAGTCCTTGCTGTGTATGATTC</u>
ABL BCR	<u>GAGGAGGCCTCAGAGTTCTCATGGGCAAAGATTCCAAAGGCGCCCCACTCTAGGGGAGGATCCTTCCTTTCTGTTCCCAGC</u>
Ital 4	
BCR	<u>TGTCTGCTTTGCTTCAGACCTTCTATCAAATTCATTGTC . . TCTGTCTTAATGGTCTTTGCAGGGTAGGCATGGCCAT</u>
BCR ABL	<u>TGTCTGCTTTGCTTCAGACCTTCTATCAAATTCATTGTCGAATCGAGGAACCTCTCAGGCTTTGTGTTGGGACGCTGA</u>
ABL	<u>CGCTGACACACTGCATTAGTACTTTCTAGACTGTGAGCAGAGGAATCGAGGAACCTCTCAGGCTTTGTGTTGGGACACTGA</u>
ABL BCR	<u>CGCTGACACACTGCATTGGTACTTTCTAGACTGTGAGCAGTTGTCTCTGTCTTAATGGTCTTTGCAGGGTAGGCATGGCCAT</u>

22852	
BCR	<u>AGCTGTGGAGGCAAGAGTGTGGGCAGGGTGACCTGATTGAGCCAAGCTGTTATTCTTCCACCTGAGGATTTTCCAGAAGAA</u>
BCR ABL	<u>AGCTGTGGAGGCAAGAGTGTGGGCAGGGTGACCTGATTGAGCCCGGCTAATTTTTGTATTTTAGTAGAGACGAGGTTTC</u>
ABL	<u>TCAGCCTCCCAAGTACTGGGACTACAGGCACCCACCACCACCCCGGCTAATTTTTGTATTTTAGTAGAGACGAGGTTTC</u>
ABL BCR	<u>TCAGCCTCCCAAGTACTGGGACTACAGGCACCCGCCACCACCCC. . TGTATTCTTCCACCTGAGGATTTTCCAGAAGAAG</u>
11941	
BCR	<u>CTTTGGATAGAAGCATCTTTCCAGACAATTGGGATCCTGTAGAAAATCTGCCCTCTGAGGCTCTTCTATGCCTTCTGTTG</u>
BCR ABL	<u>CTTTGGATAGAAGCATCTTTCCAGACAATTGGGATCCTGTCAAAAGAGTTTACTTTCATTTAAATGAAGCATGTTGTCAA</u>
ABL	<u>TGTGGGGTAAGTAAAGAAAGATGATCAAAGATGATTGACCAAAAGAGTTTACTTTCATTTAAATGAAGCATGTTGTCAA</u>
ABL BCR	<u>TGTGGGGTAAGTAAAGAAAGATGATCAAAGATGATTGA. . GAAAATCTGCCCTCTGAGGCTCTTCTATGCCTTCTGTTG</u>
28835	
BCR	<u>AGGCCCACTGCAGCCCCCGATTCCATCTCAGGTGATAAGGGCTATTTTCTCCCTGGTTCAAGGAAGACACCCCTCCCAG</u>
BCR ABL	<u>AGGCCCACTGCAGCCCCCGATTCCATCTCAGGTGATAAGGGGCTTCTTTCAGTGAGCGTAATGCCTTTGAGATCCATCTA</u>
ABL	<u>CACAGAAATGGAAGCATCTAGTATGTAACCTTTTGAGATTGGCTTCTTTCAGTGAGCGTAATGCCTTTGAGATCCATCTA</u>
ABL BCR	<u>CACAGAAATGGAAGCATCTAGTATGTAACCTTTTGAGATTGCTATTTTCTCCCTGGTTCAAGGAANACACCCCTCCCAG</u>
19635	
BCR	<u>TCCATTTTGTTTTCTCTCTTATTAGATTGGATAATTTCTATTGATCTATCTTTAAGTTCAGTACTCCATTGTCAT</u>
BCR 5'ABL	<u>TCCATTTTGTTTTCTCTCTTATTAGATTGGATAATTCGTGTCACTGCACTCCAGCCTGGGCAACAGAGCAAGACTG</u>
5'ABL	<u>CGCTTCAGCCTGGGTGGTGGAGGTTGCAATGAGCAGAGATCGTGTCACTGCACTCCAGCCTGGGCAACAGAGCAAGACTG</u>
5'ABL BCR	<u>CGCTTCAGCCTGGGTGGTGGAGGTTGCAATGAGCAGAGATCGT. . TGATCTATCTTTAAGTTCAGTACTCCATTGTCAT</u>
28624	
BCR	<u>GGTGACACAGGAAGGTCTGTGGGAAGGGGTGCAGAGCTTCCATGCCCTCTCGGGGCACTGCCACATTTTCAGCAACCTG</u>
BCR ABL	<u>GGTGACACAGGAAGGTCTGTGGGAAGGGGTGCAGAGCTTCCAGACCCAGAAGTCTTACACAGCACTTTTGCCACATAC</u>
ABL	<u>AGAAAGAGGAGAGGGAGTGCCTGCCTGTTGCTTTGAAGAACCTGACCCAGAAGTCTTACACAGCACTTTTGCCACATAC</u>
ABL BCR	<u>AGAAAGAGGAGAGGGAGTGCCTGCCTGTTGCTTTGAAGAACCT. . . CTCGGGGCACTGCCACATTTTCAGCAACCTG</u>
27523	
BCR	<u>AATGTTGCAGGTGGGAGTGCCTGAACAGGTGAGGCTTCCGGTAATTATGACAAGTGATGCTTGTGCTTCTCAGTGCCGTT</u>
BCR ABL	<u>AATGTTGCAGGTGGGAGTGCCTGAACAGGTGAGGCTTCCGGGAGAATCGCTTGAACCTGGGAGGCAGAGATTGCAGTGAG</u>
ABL	<u>GCATATGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGAGGAGAATCGCTTGAACCTGGGAGGCAGAGATTGCAGTGAG</u>
ABL BCR	MLPA normal
24350	
BCR	<u>GCATGGCCATCTGCATGTTCACTGAAGGGCCTGCCTGCATTTCCAGGGAGAGAAAGCCTTCAGGCCACTGGTGCATAAAG</u>
BCR ABL	<u>GCATGGCCATCTGCATGTTCACTGAAGGGCCTGCCTGCATTTCTAGCTGAGATTTTCATTCATTGGAAACATGCCTTATA</u>
ABL	<u>TTTTTGTGTTCCTATTTCCCCCATGAGATTTCCTACTTCTTAGCTGAGATTTTCATTCATTGGAAACATGCCTTATA</u>
ABL BCR	Not found
P747	
BCR	<u>GTCAGGATAATCGCTTGAACCCAGAAGGTAGAGGTTGCAGTAAGCCGAGATCACGCCACTGCACTCCATCCTGGGCGACA</u>
BCR ABL	<u>GTCAGGATA. TCGCTTGAACCCAGAAGGTAGAGGTTGCAGTAGAACTGAGCAGCCCTCACTGGGGTGCACTTCTAATCCT</u>
ABL	<u>CCAACAACACGTACAGCACAGTCTTTCTGTGTTTGAGCTTGGAAGTGAAGCAGCCCTCACTGGG. TGCACTTCTAATCCT</u>
ABL BCR	<u>CCAACAACACGTACAGCACAGTCTTTCTGTGTTTGAGCTTGGAAGTGAAGCAGCCCTCACTGGG. CGAGATCACGCCACTGCACTCCATCCTGGGCGACA</u>

P653	
BCR	<u>CCACCATCTCCGTTTAGTTCCGAAACATTTTCGTACCCCC..AGAAGGAAACCCCTGCCCATTAAGCAGTCACCGACCTCCT</u>
BCR	
11p (3024bp) ABL	<u>CGAAACATTTTCGTACCCCCGGGTGGGGGAGGGGGGAG..TGGCATGATTCATACAAAAATTTATGACCTGAAGTGTTTT</u>
ABL	<u>AACACCTGGAAATTAAACTTAGGCCAACACATCAAAGTC..TTTATGACTTGAAGTGTTTTAAGAATTAATGAGGCAAAGC</u>
ABL BCR	Not found

p210 ALL

P851	
BCR	<u>TTTGGGAGGCTGAGGCAGGTGGATCGCTTGAGCTCAGGAGTTGGAGACCAGCCTGACCAACATGGTGAAACCCCTGTGTCT</u>
BCR ABL	<u>TTTGGGAGGCTGAGGCAGGTGGATCGCTCGAGCTCAGGAGCACAGAGAACTGTGTTTTT.ACATCATACCTTTAGGATTTT</u>
ABL	<u>TCTGTTAGGAATCTGGAGCAAATCCTATTCTTTTGAAAGCACAGAGAACTGTGTTTTTACATTATACCTTTAGGATTTT</u>
ABL BCR	MLPA failed
26231	
BCR	<u>TGCAGGGCCCTTCTCATCGTAGGGGCTTTAGCTGGGGTTTGTGGATCGACTGAGTGAACGAATGTTGTGGGAAGTCCC</u>
BCR ABL	<u>TGCAGGGCCCTTCTCATCGTAGGGGCTTTAGCTGGGGTTTAAAGATTCTTCTGTTTTTTTTCTAGAAGATTTATAAG</u>
ABL	<u>TTACCTAACTCAAGTTTACCTTCACCTAACTCAAGTTTACAAAGATTCTTCTGTTTTTTTTCTAGAAGATTTATAAG</u>
ABL BCR	<u>TTACCT..GTGGATCGACTGAGTGAACGAATGTTGTGGGAAGTCCC</u>
P682	
BCR	<u>GTGGATCGCTTGAGCTCAGGAGTTGGAGACCAGCCTGACCAACATGGTGAAACCCCTGTGTCTACTAAAAATACAAAGATT</u>
BCR ABL	<u>GTGGATCGCTTGAGCTCAGGAGTTGGAGACCAGCCTGACCATGGTGCTGTTGGGATGTCATTTTGCTGTACCAAGGGTAA</u>
ABL	<u>ATGGTGCTGTTGGGATGTCACATTGCCTTGCCTGCAGATCATGGTGCTGTTGGGATGTCATTTTGCTGTACCAAGGGTAA</u>
ABL BCR	MLPA failed
22251	
BCR	<u>AGCAATACAGCGTGACACCCTACGCTGCCCCGTGGTCCCGGGCTTGTCTCTCCTTGCCCTCCCTGTTACCTTTCTTTCTAT</u>
BCR ABL	<u>AGCAATACAGCGTGACACCCTACGCTGCCCCGTGGTCCCGAGGTTCCACTCTTCTTGAGATTTGTGTTACTTTTATAGACA</u>
ABL	<u>TGGAACCTTAGTTACAATGTTGGAACCTTAGTTACAATGTAGGTTCCACTCTTCTTGAGATTTGTGTTACTTTTATAGACA</u>
ABL BCR	MLPA = BCR deletion
26219	
BCR	<u>GCCGAGCCAGGGTCTCCACCCAGGAAGGACTCATCGGGCAGGGTGTGGGAAACAGGGAGGTTGTTTCAGATGACCACGGGA</u>
BCR ABL	<u>GCAGAGCCAGGGTCTCCACCCAGGAAGGACTAATCGGGCATTGGTCAGGCTGGTCTCAAACCTCCGACCTCAGGTGATCTG</u>
ABL	<u>GCTAATTTTGTATTTTAGTAGAGATGGGGTTTCTCCATATTGGTCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCTG</u>
ABL BCR	MLPA = Normal
24375	
BCR	<u>GGGCTAGGCAGTGGGCACCTGTAATCACAACCTGCTTGGGAGGCTGAGGGAAGAGAATCGCTTGAACCCAGGAGCGGAGG</u>
BCR ABL	<u>GGGCTAGGCAGTGGGCACCTGTAATCACAACCTGCTTGGGAGGTTCAAGCAATTCTCTGCCTCAGCCTCCCGAGTAGCTG</u>
ABL	<u>TGGTGGTGCATCACTGCTCACTGCATCCTCCACCTCCAGGTTCAAGCAATTCTCTGCCTCAGCCTCCCGAGTAGCTG</u>
ABL BCR	No band, MLPA not done
19715	
BCR	<u>GCCTCCCTTTCCCGGACAAACAGAAGCTGACCTCTTTGATCTCTTGCGCAGATGATGAGTCTCCGGGGCTCTATGGGTTTC</u>
BCR ABL	<u>GCCTCCCTTTCCCGGACAAACAGAAGCTGACCTCTTTGATCAGAGTTCCCTAGATTACTACAGTTCCCTAGTTAGCTGGCAGA</u>
ABL	<u>TTCCAAGAAGAGACATTATTTTAGAGAGATACTAGGTTGCTAGAGTTCTAGATTACTACAGTTCTAGTTAGCTGGCAGA</u>

ABL BCR	CTGCCTTCTTTTAAACTTTGTTTGAATTTACCTGGCCTCTGGTTTGCCTGTATTGTGAAACCAACTGGATCCTGAGATCCCCAAG
11129	
BCR	<u>TATTGTGAAACCAGCTGGATCCTGAGATCCCCAAGACAGAAATCATGATGAGTATGTTTTTGGCCCATGACACTGGCTTA</u>
BCR ABL	<u>TATTGTGAAACCAACTGGATCCTGAGATCCCCAAGACAGCTTGAACACAAATGTGTTTCTCTCTGAGATCTTTAGTCTGT</u>
ABL	<u>ATACTTGTTTGGTTTATCAGGTGGTTTCTTGGCTGGGTTTGAACACAAATGTGTTTCTCTCTGAGATCTTTAGTCTGT</u>
ABL BCR	<u>TCCCCGTTTCTTCAGCACTGCTTGATATGACTGGTATGGAGGTTGATTTTGAAGCAGAGTTAGCTTGTACCTGCCTCCCT</u>
10258	
BCR	<u>AACTGCTTGGGAGGCTGAGGGAAGAGAATCGCTTGAACCC. . . .AGGAGGCGGAGGTTGCAGTGAGCCGAGCTTGTG</u>
BCR ABL	<u>AACTGCTTGGGAGGCTGAGGGAAGAGAATCGCTTGAACCTTACCCTTAAGGATTTTCTTCTCTGATTGCACTAAATCTAT</u>
ABL	<u>TTTAGAGTGGGTTTATCAGCTTCCATACCCAAACAGAAATACCTTAAGGATTTTCTTCTCTGATTGCACTAAATCTAT</u>
ABL BCR	<u>TTTAGAGTGGGTTTATCAGCTTCCATACCCAAACAGAAATAAACCAGGAGGCGGAGGTTGCAGTGAGCCGAGCTTGTG</u>
11128	
BCR	<u>AAATCATGATGAGTATGTTTTTGGCCCATGACACTGGCTTACCTTGTGCCAGGCAGATGGCAGCCACACAGTGTCCACCG</u>
BCR ABL	<u>AAATCATGATGAGTATGTTTTTGGCCCATGACACTGGCTTGGAAAAGGGATAGTTAATAGTAATCATAATAATGTTTAAT</u>
ABL	<u>AGCTTTTATCTTCCTTACACCACTGCAAGTTTATAGAAAGGAAAGGGATAGTTAATAGTAATCATAATAATGTTTAAT</u>
ABL BCR	MLPA = BCR & ABL DELETION 07.01.26
6875	
BCR	<u>CCTGGTCTTTGTAGCTCTGGATATCCCTGCAGAAAGGGTCCCCACTACCAGGCCTCTCCATCCCCAGTCTCAGGTAGTTT</u>
BCR ABL	<u>CCTGGTCTTTGTAGCTCTGGATATCCCTGCAGAAAGGGTCCGAACCTGGCTATCACTTCTCTGTTCCCACTGCATGCTG</u>
ABL	<u>AAAAGAACACCCTTGGTAGCTCCTACTTGTGCTGCAGAGGTGAACCTTGGCTATCACTTCTCTGTTCCCACTGCATGCTG</u>
ABL BCR	<u>AAAAGAACACCCTTGGT. CCATCCCCAGTCTCAGGTAGTTT</u>
10887	
BCR	<u>AATCATGATGAGTATGTTTTTGGCCCATGACACTGGCTTA. . .CCTTGTGCCAGGCAGATGGCAGCCACACAGTGTCCACC</u>
BCR ABL	<u>AATCATGATGAGTATGTTTTTGGCCCATGACACTGGCTTACGAGGTTATTGATTTGCTGGTTTGAAGAACCTCGTCTACA</u>
ABL	<u>GGATTTCACTCCACAATGTTCTTCAACTGTAAGATCTACTTGAGGTTATTTGATTTGCTGGTTTGAAGAACCTCGTCTACA</u>
ABL BCR	<u>GGATTTCACTCCACAATGTTCTTCAACTGTAAGATCTACTTGACCTTGTGCCAGGCAGATGGCAGCCACACAGTGTCCACC</u>
F22666	
BCR	<u>AGAGCAAGACTCCGCCTCAAAAAAAAAAAAAAAAAAGTTCTAGAAACAGCAAAATGTGGAGACAGAAAGCTTACCAGGGAT</u>
BCR ABL	<u>CAGAGCAAGACTCCGCCTCAAAAAAAAAAAAAAAAAAGTTCTACATCCACCATATTCGCTCCAAATCTTAACCTGATATGA</u>
ABL	<u>AGAAGGGACCTGGAAGGAGGATGGAATGGCTATGGAATTATCCATCCACCATATTCGCTCCAAATCTTAACCTGATATGA</u>
ABL BCR	<u>AGAAGGGACCTGGAAGGAGGATGGAATGGCTATGGATTATC. . CAGCAAAATGTGGAGACAGAAAGCTTACCAGGGAT</u>
F17685	
BCR	<u>AAGGGGGACTTTTtagGTGAGAGCAGTGTCGTGAAAAGACTGTGGTGCTGTTTGCCTCACATTACATTTCTAAAATTC</u>
BCR ABL	<u>AAGGGGGACTTTTcAGGTGAGCGCAGTGTCGTGAAAAGACTCAATTATGTAAGTTTAGGGCCAGGCGGGTGGCTCCTGCC</u>
ABL	<u>GGAAGCCTCCCTCCCCAGCATGCTTTTTCATGTTAAATTATCAATTATGTAAGTTTAGGGCCAGGCGGGTGGCTCCTGCC</u>
ABL BCR	MLPA = BCR & ABL DELETED 07.02.07
F17730	
BCR	<u>CATCACCCCGACCCCTCTGCTGTCCTTGGAACCTTATTACACTTCGAGTCACTGGTTTGCCTGTATTGTGAAACCAGCTG</u>
BCR ABL	<u>CGTCACCCCGACCCCTCTGCTGTCCTTGGAACCTTATTACTAACAATCAGAATAAGAATAAAAATGAATTATTTTCTTT</u>
ABL	<u>GAGATTGCAATGTCAGAAAACAGGATGCTTTTAAAAACAATAACAATCAGAATAAGAATAAAAATGAATTATTTTCTTT</u>
ABL BCR	MLPA failed

F18544	
BCR	<u>GAAGAGCTATGCTTGTTAGGGCCTCTGTCTCCTCCAGGAGTGACAAAGGTGGGTAGGAGCAGTTTCTCCCTGAGTGGCTG</u>
BCR ABL	<u>GAAGAGCTATGCCCTCCAGCCCTAGGTGACAGAGCAAGGCCCTGCCTCAAATAAATAAAATAAATTGGCCGGTTGAGCATCTGTA</u>
ABL	<u>GACTTATGACGTCTCAACTTACATTTTCAATTTTATGATGGGTTTTTCAGGATGTAACCCAATTTTAAGTTGAGCATCTGTA</u>
ABL BCR	MLPA failed
F17237	
BCR	<u>GTACACCTCTCTGTCCCCACCAGTGCAGGGCCCTTCTCATCGTAGGGGCTTTAGCTGGGGTTTGTGGATCGACTGAGTGA</u>
BCR ABL	<u>GTACACCTCTCTGTCCCCACCAGTGCAGGGCCCTTCTCATCTTGTAAACTTTTTGGGAGACTAGATTATTTCTCCTCCCC</u>
ABL	<u>CAGTATACTGCTTCTCTAGTTGTTAAGGAAGAAAAATTTCTCTTGTAAACTTTTTGGGAGACTAGATTATTTCTCCTCCCC</u>
ABL BCR	<u>CAGTATACTGCTTCTCTAGTTGTTAAGGAAGAAAAATTTCTCGTAGGGCTTTAGCTGGGGTTTGTGGATCGACTGAGTGAA</u>
F16604	
BCR	<u>CCTGCCAGCCGGCACTTTTGGTCAAGCTGTTTTGCATTCACTGTTGCACATATGCTCAGTCACACACACAGCATAACGCTA</u>
BCR ABL	<u>CCTGCCAGCCGGCACTTTTGGTCAAGCTGTTTTGCATTCACTGTTGCACATATGCTCAGTCACACACACAGCATAACGCTA</u>
ABL	<u>ATTTTTTGCGTTTTTAATAGAGATAGGGTTTACCATAATTGGTTAGGCTGGTCTCAAACCTCGACCTCGTGATCCGCCC</u>
ABL BCR	MLPA = ABL & BCR Exon 16 DELETION
F15010	
BCR	<u>TGGTTTGCCTGTATTGTGAAACCAGCTGGATCCTGAGATCCCCAAGACAGAAATCATGATGAGTATGTTTTTGGCCCATG</u>
BCR ABL	<u>TGGTTTGCCTGTATTGTGAAACCAGCTGGATCCTGAGATCCCCACTGACCTGTCTTTAGACAGGCAGGACAGAGGATCT</u>
ABL	<u>AAGTAGGGGATGAAGGGGAAGCCCTTCAGTTAAACAAAGTTACTGACCTGTCTTTAGACAGGCAGGACAGAGGATCT</u>
ABL BCR	MLPA failed
F22757	
BCR	<u>CTTGGGAGGCTGAGGGAAGAGAATCGCTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGAGCTTGTGCCACTGCATTCCA</u>
BCR ABL	<u>CTTGGGAGGCTGAGGGAAGAGAATCGCTTGAACCCAGGAGTTGTGTCTTTTGTGTTGGCTTTTGTGTTTTGTTTTGCCCC</u>
ABL	<u>AGAGGTTGACAAGTTTTTTGGTATTGGTTTTAGTTTTGGGTTGTGTCTTTTGTGTTGGCTTTTGTGTTTTGTTTTGCCCC</u>
ABL BCR	MLPA failed
F14472	
BCR	<u>CGGATGGTTGATTTTGAAGCAGAGTTAGCTTGTACCTGCCTCCCTTTCCCGGGACAACAGAAGCTGACCTCTTTGATCT</u>
BCR ABL	<u>CGGATGGTTGATTTTGAAGCAGAGTTAGCTTGTACCTGCCTCCCGCCACCTCGGCCTCCCAAAGCCCTGGGATTACAGGCG</u>
ABL	<u>CATGTTAGCCAGGATGGTCTCCATCTCCTGACCTCGTGATCCGCCACCTCGGCCTCCCAAAGCCCTGGGATTACAGGCG</u>
ABL BCR	<u>CATGTTAGCCAGGATGGTCTCCATCTCCTGACCTCGTGATTTCCCTTTCCCGGGACAACAGAAGCTGACCTCTTTGATCT</u>
F11329	
BCR	<u>GTGTCCACACACACCCACCCACATCCACATCACCCCGACCCCTCTGCTGTCTTGGAACCTTATTACACTTCGAGTC</u>
BCR ABL	<u>GTGTCCACACACACCCACCCACATCCACATCACCCCGACAAGTGCATATGGAACGTGTTCCAAGACAAATCATTTTCT</u>
ABL	<u>AGAACATTGACCCCTAAACAGCAGAATACAGTTCTTTTCAAGTGCATATGGAACGTGTTCCAAGACAAATCATTTTCT</u>
ABL BCR	MLPA = Normal
F11751	
BCR	<u>TTAAGCAG . AGTTCAAGTAAG . TACTGGTTTGGGGAGGAGGGTTGCAGCGGCCGAGCCAGGGTCTCCACCCAGGAAGGACT</u>
BCR ABL	<u>TTAAGCAGTAGTTCAAGTAAGATACTGGTTTGGGGAGGAGCTTCTGGAGGCCACCTGCATTGCTTGGTTTGTGACCCCTT</u>
ABL	<u>CCAGAGACTCCAGGAGAGAAATACTTCCTGGCCTCTCTGGCTTCTGGAGGCCACCTGCATTGCTTGGTTTGTGACCCCTT</u>
ABL BCR	<u>CCAGAGACTCCAGGAGAGAAATACTTCCTGGCCTCT . .GGGTTGCAGCGGCCGAGCCAGGGTCTCCACCCAGGAAGGAC</u>

F24559	
BCR	<u>ACATCACCCCGACCCCTCTGCTGTCTTGGAACTTATTACACTTCGAGTCACTGGTTGCGCTGTATTGTGAACCAGC</u>
BCR ABL	<u>ACGTCACCCCGACCCCTCTGCTGTCTTGGAACTTATTACGCTTGGCCTCTATATAAGCTTTAATGAATAATAAGA</u>
ABL	<u>TTGGCTTCCCAAAGTCTTGGGATTGCAGGCGTGAGCCACTGCGCTTGGCCTCTATATAAGCTTTAATGAATAATAAGA</u>
ABL BCR	MLPA failed
F20099	
BCR	<u>CTAAATGCAAACCCACCTGCAACTTACGCCCACAGCCAGCCACTCTTCTCCAGGCTCGCCTCCCTCCCTTCCC</u>
BCR ABL	<u>CTAAATGCAAACCCACCTGCAACTTACGCCCACAGCAATAACCCATCTTAGCTAGACTCCAGGTTTCGTACATAT</u>
ABL	<u>GGAGGCCATTACTAGATTACAGATGATGGGCGTGCTTAAATAACCCATCTTAGCTAGACTCCAGGTTTCGTACATAT</u>
ABL BCR	MLPA = ABL1 & EXSOC2 DELETED

p210 CML

NP241	
BCR	<u>TGCACGGCTTCTGTTCTTAGTCACAAGGCTGCAGCAGACGCTCCTCAGATGCTCTGTGCGCTTGGATCTGGCCCCACTCCC</u>
BCR ABL	<u>TGCACGGCTTCTGTTCTTAGTCACAAGGCTGCAGCAGACGCTCCACTAAAAATACAAAAAATTCTCCGGGCGTGGTGCGG</u>
ABL	<u>GAGACCGAGACCATCCTTGCTAACACGGTGAAACCCGCTCTCCACTAAAAATACAAAAAATTCTCCGGGCGTGGTGCGG</u>
ABL BCR	MLPA = BCR & ABL DELETED
E656	
BCR	<u>GTAAGTACTGGTTTGGGGAGGAGGGTTGCAGCGGCCGAGC</u>
BCR ABL	<u>GTAAGTACTGGTTTGGGGAGCAGGTACAGGGGTGAGCCTC</u>
ABL	<u>CCTCAGCCTCCCAAAATGCTGGGATTACAGGGGTGAGCCTC</u>
ABL BCR	Not found, MLPA not done
1538	
BCR	<u>TCCACCCAGGAAGGACTCATCGGGCAGGGTGTGGGGAAACAGGGAGGTTGTTTCAGATGACCACGGGACACCTTTGACCCCT</u>
BCR ABL	<u>TCCACCCAGGAAGGACTCATCGGGCAGGGTGTGGGGAAACGGCATCGAAAGCTGTTGACCTTCATCAACAATGAGCACA</u>
ABL	<u>TAATCAACTCCCCTTTTAATTTTTTCAAAAGTCATTATTAGGCATCGAAAGCTGTTGACCTTCATCAACAATGAGCACA</u>
ABL BCR	MLPA failed
1716	
BCR	<u>TGGTCCCGGGCTTGTCTCTCCTTGCCTCCCTGTTACCTTTCTTTCTATCTCTTCCCTTGCCCCGTGCACTCAACCTTGCACT</u>
BCR ABL	<u>TGGTCCCGGGCTTGTCTCTCCTTGCCTCCCTGTTACCTTTGATTTTTTAGTAGATTATAAGTTTTAATTAACTTTTAG</u>
ABL	<u>AAATCTTCACCTAACTCAAGTTTACAAAGATTCTTCTGTTTTTTTCTAGAAGATTATAAGTTTTAATTAACTTTTAG</u>
ABL BCR	<u>AAGATTATAAGTTTTAATTAACTTTTAGGCTTATAAATCTTGGGAGGCTGAGGGAAGAGAATCGCTTGANCCAGGAG</u>
1578	
BCR	<u>TGTTTTTGCCCATGACACTGGCTTACCTTGTGCCAGGCAGATGGCAGCCACACAGTGTCCACCGGATGGTTGATTTTGA</u>
BCR ABL	<u>TGTTTTTGCCCATGACACTGGCTTACCTTGTGCCAGTGCCACACCACCATGCCCGGCTACTTTTTTTTGTATTTTAG</u>
ABL	<u>TCTATTGCCTCAGCTCCCGTAGCTGGGACTACGGGCACACACCACCATGCCCGGCTACTTTTTTTTGTATTTTAG</u>
ABL BCR	MLPA = ABL DELETED 07.01.26
1349	
BCR	<u>AATTCCACAGAGCGGGCAGGGGCATCGCATGAGGTGCTGGTGTTCACGCCAGACCACAATTAGGTGTTAATTTTTAAAA</u>
BCR ABL	<u>AATTCCACAGAGCGGGCAGGGGCATCGCATGAGGTGCTGGTGTGGGTTCTATTGCTTCCACATCAGGAACCTATAATTACT</u>
ABL	<u>ATTTAAAGTAGGTGCCCTAAGACATTTTTTCAGGGTGCAATTTGGGTTCTATTGCTTCCACATCAGGAACCTATAATTACT</u>

ABL BCR	MLPA = BCR & ABL DELETED 07.01.26
1103	Break in BCR exon 4
BCR	<u>cccacagATCTGTACTGCACCCCTGGAGGTGGATTCCCTTTGGGTATTTTGTGAATAAAGCAAAGACGCGCGTCTACAGGGA</u>
BCR ABL	<u>CCCACAGATCTGTACTGCACCCCTGGAGGTGGATTCCCTTTGGATAAGTATATGCTTAACTTAGGAGACCCCTGCCACAGTAT</u>
ABL	<u>TCTTCGGTAAATAATTAGGAGTAGAATTGCTAGGTCGTAGGATAAGTATATGCTTAACTTAGGAGACCCCTGCCACAGTAT</u>
ABL BCR	<u>CACACCATATGTATTACACCAGTTGCTCCATTGGTCTT. AAGACGCGCGTCTACAGGGA</u>
1810	
BCR	<u>CCATCAATAAGGAAGGTGGGCCCCCCC.GTTTCCTGTACAGGGCACCTGCAGGGAGGGCAGGCAGCTAGCCTGAAGGCT</u>
BCR ABL	<u>CCATCAACAAGGAAGGTGGGCCCCCCCCGTCTCCGTGTACAGTGAAACCTGAATAATCTGGCTGGGCGTGGTGGCTCAC</u>
ABL	<u>CTCCTGCTTTCAGTCTTCCTCCTGCTAGTTATTTTTCATATGAAACCTGAATAATCTGGCTGGGCGTGGTGGCTCAC</u>
ABL BCR	No band by PCR MLPA not done
1569	
BCR	<u>TCCAGCCTGGGCGACAGAGCAAGACTCCGCCTCAAAAAAAAAAAAAAAAAAGTTCCCTAGAAACAGCAAAATGTGGAGACAG</u>
BCR ABL	<u>TCCAGCCTGGGCGACAGAGCAAGACTCCACCTCAAAAAAAAAAAAAAAAA. . .TTTACTATCATGTAAAGGGGAGTT</u>
ABL	<u>TCCAGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAATTTACTATCATGTAAAGGGGAGTT</u>
ABL BCR	<u>CTGAGGCAGGAGAATGGCGTGAACCTGGGAGGCGGAGCTTGGGAGAGAGGACTAACTGCAGATGAACCCAAGGGGGACTT</u>
1460	
BCR	<u>TCCAGCCTGGGCGACAGAGCAAGACTCCGCCTCAAAAAAAAAAAAAAAAAAGTTCCCTAGAAACAGCAAAATGTGGAGACAG</u>
BCR ABL	<u>TCCAGCCTGGGCGACAGAGCAAGACTCCACCTCAAAAAAAAAAAATCAAACCAACAATAGCAAAACTCATGATACCTCCTG</u>
ABL	<u>CCTGCTGTTTGATTTTCAGTTTCTGTAGAAAGGAAAAAAAAAAATCAAACCAACAATAGCAAAACTCATGATACCTCCTG</u>
ABL BCR	<u>GAGCTCAGGAGTTGGAGACCAGCCTGACCAACATGGTGAAACCTGTGTNTANTAAAAATACAAAGATTAGCCAGGNTAG</u>
230	
BCR	<u>AGGGCTGTGAGCAGTGCACCTTCACCCACAGCAGAGCAGATTGGCTGCTCTGTCTGAGCTGGATGATACTACTTTTTT</u>
BCR ABL	<u>AGGGCTGTGAGCAGTGCACCTTCACCCACAGCAGAGCAGATTATTTTTCTTTCCTGTTGTAGTCTTTAGTTTATA</u>
ABL	<u>TTCTTTCTACCTACTCTCCTTCTCTAATTACATCTGTTTTATTTTATTTTCTTTCCTGTTGTAGTCTTTAGTTTATA</u>
ABL BCR	MLPA = NOT DELETED
248	
BCR	<u>GCACTTTTGATGGGACTAGTGGACTTTGGTTCAGAAGGA. .AGAGCTATGCTTGTAGGGCCTCTTGTCTCCTCCCAGG</u>
BCR ABL	<u>GCACTTTTGATGGGACTAGTGGACTTTGGTTCAGAAGGAAGACCCACCTACATGGGAGGCTGAGGCAGAGGGATCACGTG</u>
ABL	<u>AAAATACAAAATTAGTCGGCATGGTGGCACACACCTATAGTCCACCTACATGGGAGGCTGAGGCAGAGGGATCACGTG</u>
ABL BCR	<u>AAAATACAAAATTAGTCGGCATGGTGGCACACACCTATAGTAGAGCTATGCTTGTAGGGCCTCTTGTCTCCTCCCAGG</u>
266	
BCR	<u>ACACCCACCCACATCCCACATCACCCGACCCCTCTGTCTGCTTGGAAACCTTATTACACTTCGAGTCACTGGTTTGC</u>
BCR ABL	<u>ACACCCACCCACATCCCACATCACCCGACCCCTCTGTGATCTACCCACCTTGGCCTCCCAAGGTGCTGGGATTACA</u>
ABL	<u>ACCGTGTGGCCAGGCTGGTCTCAAACCTCTGACCTCAGTGATCTACCCACCTTGGCCTCCCAAGGTGCTGGGATTACA</u>
ABL BCR	<u>GCTGAAGGACAGTGCTCCAGGGTCCTAGAGCACAGAACCTCTTGGTTCAGAAGGAAGAGCTATGCTTGTAGGGCCTCTT</u>
316	
BCR	<u>GTTTTGCATTCACTGTTGCACATATGCTCAGTCACACACACAGCATACGCTATGCACATGTGTCCACACACCCCCACCA</u>
BCR ABL	<u>GTTTTGCATTCACTGTTGCACATATGCTCAGTCACACACACACAGAGATGAAAAATACACTGAAGACTGTGAGAATAAACA</u>
ABL	<u>GGGCGAGAGAGCAAGATTCGCTCTCCACAAAAAGAAAGATACAGAGATGAAAAATACACTGAAGACTGTGAGAATAAACA</u>
ABL BCR	<u>GGGCGAGAGAGCAAGATTCGCTC.CATGTGTCCACACACCCCCACCA</u>

403	
BCR	<u>CTTGGATCTGGCCCCACTCCCGTCTCCAGCCCTCCTCTCCTC . CAGCTACCTGCCAGCCGGCACTTTTGGTCAAGCTGT</u>
BCR ABL	<u>CTTGGATCTGGCCCCACTCCCGTCTCCAGCCCTCCTCTCCATAGAGATGGGTCTCACTATGTCGCCCAGGCTGGTCTC</u>
ABL	<u>TAGTGTCAATTTTAAAACTATCTCTTTTAAATTAAAAAAATAGAGATGGGTCTCACTATGTCGCCCAGGCTGGTCTC</u>
ABL BCR	<u>TAGTGTCAATTTTAAAACTATCTCTTTTAAATTAAAAAAATACAGCTACCTGCCAGCCGGCACTTTTGGTCAGCTAC</u>
440	
BCR	<u>GTCACACACACAGCATACGCTATGCACATGTGTCCACACACACCCACCCACATCCCACATCACCCGACCCCTCTGCT</u>
BCR ABL	<u>GTCACACACACAGCATACGCTATGCACATGTGTCCACACACAGCCTGGGCGACAGAGCGGACTCCATCTCAAAAAAAG</u>
ABL	<u>GGCAGAGCTTGCAGTGAGCCAAGATCGCGCACTGCACTCCAGCCTGGGCGACAGAGCGGACTCCATCTCAAAAAAAG</u>
ABL BCR	MLPA NOT DELETED
446	
BCR	<u>TGGTTGATTTTGAAGCAGAGTTAGCTTGTACCTGCCTCCCTTTCCCGGACAAACAGAAGCTGACCTCTTTGATCTCTTG</u>
BCR ABL	<u>TGGTTGATTTTGAAGCAGAGTTAGCTTGTACCTGCCTCCCTGAGCCACTGTGCCCGGCCGGTTGAACGTTTTTAACATG</u>
ABL	<u>TCTGCCACCTCGGCCTCCCAAAGTGTGGGATTACAGGCGTGAGCCACTGTGCCCGGCCGGTTGAACGTTTTTAACATG</u>
ABL BCR	MLPA NOT DELETED
483	
BCR	<u>CTTTGGTTCAGAAGGAAGAGCTATGCTTGTAGGGCCTCTTGTCTCTCCAGGAGTGACAAGGTGGGTAGGAGCAGT</u>
BCR ABL	<u>CTTTGGTTCAGAAGGAAGAGCTATGCTTGTAGGGCCTCTTACCATGTTGATCAGGCTGGTCTCAAACCTCTGACCTCA</u>
ABL	<u>CATGCCTGGCAAATTTTGTATTTTAGTAGAGACGAGGTTTACCATGTTGATCAGGCTGGTCTCAAACCTCTGACCTCA</u>
ABL BCR	<u>CATGCCTGGCAAATTTTGTATTTTAGTAGAG. TGGACAAGGTGGGTAGGAGCAGT</u>
495	
BCR	<u>TAGCCTGAAGGCTGATCCCCCTTCTGTGTAGCACTTTTGATGGGACTAGTGGACTTTGGTTCAGAAGGAAGAGCTATGC</u>
BCR ABL	<u>TAGCCTGAAGGCTGATCCCCCTTCTGTGTAGCACTTTTGATTTTTTCTGTTAATGATAAAGGCTAAATGATTCCTGTC</u>
ABL	<u>AGAATTCTTATTGATGAAAGAAAGCTATTTTCAGGAACTTTTTTTTCTGTTAATGATAAAGGCTAAATGATTCCTGTC</u>
ABL BCR	MLPA NOT DELETED
524	
BCR	<u>ATTCTCCATCAGTGAGGCTTCTTAGTCATCTCTGGCTGCCTGGCCAGGCCCTGGCTGTGGCCTCTCCCTGGTCTTTGTA</u>
BCR ABL	<u>ATTCTCCATCAGTGAGGCTTCTTAGTCATCTCTGGCTGCCTGTCCAAAAACAGAATTATATAAATCTCTGGATTATGAG</u>
ABL	<u>TAAATGTAAGGAAAAATTCAGAGTTTAGTGAAGTGAGTTTCCAAAAACAGAATTATATAAATCTCTGGATTATGAG</u>
ABL BCR	<u>TAAATGTAAGGAAAAATTCAGAGTTTAGTGAAGTGAG. . . CCAGGCCCTGGCTGTGGCCTCTCCCTGGTCTTTGTA</u>
545	
BCR	<u>TCCTCAGATGCTCTGTGCCTTGGATCTGGCCCCACTCCCGTCTCCAGCCCTCCTCTCCTCCAGCTACCTGCCAGCCGG</u>
BCR ABL	<u>TCCTCAGATGCTCTGTGCCTTGGATCTGGCCCCACTCCCGTCTCCCAAATTCAGGATTTAAAAAGTGAATTATTATAC</u>
ABL	<u>TTTGTTTTTAGGCAGGCAAAATATGGTAATGTGTGTATGATGGTTCCCAAATTCAGGATTTAAAAAGTGAATTATTATAC</u>
ABL BCR	MLPA NOT DELETED
282	
BCR	<u>CCTCTAAGTGGGGGTCTCCCCAGCTACTGGAGCTGTCAGAACAGTGAAGGCTGGTAACACATGAGTTGCACTGTGTAAAA</u>
BCR ABL	<u>CCTCTAAGTGGGGGTCTCCCCAGCTACTGGAGCTGTCAGACTGGGCAACATGGAGAAACCTCATCTGTTAAAAAACAAA</u>
ABL	<u>GAGGTGGGCGGATTGCCTGAGCTCAGGAGTTCGTGACCAGACTGGGCAACATGGAGAAACCTCATCTGTTAAAAAACAAA</u>
ABL BCR	MLPA DELETED

286	
BCR	<u>GACACCCTACGCTGCCCCGTGGTCCCGGGCTTGCTCTCCTTGCCCTCCCTGTTACCTTTCTTTCTATCTCTTCTCCTTGCCCC</u>
BCR 5' ABL	<u>GACACCCTACGCTGCCCCGTGGTCCCGGGCTTGCTCTCCTTGCTGGCAATGCATTCCCTTTTCGTGAGTCGAGGGCAAAC</u>
ABL	<u>GATTGGTTCCACAGGCCTTTTAAAAAGCGGACTTAAAAGTTGCTGGCAATGCATTCCCTTTTCGTGAGTCGAGGGCAAAC</u>
ABL BCR	<u>GTTTCTTAGGGTCCCTGGTTCTATGATTCTGCTCTAACCGACACCTTCATAACATAATCTTTCTCCTGGGCCCCGTGCTCT</u>
287	
BCR	<u>GCAGAGTTCAAGTAAGTACTGGTTTGGGGAGGAGGGTTGCAGCGGCCGAGCCAGGGTCTCCACCAGGAAGGACTCATCG</u>
BCR ABL	<u>GCAGAGTTCAAGTAAGTCTGGTTTGGGGAGGAGGGTTGCACATTAATTACCACACCTAGTATTTTTTTTTTATTTTTAT</u>
ABL	<u>AACTTCTGGACTCATGCAGTCCTCCTGAGTAGTTGGGACTACATTAATTACCACACCTAGTATTTTTTTTTTATTTTTAT</u>
ABL BCR	Not found
291	
BCR	<u>GGTTTGGCAAGGACTTTGACAGACATCCCCAGGGGTGCCCGGAGTGTGGGGTCCAAGCCAGGAGGGCTGTGACAGTGTC</u>
BCR ABL	<u>GGTTTGGCAAGGACTTTGACAGACATCCCCAGGGGTGCCCGAGCATTTTGTTGATCTGCAGGTTAGACTGGCAGTTGAAC</u>
ABL	<u>TGAGATGATTATAATAGCTGCTTAAATCTTTTGCTCTATTCCAGCATTTTGTGATCTGCAGGTTAGACTGGCAGTTGAAC</u>
ABL BCR	MLPA NOT DELETED
	—
300	
BCR	<u>TTTTTATTTTTTCTGATTCTGCAAATAACACCTGCTCTTACAGACCATGTGGGTGATGTGGAAAAGACCTGTGACCTTCT</u>
BCR ABL	<u>TTTTTATTTTTTCTGATTCTGCAAATAACACCTGCTCTTACATGTGAAGTTGTATAATGTCAATTGAAGGTAGACTGTGA</u>
ABL	<u>AGTGGGAGAACAAATATACTCTTGTAAGCTTTTACACTACATGTGAAGTTGTATAATGTCAATTGAAGGTAGACTGTGA</u>
ABL BCR	MLPA = BCR DELETED
317	
BCR	<u>TGGACTTTGGTTTCTGAGGAAGAGCTATGCTTGTTAGGGCCTCTTGCTCTCTCCAGGAGTGGACAAGGTGGTTAGGAG</u>
BCR ABL	<u>TGGACTTTGGTTTCTGAGGAAGAGCTATGCTTGTTAGGGCCCGTTTCTCTGAAATCTTACCCACACTGGCTGTGAGCGTC</u>
ABL	<u>CTCAATTTATACACCTACTACTAGCATTGTATGAGAGTGCCCGTTTCTCTGAAATCTTACCCACACTGGCTGTGAGCGTC</u>
ABL BCR	MLPA = NOT DELETED
375	
BCR	<u>TGCAGAAAGGGTCCCCACTACCAGGCCTCTCCATCCCCAGTCTCAGGTAGTTTTTCTAAAAATGCAAACCCACCTGCAA</u>
BCR ABL	<u>TGCAGAAAGGGTCCCCACTACCAGGCCTCTCCATCCCCAGTTTTTCCCTCTCTCTTTCTCTCCCTTCCCTCTCTGATT</u>
ABL	<u>AAACACTGCATCAAGAGACTGTGCCCCATGTCTTTTCACTTTTTCCTCTCTCTTTCTCTCCCTTCCCTCTCTGATT</u>
ABL BCR	<u>AAACACTGCATCAAGAGACTGTGCCCCATAGTTTTT.CTAAAAATGCAAACCCACCTGCAA</u>
386	
BCR	<u>TTCTCGAGGCCGGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCAGGTGGATCGCTTGAGCTCA</u>
BCR ABL	<u>TTCTCGAGGCCGGGCGCAGTGGCTCATGCCTGTAATCCCAGTCACACACACCTTTTCAGTTAGTTTTGCTGAGGGTTTAC</u>
ABL	<u>CCTCACATTGGTAATTGTGTTTTTCTCTCTGTTTCTCTCTCACACACACCTTTTCAGTTAGTTTTGCTGAGGGTTTAC</u>
ABL BCR	<u>ATTTCTAGCTTCTTAAAGTAGAAATTTAGATCGATTCCAAACACCCACAGCAGAGCAGATTTGGCTGCTCTGTGAGCT</u>
436	
BCR	<u>CTATTATTTCATGGACCCCAAACCTTGTTCTCTTATGTCCTGTCCCTTTGAGGGGCACCACCATCCACCCGCATGGCCAAG</u>
BCR ABL	<u>CTATTATTTCATGGACCCCAAACCTTGTTCTCTTATGTCCTGTCCCTTTGAGGGGTGAGATTACGAGGTCAGGAGATCGAGACCATCCT</u>
ABL	<u>TGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGTGGATTACGAGGTCAGGAGATCGAGACCATCCT</u>
ABL BCR	<u>TGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCC. . . TTTGAGGGGCACCACCATCCACCCGCATGGCCAAG</u>

389	
BCR	<u>AATTGCAGGGGTTTGGCAAGGACTTTGACAGACATCCCCAGGGGTGCCCCGGGAGTGTGGGGTCCAAGCCAGGAGGGCTGT</u>
BCR ABL	<u>AATTGCAGGGGTTTGGCAAGGACTTTGACAGACATCCCCAGGACCACGAGAGGCTCAGAGATATAGTGAGAAGGTCCAAC</u>
ABL	GAAATAACACAGACTGAAGCTAATAGAAACAAAGGGATGTAAACCACGAGAGGCTCAGAGATATAGTGAGAAGGTCCAAC
ABL BCR	GAAATAACACAGACTGAAGCTAATAG. <u>.GAGTGTGGGGTCCCAGCCAGGAGGGCTGT</u>
470	
BCR	<u>GGCTGCCCTCTGCTGTGGCATCACTGTGTAACAATGGCGTGTACACCTCTCTGTCCCCACCAGTGCAGGGCCCTTCTCAT</u>
BCR ABL	<u>GGCTGCCCTCTGCTGTGGCATCACTGTGTAACAATGGCGTGTCTGTTTCCAGGCTTGCAGAGAACTGCTGCTCAGGAG</u>
ABL	CAGATGGAGATGAAGTGGGTGCAGGTCATCGCCACGAC. .CTGTTTCCAGGCTTGCAGAGAACTGCTGCTCAGGAG
ABL BCR	GCAGATGGAGATGAAGTGGGTGCAGGTCATCGCCACGACCT <u>TACACCTCTCTGTCCCCACCAGTGCAGGGCCCTTCTCAT</u>

Appendix IV

Reference sequences

NM_004327 BCR from ATG

NM_007313 ABL from ATG

Locations of recombination sequence motifs in 50 bp either side of BCR and ABLforward breakpoint in p210 ALL

Patient ID	BCR bpt	Fuzznuc 7-mer	Fuzznuc 13-mer	ABL bpt	Fuzznuc 7-mer	Fuzznuc 13-mer
22251	110640	72 to 78	None	80502	None	None
19715	109368	16 to 22	None	60782	None	None
6875	110950	11 to 17	None	39021	None	None
F18544	108792	None	97 to 85 (ccacctgtccac)	99563	None	None
F20099	111031	81 to 87 & 85 to 91	85 to 97 (cctccctccccc)	28149	None	None

Locations of recombination sequence motifs in 50 bp either side of BCR and ABLforward breakpoint in p210 CML

Patient ID	BCR bpt	Fuzznuc 7-mer	Fuzznuc 13-mer	ABL bpt	Fuzznuc 7-mer	Fuzznuc 13-mer
1716	110672	40 to 46	None	104268	None	None
1810	108688	70-64	None	48293	None	None
1338	109540	62-56	None	47245	None	None
266	109131	None	None	55259	None	61-73 (ccacctggcctc)
446	109294	90-96	None	22454	None	None
483	108807	None	82-70 (ccacctgtccac)	107887	None	66-78 (ccacctggcctc)
524	110884	77-83	None	114565	None	None
286	110614	1-7, 98-104	None	-3162 (upstream ABL)		
287	109483	None	None	76292	None	None
317	108803	None	86-74 (ccacctgtccac)	135075	None	None

From Hammersmith						
S9	108995	None	None	-6872 (US ABL)		
S14	111042	70-76, 74-80	74-86 (cctcccttcccc)	122256	None	None
S15	109345	39-45	None	85519	None	None
S19	109374	10-16	None	106169	None	None

**Locations of recombination sequence motifs in 50 bp either side of BCR and ABLforward
breakpoint in p190 ALL**

There were no 7-mers or 13-mers in any of the p190 breakpoints.

Appendix V

FIP1L1-PDGFR α sample calculation

E number	Copies FP / μ l in samples	Am gDNA in sample ng/ μ l	Copies FP / ng DNA	Number haploid genomes	% haploid genomes
	FP	gDNA	FP/gDNA	(FP/gDNA)x330	x100
E243	5060	61.93	81.7	0.248	24.8
E43	341	40.295	8.5	0.026	2.6

.

References

1. Domen J, Weissman IL. Self-renewal, differentiation or death: regulation and manipulation of hematopoietic stem cell fate. *Mol.Med Today* 1999; **5**(5): 201-208.
2. Weissman IL. Stem cells: units of development, units of regeneration, and units in evolution. *Cell* 2000; **100**(1): 157-168.
3. Metcalf D. The molecular control of cell division, differentiation commitment and maturation in haemopoietic cells. *Nature* 1989; **339**(6219): 27-30.
4. Bazan JF. Structural design and molecular evolution of a cytokine receptor superfamily. *Proc Natl Acad Sci U S A* 1990; **87**(18): 6934-8.
5. Robb L. Cytokine receptors and hematopoietic differentiation. *Oncogene* 2007; **26**(47): 6715-23.
6. Darnell JE, Jr., Kerr IM, Stark GR. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* 1994; **264**(5164): 1415-21.
7. Ihle JN. Cytokine receptor signalling. *Nature* 1995; **377**(6550): 591-4.
8. Endo TA, Masuhara M, Yokouchi M, Suzuki R, Sakamoto H, Mitsui K *et al.* A new protein containing an SH2 domain that inhibits JAK kinases. *Nature* 1997; **387**(6636): 921-4.
9. Klingmuller U, Lorenz U, Cantley LC, Neel BG, Lodish HF. Specific recruitment of SH-PTP1 to the erythropoietin receptor causes inactivation of JAK2 and termination of proliferative signals. *Cell* 1995; **80**(5): 729-38.
10. Starr R, Willson TA, Viney EM, Murray LJ, Rayner JR, Jenkins BJ *et al.* A family of cytokine-inducible inhibitors of signalling. *Nature* 1997; **387**(6636): 917-21.

11. Wormald S, Hilton DJ. Inhibitors of cytokine signal transduction. *J Biol Chem* 2004; **279**(2): 821-4.
12. Valentino L, Pierre J. JAK/STAT signal transduction: regulators and implication in hematological malignancies. *Biochem Pharmacol* 2006; **71**(6): 713-21.
13. Liu B, Liao J, Rao X, Kushner SA, Chung CD, Chang DD *et al.* Inhibition of Stat1-mediated gene activation by PIAS1. *Proc Natl Acad Sci U S A* 1998; **95**(18): 10626-31.
14. Sallmyr A, Fan J, Rassool FV. Genomic instability in myeloid malignancies: Increased reactive oxygen species (ROS), DNA double strand breaks (DSBs) and error-prone repair. *Cancer Lett* 2008.
15. Stoklosa T, Poplawski T, Koptyra M, Nieborowska-Skorska M, Basak G, Slupianek A *et al.* BCR/ABL inhibits mismatch repair to protect from apoptosis and induce point mutations. *Cancer Res* 2008; **68**(8): 2576-80.
16. de The H, Vivanco-Ruiz MM, Tiollais P, Stunnenberg H, Dejean A. Identification of a retinoic acid responsive element in the retinoic acid receptor beta gene. *Nature* 1990; **343**(6254): 177-80.
17. Gao J, Erickson P, Gardiner K, Le Beau MM, Diaz MO, Patterson D *et al.* Isolation of a yeast artificial chromosome spanning the 8;21 translocation breakpoint t(8;21)(q22;q22.3) in acute myelogenous leukemia. *Proc Natl Acad Sci U S A* 1991; **88**(11): 4882-6.
18. Melo JV. The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype. *Blood* 1996; **88**(7): 2375-2384.
19. Nowell PC, Hungerford DA. Chromosome studies on normal and leukemic human leukocytes. *J Natl. Cancer Inst.* 1960; **25**: 85-109.

20. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* 1973; **243**(5405): 290-293.
21. Ben-Neriah Y, Daley GQ, Mes-Masson AM, Witte ON, Baltimore D. The chronic myelogenous leukemia-specific P210 protein is the product of the bcr/abl hybrid gene. *Science* 1986; **233**(4760): 212-214.
22. Daley GQ, Van Etten RA, Baltimore D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science* 1990; **247**(4944): 824-830.
23. Verfaillie CM. Biology of chronic myelogenous leukemia. *Hematol Oncol Clin North Am* 1998; **12**(1): 1-29.
24. Westbrook CA, Hooberman AL, Spino C, Dodge RK, Larson RA, Davey F *et al.* Clinical significance of the BCR-ABL fusion gene in adult acute lymphoblastic leukemia: a Cancer and Leukemia Group B Study (8762). *Blood* 1992; **80**(12): 2983-2990.
25. Suryanarayan K, Hunger SP, Kohler S, Carroll AJ, Crist W, Link MP *et al.* Consistent involvement of the bcr gene by 9;22 breakpoints in pediatric acute leukemias. *Blood* 1991; **77**(2): 324-330.
26. Paietta E, Racevskis J, Bennett JM, Neuberg D, Cassileth PA, Rowe JM *et al.* Biologic heterogeneity in Philadelphia chromosome-positive acute leukemia with myeloid morphology: the Eastern Cooperative Oncology Group experience. *Leukemia* 1998; **12**(12): 1881-5.
27. Abelson HT, Rabstein LS. Lymphosarcoma: virus-induced thymic-independent disease in mice. *Cancer Res* 1970; **30**(8): 2213-2222.
28. Rosenberg N, Baltimore D, Scher CD. In vitro transformation of lymphoid cells by Abelson murine leukemia virus. *Proc.Natl.Acad.Sci.U.S.A* 1975; **72**(5): 1932-1936.

29. Scher CD, Siegler R. Direct transformation of 3T3 cells by Abelson murine leukaemia virus. *Nature* 1975; **253**(5494): 729-731.
30. Van dV, Reynolds FH, Jr., Nalewaik RP, Stephenson JR. The nonstructural component of the Abelson murine leukemia virus polyprotein P120 is encoded by newly acquired genetic sequences. *J Virol.* 1979; **32**(3): 1041-1045.
31. Witte ON, Goff S, Rosenberg N, Baltimore D. A transformation-defective mutant of Abelson murine leukemia virus lacks protein kinase activity. *Proc.Natl.Acad.Sci.U.S.A* 1980; **77**(8): 4993-4997.
32. Wang JY, Ledley F, Goff S, Lee R, Groner Y, Baltimore D. The mouse c-abl locus: molecular cloning and characterization. *Cell* 1984; **36**(2): 349-356.
33. Heisterkamp N, Groffen J, Stephenson JR, Spurr NK, Goodfellow PN, Solomon E *et al.* Chromosomal localization of human cellular homologues of two viral oncogenes. *Nature* 1982; **299**(5885): 747-749.
34. Bartram CR, de KA, Hagemeijer A, van AT, Geurts van KA, Bootsma D *et al.* Translocation of c-abl oncogene correlates with the presence of a Philadelphia chromosome in chronic myelocytic leukaemia. *Nature* 1983; **306**(5940): 277-280.
35. de KA, van Kessel AG, Grosveld G, Bartram CR, Hagemeijer A, Bootsma D *et al.* A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia. *Nature* 1982; **300**(5894): 765-767.
36. Groffen J, Stephenson JR, Heisterkamp N, Bartram C, de KA, Grosveld G. The human c-abl oncogene in the Philadelphia translocation. *J Cell Physiol Suppl* 1984; **3**: 179-191.
37. Heisterkamp N, Stephenson JR, Groffen J, Hansen PF, de KA, Bartram CR *et al.* Localization of the c-abl oncogene adjacent to a translocation break point in chronic myelocytic leukaemia. *Nature* 1983; **306**(5940): 239-242.

38. Blume-Jensen P, Hunter T. Oncogenic kinase signalling. *Nature* 2001; **411**(6835): 355-365.
39. McWhirter JR, Wang JY. Activation of tyrosinase kinase and microfilament-binding functions of c-abl by bcr sequences in bcr/abl fusion proteins. *Mol.Cell Biol.* 1991; **11**(3): 1553-1565.
40. Ito Y, Pandey P, Mishra N, Kumar S, Narula N, Kharbanda S *et al.* Targeting of the c-Abl tyrosine kinase to mitochondria in endoplasmic reticulum stress-induced apoptosis. *Mol Cell Biol* 2001; **21**(18): 6233-42.
41. Shaul Y. c-Abl: activation and nuclear targets. *Cell Death Differ* 2000; **7**(1): 10-6.
42. Yoshida K. Regulation for nuclear targeting of the Abl tyrosine kinase in response to DNA damage. *Adv Exp Med Biol* 2007; **604**: 155-65.
43. Pendergast AM. The Abl family kinases: mechanisms of regulation and signaling. *Adv.Cancer Res* 2002; **85**: 51-100.
44. Saglio G, Cilloni D. Abl: the prototype of oncogenic fusion proteins. *Cell Mol.Life Sci.* 2004; **61**(23): 2897-2911.
45. Qi X, Mochly-Rosen D. The PKCdelta -Abl complex communicates ER stress to the mitochondria - an essential step in subsequent apoptosis. *J Cell Sci* 2008; **121**(Pt 6): 804-13.
46. Goldman JM, Melo JV. Chronic myeloid leukemia--advances in biology and new approaches to treatment. *N Engl J Med* 2003; **349**(15): 1451-1464.
47. Shtivelman E, Lifshitz B, Gale RP, Roe BA, Canaani E. Alternative splicing of RNAs transcribed from the human abl gene and from the bcr-abl fused gene. *Cell* 1986; **47**(2): 277-284.

48. Hantschel O, Nagar B, Guettler S, Kretzschmar J, Dorey K, Kuriyan J *et al.* A myristoyl/phosphotyrosine switch regulates c-Abl. *Cell* 2003; **112**(6): 845-57.
49. Nagar B, Hantschel O, Young MA, Scheffzek K, Veach D, Bornmann W *et al.* Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* 2003; **112**(6): 859-71.
50. Xiong X, Cui P, Hossain S, Xu R, Warner B, Guo X *et al.* Allosteric inhibition of the nonMyristoylated c-Abl tyrosine kinase by phosphopeptides derived from Abi1/Hssh3bp1. *Biochim Biophys Acta* 2008; **1783**(5): 737-47.
51. Scheijen B, Griffin JD. Tyrosine kinase oncogenes in normal hematopoiesis and hematological disease. *Oncogene* 2002; **21**(21): 3314-3333.
52. Heldin CH. Dimerization of cell surface receptors in signal transduction. *Cell* 1995; **80**(2): 213-223.
53. Pawson T. Protein modules and signalling networks. *Nature* 1995; **373**(6515): 573-580.
54. Lewis JM, Baskaran R, Taagepera S, Schwartz MA, Wang JY. Integrin regulation of c-Abl tyrosine kinase activity and cytoplasmic-nuclear transport. *Proc.Natl.Acad.Sci.U.S.A* 1996; **93**(26): 15174-15179.
55. Van Etten RA, Jackson PK, Baltimore D, Sanders MC, Matsudaira PT, Janmey PA. The COOH terminus of the c-Abl tyrosine kinase contains distinct F- and G-actin binding domains with bundling activity. *J Cell Biol.* 1994; **124**(3): 325-340.
56. Wong S, Witte ON. The BCR-ABL story: bench to bedside and back. *Annu.Rev.Immunol.* 2004; **22**: 247-306.
57. Wang JY, Baltimore D. Cellular RNA homologous to the Abelson murine leukemia virus transforming gene: expression and relationship to the viral sequence. *Mol.Cell Biol.* 1983; **3**(5): 773-779.

58. McWhirter JR, Galasso DL, Wang JY. A coiled-coil oligomerization domain of Bcr is essential for the transforming function of Bcr-Abl oncoproteins. *Mol. Cell Biol.* 1993; **13**(12): 7587-7595.
59. Davis RL, Konopka JB, Witte ON. Activation of the c-abl oncogene by viral transduction or chromosomal translocation generates altered c-abl proteins with similar in vitro kinase properties. *Mol. Cell Biol.* 1985; **5**(1): 204-213.
60. Zhang X, Ren R. Bcr-Abl efficiently induces a myeloproliferative disease and production of excess interleukin-3 and granulocyte-macrophage colony-stimulating factor in mice: a novel model for chronic myelogenous leukemia. *Blood* 1998; **92**(10): 3829-3840.
61. Druker BJ, Tamura S, Buchdunger E, Ohno S, Segal GM, Fanning S *et al.* Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat Med* 1996; **2**(5): 561-566.
62. Muller AJ, Young JC, Pendergast AM, Pondel M, Landau NR, Littman DR *et al.* BCR first exon sequences specifically activate the BCR/ABL tyrosine kinase oncogene of Philadelphia chromosome-positive human leukemias. *Mol. Cell Biol.* 1991; **11**(4): 1785-1792.
63. He Y, Wertheim JA, Xu L, Miller JP, Karnell FG, Choi JK *et al.* The coiled-coil domain and Tyr177 of bcr are required to induce a murine chronic myelogenous leukemia-like disease by bcr/abl. *Blood* 2002; **99**(8): 2957-2968.
64. Million RP, Van Etten RA. The Grb2 binding site is required for the induction of chronic myeloid leukemia-like disease in mice by the Bcr/Abl tyrosine kinase. *Blood* 2000; **96**(2): 664-670.
65. Sattler M, Mohi MG, Pride YB, Quinnan LR, Malouf NA, Podar K *et al.* Critical role for Gab2 in transformation by BCR/ABL. *Cancer Cell* 2002; **1**(5): 479-492.

66. Goga A, McLaughlin J, Afar DE, Saffran DC, Witte ON. Alternative signals to RAS for hematopoietic transformation by the BCR-ABL oncogene. *Cell* 1995; **82**(6): 981-988.
67. Roumiantsev S, de AI, Varticovski L, Ilaria RL, Van Etten RA. The src homology 2 domain of Bcr/Abl is required for efficient induction of chronic myeloid leukemia-like disease in mice but not for lymphoid leukemogenesis or activation of phosphatidylinositol 3-kinase. *Blood* 2001; **97**(1): 4-13.
68. Zhang X, Wong R, Hao SX, Pear WS, Ren R. The SH2 domain of bcr-Abl is not required to induce a murine myeloproliferative disease; however, SH2 signaling influences disease latency and phenotype. *Blood* 2001; **97**(1): 277-287.
69. Cortez D, Reuther G, Pendergast AM. The Bcr-Abl tyrosine kinase activates mitogenic signaling pathways and stimulates G1-to-S phase transition in hematopoietic cells. *Oncogene* 1997; **15**(19): 2333-2342.
70. Sawyers CL, McLaughlin J, Witte ON. Genetic requirement for Ras in the transformation of fibroblasts and hematopoietic cells by the Bcr-Abl oncogene. *J Exp.Med* 1995; **181**(1): 307-313.
71. Sonoyama J, Matsumura I, Ezoe S, Satoh Y, Zhang X, Kataoka Y *et al.* Functional cooperation among Ras, STAT5, and phosphatidylinositol 3-kinase is required for full oncogenic activities of BCR/ABL in K562 cells. *J Biol.Chem* 2002; **277**(10): 8076-8082.
72. Janssen JW, Steenvoorden AC, Lyons J, Anger B, Bohlke JU, Bos JL *et al.* RAS gene mutations in acute and chronic myelocytic leukemias, chronic myeloproliferative disorders, and myelodysplastic syndromes. *Proc.Natl.Acad.Sci.U.S.A* 1987; **84**(24): 9228-9232.
73. Sawyers CL, Denny CT. Chronic myelomonocytic leukemia: Tel-a-kinase what Ets all about. *Cell* 1994; **77**(2): 171-173.

74. Pendergast AM, Quilliam LA, Cripe LD, Bassing CH, Dai Z, Li N *et al.* BCR-ABL-induced oncogenesis is mediated by direct interaction with the SH2 domain of the GRB-2 adaptor protein. *Cell* 1993; **75**(1): 175-185.
75. Raitano AB, Halpern JR, Hambuch TM, Sawyers CL. The Bcr-Abl leukemia oncogene activates Jun kinase and requires Jun for transformation. *Proc.Natl.Acad.Sci.U.S.A* 1995; **92**(25): 11746-11750.
76. Cortez D, Stoica G, Pierce JH, Pendergast AM. The BCR-ABL tyrosine kinase inhibits apoptosis by activating a Ras-dependent signaling pathway. *Oncogene* 1996; **13**(12): 2589-2594.
77. Kinoshita T, Yokota T, Arai K, Miyajima A. Regulation of Bcl-2 expression by oncogenic Ras protein in hematopoietic cells. *Oncogene* 1995; **10**(11): 2207-2212.
78. Yang MY, Liu TC, Chang JG, Lin PM, Lin SF. JunB gene expression is inactivated by methylation in chronic myeloid leukemia. *Blood* 2003; **101**(8): 3205-3211.
79. Passegue E, Jochum W, Schorpp-Kistner M, Mohle-Steinlein U, Wagner EF. Chronic myeloid leukemia with increased granulocyte progenitors in mice lacking junB expression in the myeloid lineage. *Cell* 2001; **104**(1): 21-32.
80. Skorski T, Kanakaraj P, Nieborowska-Skorska M, Ratajczak MZ, Wen SC, Zon G *et al.* Phosphatidylinositol-3 kinase activity is regulated by BCR/ABL and is required for the growth of Philadelphia chromosome-positive cells. *Blood* 1995; **86**(2): 726-736.
81. Vivanco I, Sawyers CL. The phosphatidylinositol 3-Kinase AKT pathway in human cancer. *Nat Rev.Cancer* 2002; **2**(7): 489-501.
82. Sattler M, Mohi MG, Pride YB, Quinnan LR, Malouf NA, Podar K *et al.* Critical role for Gab2 in transformation by BCR/ABL. *Cancer Cell* 2002; **1**(5): 479-92.

83. Skorski T, Bellacosa A, Nieborowska-Skorska M, Majewski M, Martinez R, Choi JK *et al.* Transformation of hematopoietic cells by BCR/ABL requires activation of a PI-3k/Akt-dependent pathway. *EMBO J* 1997; **16**(20): 6151-6161.
84. Ly C, Arechiga AF, Melo JV, Walsh CM, Ong ST. Bcr-Abl kinase modulates the translation regulators ribosomal protein S6 and 4E-BP1 in chronic myelogenous leukemia cells via the mammalian target of rapamycin. *Cancer Res* 2003; **63**(18): 5716-22.
85. Mohi MG, Boulton C, Gu TL, Sternberg DW, Neuberg D, Griffin JD *et al.* Combination of rapamycin and protein tyrosine kinase (PTK) inhibitors for the treatment of leukemias caused by oncogenic PTKs. *Proc Natl Acad Sci U S A* 2004; **101**(9): 3130-5.
86. Prabhu S, Saadat D, Zhang M, Halbur L, Fruehauf JP, Ong ST. A novel mechanism for Bcr-Abl action: Bcr-Abl-mediated induction of the eIF4F translation initiation complex and mRNA translation. *Oncogene* 2007; **26**(8): 1188-200.
87. Gera JF, Mellinghoff IK, Shi Y, Rettig MB, Tran C, Hsu JH *et al.* AKT activity determines sensitivity to mammalian target of rapamycin (mTOR) inhibitors by regulating cyclin D1 and c-myc expression. *J Biol Chem* 2004; **279**(4): 2737-46.
88. Carayol N, Katsoulidis E, Sassano A, Altman JK, Druker BJ, Platanias LC. Suppression of programmed cell death 4 (PDCD4) protein expression by BCR-ABL-regulated engagement of the mTOR/p70 S6 kinase pathway. *J Biol Chem* 2008; **283**(13): 8601-10.
89. Neshat MS, Raitano AB, Wang HG, Reed JC, Sawyers CL. The survival function of the Bcr-Abl oncogene is mediated by Bad-dependent and -independent pathways: roles for phosphatidylinositol 3-kinase and Raf. *Mol. Cell Biol.* 2000; **20**(4): 1179-1186.
90. Sanchez-Garcia I, Grutz G. Tumorigenic activity of the BCR-ABL oncogenes is mediated by BCL2. *Proc. Natl. Acad. Sci. U.S.A* 1995; **92**(12): 5287-5291.

91. Danial NN, Pernis A, Rothman PB. Jak-STAT signaling induced by the v-abl oncogene. *Science* 1995; **269**(5232): 1875-7.
92. Wilson-Rawls J, Xie S, Liu J, Laneuville P, Arlinghaus RB. P210 Bcr-Abl interacts with the interleukin 3 receptor beta(c) subunit and constitutively induces its tyrosine phosphorylation. *Cancer Res* 1996; **56**(15): 3426-30.
93. Ilaria RL, Jr., Van Etten RA. P210 and P190(BCR/ABL) induce the tyrosine phosphorylation and DNA binding activity of multiple specific STAT family members. *J Biol.Chem* 1996; **271**(49): 31704-31710.
94. Chai SK, Nichols GL, Rothman P. Constitutive activation of JAKs and STATs in BCR-Abl-expressing cell lines and peripheral blood cells derived from leukemic patients. *J Immunol* 1997; **159**(10): 4720-8.
95. Xie S, Wang Y, Liu J, Sun T, Wilson MB, Smithgall TE *et al.* Involvement of Jak2 tyrosine phosphorylation in Bcr-Abl transformation. *Oncogene* 2001; **20**(43): 6188-95.
96. Tao WJ, Lin H, Sun T, Samanta AK, Arlinghaus R. BCR-ABL oncogenic transformation of NIH 3T3 fibroblasts requires the IL-3 receptor. *Oncogene* 2008; **27**(22): 3194-200.
97. Sawyers CL. The role of myc in transformation by BCR-ABL. *Leuk.Lymphoma* 1993; **11 Suppl 1**: 45-46.
98. Zou X, Rudchenko S, Wong K, Calame K. Induction of c-myc transcription by the v-Abl tyrosine kinase requires Ras, Raf1, and cyclin-dependent kinases. *Genes Dev.* 1997; **11**(5): 654-662.
99. Samanta AK, Lin H, Sun T, Kantarjian H, Arlinghaus RB. Janus kinase 2: a critical target in chronic myelogenous leukemia. *Cancer Res* 2006; **66**(13): 6468-72.

100. Ress A, Moelling K. Bcr is a negative regulator of the Wnt signalling pathway. *EMBO Rep* 2005; **6**(11): 1095-1100.
101. Gesbert F, Griffin JD. Bcr/Abl activates transcription of the Bcl-X gene through STAT5. *Blood* 2000; **96**(6): 2269-2276.
102. Sexl V, Piekorz R, Moriggl R, Rohrer J, Brown MP, Bunting KD *et al.* Stat5a/b contribute to interleukin 7-induced B-cell precursor expansion, but abl- and bcr/abl-induced transformation are independent of stat5. *Blood* 2000; **96**(6): 2277-2283.
103. Bunting KD, Bradley HL, Hawley TS, Moriggl R, Sorrentino BP, Ihle JN. Reduced lymphomyeloid repopulating activity from adult bone marrow and fetal liver of mice lacking expression of STAT5. *Blood* 2002; **99**(2): 479-87.
104. Hoelbl A, Kovacic B, Kerenyi MA, Simma O, Warsch W, Cui Y *et al.* Clarifying the role of Stat5 in lymphoid development and Abelson-induced transformation. *Blood* 2006; **107**(12): 4898-906.
105. Scherr M, Chaturvedi A, Battmer K, Dallmann I, Schultheis B, Ganser A *et al.* Enhanced sensitivity to inhibition of SHP2, STAT5, and Gab2 expression in chronic myeloid leukemia (CML). *Blood* 2006; **107**(8): 3279-87.
106. Baum KJ, Ren R. Effect of Ras Inhibition in Hematopoiesis and BCR/ABL Leukemogenesis. *J Hematol Oncol* 2008; **1**(1): 5.
107. Danhauser-Riedl S, Warmuth M, Druker BJ, Emmerich B, Hallek M. Activation of Src kinases p53/56lyn and p59hck by p210bcr/abl in myeloid cells. *Cancer Res* 1996; **56**(15): 3589-96.
108. Hu Y, Liu Y, Pelletier S, Buchdunger E, Warmuth M, Fabbro D *et al.* Requirement of Src kinases Lyn, Hck and Fgr for BCR-ABL1-induced B-lymphoblastic leukemia but not chronic myeloid leukemia. *Nat Genet* 2004; **36**(5): 453-61.

109. Schmidt M, Nagel S, Proba J, Thiede C, Ritter M, Waring JF *et al.* Lack of interferon consensus sequence binding protein (ICSBP) transcripts in human myeloid leukemias. *Blood* 1998; **91**(1): 22-29.
110. Holtschke T, Lohler J, Kanno Y, Fehr T, Giese N, Rosenbauer F *et al.* Immunodeficiency and chronic myelogenous leukemia-like syndrome in mice with a targeted mutation of the ICSBP gene. *Cell* 1996; **87**(2): 307-317.
111. Zandy NL, Pendergast AM. Abl tyrosine kinases modulate cadherin-dependent adhesion upstream and downstream of Rho family GTPases. *Cell Cycle* 2008; **7**(4): 444-8.
112. Gordon MY, Dowding CR, Riley GP, Goldman JM, Greaves MF. Altered adhesive interactions with marrow stroma of haematopoietic progenitor cells in chronic myeloid leukaemia. *Nature* 1987; **328**(6128): 342-344.
113. Aiuti A, Turchetto L, Cota M, Cipponi A, Brambilla A, Arcelloni C *et al.* Human CD34(+) cells express CXCR4 and its ligand stromal cell-derived factor-1. Implications for infection by T-cell tropic human immunodeficiency virus. *Blood* 1999; **94**(1): 62-73.
114. Zou YR, Kottmann AH, Kuroda M, Taniuchi I, Littman DR. Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development. *Nature* 1998; **393**(6685): 595-9.
115. Salgia R, Quackenbush E, Lin J, Souchkova N, Sattler M, Ewaniuk DS *et al.* The BCR/ABL oncogene alters the chemotactic response to stromal-derived factor-1alpha. *Blood* 1999; **94**(12): 4233-46.
116. Chen YY, Malik M, Tomkowicz BE, Collman RG, Ptasznik A. BCR-ABL1 alters SDF-1alpha-mediated adhesive responses through the beta2 integrin LFA-1 in leukemia cells. *Blood* 2008; **111**(10): 5182-6.

117. Bueno MJ, Perez de Castro I, Gomez de Cedron M, Santos J, Calin GA, Cigudosa JC *et al.* Genetic and epigenetic silencing of microRNA-203 enhances ABL1 and BCR-ABL1 oncogene expression. *Cancer Cell* 2008; **13**(6): 496-506.
118. Ambros V. The functions of animal microRNAs. *Nature* 2004; **431**(7006): 350-5.
119. Mullighan CG, Miller CB, Radtke I, Phillips LA, Dalton J, Ma J *et al.* BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* 2008; **453**(7191): 110-4.
120. Melo JV. BCR-ABL gene variants. *Baillieres Clin Haematol.* 1997; **10**(2): 203-222.
121. Hochhaus A, Reiter A, Skladny H, Melo JV, Sick C, Berger U *et al.* A novel BCR-ABL fusion gene (e6a2) in a patient with Philadelphia chromosome-negative chronic myelogenous leukemia. *Blood* 1996; **88**(6): 2236-2240.
122. Burmeister T, Schwartz S, Taubald A, Jost E, Lipp T, Schneller F *et al.* Atypical BCR-ABL mRNA transcripts in adult acute lymphoblastic leukemia. *Haematologica* 2007; **92**(12): 1699-702.
123. Fujisawa S, Nakamura S, Naito K, Kobayashi M, Ohnishi K. A variant transcript, e1a3, of the minor BCR-ABL fusion gene in acute lymphoblastic leukemia: case report and review of the literature. *Int J Hematol* 2008; **87**(2): 184-8.
124. Melo JV, Gordon DE, Cross NC, Goldman JM. The ABL-BCR fusion gene is expressed in chronic myeloid leukemia. *Blood* 1993; **81**(1): 158-165.
125. Huntly BJ, Reid AG, Bench AJ, Campbell LJ, Telford N, Shepherd P *et al.* Deletions of the derivative chromosome 9 occur at the time of the Philadelphia translocation and provide a powerful and independent prognostic indicator in chronic myeloid leukemia. *Blood* 2001; **98**(6): 1732-1738.

126. Kolomietz E, Al-Maghrabi J, Brennan S, Karaskova J, Minkin S, Lipton J *et al.* Primary chromosomal rearrangements of leukemia are frequently accompanied by extensive submicroscopic deletions and may lead to altered prognosis. *Blood* 2001; **97**(11): 3581-8.
127. Sinclair PB, Nacheva EP, Leversha M, Telford N, Chang J, Reid A *et al.* Large deletions at the t(9;22) breakpoint are common and may identify a poor-prognosis subgroup of patients with chronic myeloid leukemia. *Blood* 2000; **95**(3): 738-743.
128. Kreil S, Pfirrmann M, Haferlach C, Waghorn K, Chase A, Hehlmann R *et al.* Heterogeneous prognostic impact of derivative chromosome 9 deletions in chronic myelogenous leukemia. *Blood* 2007; **110**(4): 1283-90.
129. Savage DG, Antman KH. Imatinib mesylate--a new oral targeted therapy. *N Engl J Med* 2002; **346**(9): 683-693.
130. Buchdunger E, Zimmermann J, Mett H, Meyer T, Muller M, Druker BJ *et al.* Inhibition of the Abl protein-tyrosine kinase in vitro and in vivo by a 2-phenylaminopyrimidine derivative. *Cancer Res* 1996; **56**(1): 100-104.
131. Deininger MW, Goldman JM, Lydon N, Melo JV. The tyrosine kinase inhibitor CGP57148B selectively inhibits the growth of BCR-ABL-positive cells. *Blood* 1997; **90**(9): 3691-3698.
132. Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, Ford JM *et al.* Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med* 2001; **344**(14): 1031-1037.
133. Kantarjian H, Sawyers C, Hochhaus A, Guilhot F, Schiffer C, Gambacorti-Passerini C *et al.* Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. *N Engl J Med* 2002; **346**(9): 645-652.

134. O'Brien SG, Guilhot F, Larson RA, Gathmann I, Baccarani M, Cervantes F *et al.* Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. *N Engl J Med* 2003; **348**(11): 994-1004.
135. Druker BJ, Guilhot F, O'Brien SG, Gathmann I, Kantarjian H, Gattermann N *et al.* Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N Engl J Med* 2006; **355**(23): 2408-17.
136. Hughes TP, Kaeda J, Branford S, Rudzki Z, Hochhaus A, Hensley ML *et al.* Frequency of major molecular responses to imatinib or interferon alfa plus cytarabine in newly diagnosed chronic myeloid leukemia. *N Engl J Med* 2003; **349**(15): 1423-32.
137. Goldman J. Monitoring minimal residual disease in BCR-ABL-positive chronic myeloid leukemia in the imatinib era. *Curr Opin Hematol* 2005; **12**(1): 33-9.
138. Branford S, Rudzki Z, Walsh S, Grigg A, Arthur C, Taylor K *et al.* High frequency of point mutations clustered within the adenosine triphosphate-binding region of BCR/ABL in patients with chronic myeloid leukemia or Ph-positive acute lymphoblastic leukemia who develop imatinib (STI571) resistance. *Blood* 2002; **99**(9): 3472-3475.
139. O'Hare T, Eide CA, Deininger MW. Bcr-Abl kinase domain mutations, drug resistance, and the road to a cure for chronic myeloid leukemia. *Blood* 2007; **110**(7): 2242-9.
140. Azam M, Latek RR, Daley GQ. Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL. *Cell* 2003; **112**(6): 831-843.
141. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* 2000; **289**(5486): 1938-42.

142. Druker BJ, Sawyers CL, Kantarjian H, Resta DJ, Reese SF, Ford JM *et al.* Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *N Engl J Med* 2001; **344**(14): 1038-1042.
143. Soverini S, Colarossi S, Gnani A, Rosti G, Castagnetti F, Poerio A *et al.* Contribution of ABL kinase domain mutations to imatinib resistance in different subsets of Philadelphia-positive patients: by the GIMEMA Working Party on Chronic Myeloid Leukemia. *Clin Cancer Res* 2006; **12**(24): 7374-9.
144. Wu J, Meng F, Lu H, Kong L, Bornmann W, Peng Z *et al.* Lyn regulates BCR-ABL and Gab2 tyrosine phosphorylation and c-Cbl protein stability in imatinib-resistant chronic myelogenous leukemia cells. *Blood* 2008; **111**(7): 3821-9.
145. Meyn MA, 3rd, Wilson MB, Abdi FA, Fahey N, Schiavone AP, Wu J *et al.* Src family kinases phosphorylate the Bcr-Abl SH3-SH2 region and modulate Bcr-Abl transforming activity. *J Biol Chem* 2006; **281**(41): 30907-16.
146. Hu Y, Swerdlow S, Duffy TM, Weinmann R, Lee FY, Li S. Targeting multiple kinase pathways in leukemic progenitors and stem cells is essential for improved treatment of Ph⁺ leukemia in mice. *Proc Natl Acad Sci U S A* 2006; **103**(45): 16870-5.
147. Melnick JS, Janes J, Kim S, Chang JY, Sipes DG, Gunderson D *et al.* An efficient rapid system for profiling the cellular activities of molecular libraries. *Proc Natl Acad Sci U S A* 2006; **103**(9): 3153-8.
148. Shah NP, Tran C, Lee FY, Chen P, Norris D, Sawyers CL. Overriding imatinib resistance with a novel ABL kinase inhibitor. *Science* 2004; **305**(5682): 399-401.
149. Tokarski JS, Newitt JA, Chang CY, Cheng JD, Wittekind M, Kiefer SE *et al.* The structure of Dasatinib (BMS-354825) bound to activated ABL kinase domain elucidates its inhibitory activity against imatinib-resistant ABL mutants. *Cancer Res* 2006; **66**(11): 5790-7.

150. Baccarani M, Saglio G, Goldman J, Hochhaus A, Simonsson B, Appelbaum F *et al.* Evolving concepts in the management of chronic myeloid leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. *Blood* 2006; **108**(6): 1809-20.
151. Kean SJ. Dasatinib: in chronic myeloid leukemia and Philadelphia chromosome-positive acute lymphoblastic leukemia. *BioDrugs* 2008; **22**(1): 59-69.
152. Cortes J, Rousselot P, Kim DW, Ritchie E, Hamerschlak N, Coutre S *et al.* Dasatinib induces complete hematologic and cytogenetic responses in patients with imatinib-resistant or -intolerant chronic myeloid leukemia in blast crisis. *Blood* 2007; **109**(8): 3207-13.
153. Hochhaus A. Management of Bcr-Abl-positive leukemias with dasatinib. *Expert Rev Anticancer Ther* 2007; **7**(11): 1529-36.
154. Plosker GL, Robinson DM. Nilotinib. *Drugs* 2008; **68**(4): 449-59; discussion 460-1.
155. Weisberg E, Manley PW, Breitenstein W, Bruggen J, Cowan-Jacob SW, Ray A *et al.* Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. *Cancer Cell* 2005; **7**(2): 129-41.
156. Adrian FJ, Ding Q, Sim T, Velentza A, Sloan C, Liu Y *et al.* Allosteric inhibitors of Bcr-abl-dependent cell proliferation. *Nat Chem Biol* 2006; **2**(2): 95-102.
157. Giles FJ, Cortes J, Jones D, Bergstrom D, Kantarjian H, Freedman SJ. MK-0457, a novel kinase inhibitor, is active in patients with chronic myeloid leukemia or acute lymphocytic leukemia with the T315I BCR-ABL mutation. *Blood* 2007; **109**(2): 500-2.

158. Harrington EA, Bebbington D, Moore J, Rasmussen RK, Ajose-Adeogun AO, Nakayama T *et al.* VX-680, a potent and selective small-molecule inhibitor of the Aurora kinases, suppresses tumor growth in vivo. *Nat Med* 2004; **10**(3): 262-7.
159. Noronha G, Cao J, Chow CP, Dneprovskaja E, Fine RM, Hood J *et al.* Inhibitors of ABL and the ABL-T315I mutation. *Curr Top Med Chem* 2008; **8**(10): 905-21.
160. Holyoake T, Jiang X, Eaves C, Eaves A. Isolation of a highly quiescent subpopulation of primitive leukemic cells in chronic myeloid leukemia. *Blood* 1999; **94**(6): 2056-64.
161. Bhatia R, Holtz M, Niu N, Gray R, Snyder DS, Sawyers CL *et al.* Persistence of malignant hematopoietic progenitors in chronic myelogenous leukemia patients in complete cytogenetic remission following imatinib mesylate treatment. *Blood* 2003; **101**(12): 4701-4707.
162. Graham SM, Jorgensen HG, Allan E, Pearson C, Alcorn MJ, Richmond L *et al.* Primitive, quiescent, Philadelphia-positive stem cells from patients with chronic myeloid leukemia are insensitive to STI571 in vitro. *Blood* 2002; **99**(1): 319-325.
163. Holtz MS, Forman SJ, Bhatia R. Nonproliferating CML CD34+ progenitors are resistant to apoptosis induced by a wide range of proapoptotic stimuli. *Leukemia* 2005; **19**(6): 1034-41.
164. Konig H, Holtz M, Modi H, Manley P, Holyoake TL, Forman SJ *et al.* Enhanced BCR-ABL kinase inhibition does not result in increased inhibition of downstream signaling pathways or increased growth suppression in CML progenitors. *Leukemia* 2008; **22**(4): 748-55.
165. Copland M, Hamilton A, Elrick LJ, Baird JW, Allan EK, Jordanides N *et al.* Dasatinib (BMS-354825) targets an earlier progenitor population than imatinib in primary CML but does not eliminate the quiescent fraction. *Blood* 2006; **107**(11): 4532-9.

166. Elrick LJ, Jorgensen HG, Mountford JC, Holyoake TL. Punish the parent not the progeny. *Blood* 2005; **105**(5): 1862-1866.
167. Buchdunger E, O'Reilly T, Wood J. Pharmacology of imatinib (STI571). *Eur.J Cancer* 2002; **38 Suppl 5**: S28-S36.
168. Cross NC, Reiter A. Fibroblast growth factor receptor and platelet-derived growth factor receptor abnormalities in eosinophilic myeloproliferative disorders. *Acta Haematol* 2008; **119**(4): 199-206.
169. Longley BJ, Tyrrell L, Lu SZ, Ma YS, Langley K, Ding TG *et al.* Somatic c-KIT activating mutation in urticaria pigmentosa and aggressive mastocytosis: establishment of clonality in a human mast cell neoplasm. *Nat Genet* 1996; **12**(3): 312-314.
170. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood* 2002; **100**(7): 2292-302.
171. James C, Ugo V, Le Couedic JP, Staerk J, Delhommeau F, Lacout C *et al.* A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature* 2005; **434**(7037): 1144-8.
172. Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJ *et al.* Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* 2005; **7**(4): 387-397.
173. Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S *et al.* Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* 2005; **365**(9464): 1054-61.
174. Saharinen P, Vihinen M, Silvennoinen O. Autoinhibition of Jak2 tyrosine kinase is dependent on specific regions in its pseudokinase domain. *Mol.Biol.Cell* 2003; **14**(4): 1448-1459.

175. Lindauer K, Loerting T, Liedl KR, Kroemer RT. Prediction of the structure of human Janus kinase 2 (JAK2) comprising the two carboxy-terminal domains reveals a mechanism for autoregulation. *Protein Eng* 2001; **14**(1): 27-37.
176. Jones AV, Kreil S, Zoi K, Waghorn K, Curtis C, Zhang L *et al*. Widespread occurrence of the JAK2 V617F mutation in chronic myeloproliferative disorders. *Blood* 2005; **106**(6): 2162-8.
177. Levine RL, Loriaux M, Huntly BJ, Loh ML, Beran M, Stoffregen E *et al*. The JAK2V617F activating mutation occurs in chronic myelomonocytic leukemia and acute myeloid leukemia, but not in acute lymphoblastic leukemia or chronic lymphocytic leukemia. *Blood* 2005; **106**(10): 3377-9.
178. Steensma DP, Dewald GW, Lasho TL, Powell HL, McClure RF, Levine RL *et al*. The JAK2 V617F activating tyrosine kinase mutation is an infrequent event in both "atypical" myeloproliferative disorders and myelodysplastic syndromes. *Blood* 2005; **106**(4): 1207-9.
179. Scott LM, Scott MA, Campbell PJ, Green AR. Progenitors homozygous for the V617F mutation occur in most patients with polycythemia vera, but not essential thrombocythemia. *Blood* 2006; **108**(7): 2435-7.
180. Kralovics R, Passamonti F, Buser AS, Teo SS, Tiedt R, Passweg JR *et al*. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med* 2005; **352**(17): 1779-90.
181. Mesa RA, Powell H, Lasho T, Dewald G, McClure R, Tefferi A. JAK2(V617F) and leukemic transformation in myelofibrosis with myeloid metaplasia. *Leuk Res* 2006; **30**(11): 1457-60.
182. Campbell PJ, Green AR. The myeloproliferative disorders. *N Engl J Med* 2006; **355**(23): 2452-66.

183. Baxter EJ, Hochhaus A, Bolufer P, Reiter A, Fernandez JM, Senent L *et al.* The t(4;22)(q12;q11) in atypical chronic myeloid leukaemia fuses BCR to PDGFRA. *Hum.Mol.Genet* 2002; **11**(12): 1391-1397.
184. Cools J, DeAngelo DJ, Gotlib J, Stover EH, Legare RD, Cortes J *et al.* A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. *N Engl J Med* 2003; **348**(13): 1201-1214.
185. Fauci AS, Harley JB, Roberts WC, Ferrans VJ, Gralnick HR, Bjornson BH. NIH conference. The idiopathic hypereosinophilic syndrome. Clinical, pathophysiologic, and therapeutic considerations. *Ann Intern Med* 1982; **97**(1): 78-92.
186. Brito-Babapulle F. Clonal eosinophilic disorders and the hypereosinophilic syndrome. *Blood Rev.* 1997; **11**(3): 129-145.
187. Cools J, Stover EH, Wlodarska I, Marynen P, Gilliland DG. The FIP1L1-PDGFRalpha kinase in hypereosinophilic syndrome and chronic eosinophilic leukemia. *Curr Opin.Hematol* 2004; **11**(1): 51-57.
188. Bain BJ. Eosinophilic leukaemias and the idiopathic hypereosinophilic syndrome. *Br.J Haematol.* 1996; **95**(1): 2-9.
189. Chang HW, Leong KH, Koh DR, Lee SH. Clonality of isolated eosinophils in the hypereosinophilic syndrome. *Blood* 1999; **93**(5): 1651-1657.
190. Gleich GJ, Leiferman KM, Pardanani A, Tefferi A, Butterfield JH. Treatment of hypereosinophilic syndrome with imatinib mesilate. *Lancet* 2002; **359**(9317): 1577-1578.
191. Buchdunger E, Cioffi CL, Law N, Stover D, Ohno-Jones S, Druker BJ *et al.* Abl protein-tyrosine kinase inhibitor STI571 inhibits in vitro signal transduction mediated by c-kit and platelet-derived growth factor receptors. *J Pharmacol.Exp.Ther.* 2000; **295**(1): 139-145.

192. Gorre ME, Sawyers CL. Molecular mechanisms of resistance to STI571 in chronic myeloid leukemia. *Curr Opin.Hematol* 2002; **9**(4): 303-307.
193. Griffin JH, Leung J, Bruner RJ, Caligiuri MA, Briesewitz R. Discovery of a fusion kinase in EOL-1 cells and idiopathic hypereosinophilic syndrome. *Proc.Natl.Acad.Sci.U.S.A* 2003; **100**(13): 7830-7835.
194. Baccarani M, Cilloni D, Rondoni M, Ottaviani E, Messa F, Merante S *et al.* The efficacy of imatinib mesylate in patients with FIP1L1-PDGFRalpha-positive hypereosinophilic syndrome. Results of a multicenter prospective study. *Haematologica* 2007; **92**(9): 1173-9.
195. Jovanovic JV, Score J, Waghorn K, Cilloni D, Gottardi E, Metzgeroth G *et al.* Low-dose imatinib mesylate leads to rapid induction of major molecular responses and achievement of complete molecular remission in FIP1L1-PDGFRalpha-positive chronic eosinophilic leukemia. *Blood* 2007; **109**(11): 4635-4640.
196. Cools J, Stover EH, Gilliland DG. Detection of the FIP1L1-PDGFRalpha fusion in idiopathic hypereosinophilic syndrome and chronic eosinophilic leukemia. *Methods Mol.Med.* 2006; **125**: 177-187.
197. Vigna E, Lucia E, Gentile M, Mazzone C, Bisconte MG, Gentile C *et al.* PDGFRalpha/FIP1L1-positive chronic eosinophilic leukemia presenting with retro-orbital localization: efficacy of imatinib treatment. *Cancer Chemother.Pharmacol.* 2008; **61**(4): 713-716.
198. Rothenberg ME, Klion AD, Roufosse FE, Kahn JE, Weller PF, Simon HU *et al.* Treatment of patients with the hypereosinophilic syndrome with mepolizumab. *N.Engl.J.Med.* 2008; **358**(12): 1215-1228.
199. Richardson C, Jasin M. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* 2000; **405**(6787): 697-700.

200. Neves H, Ramos C, da Silva MG, Parreira A, Parreira L. The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. *Blood* 1999; **93**(4): 1197-1207.
201. Roix JJ, McQueen PG, Munson PJ, Parada LA, Misteli T. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nat Genet* 2003; **34**(3): 287-291.
202. Penserga ET, Skorski T. Fusion tyrosine kinases: a result and cause of genomic instability. *Oncogene* 2007; **26**(1): 11-20.
203. Khanna KK, Jackson SP. DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet* 2001; **27**(3): 247-254.
204. Tycko B, Sklar J. Chromosomal translocations in lymphoid neoplasia: a reappraisal of the recombinase model. *Cancer Cells* 1990; **2**(1): 1-8.
205. Bergsagel PL, Kuehl WM. Chromosome translocations in multiple myeloma. *Oncogene* 2001; **20**(40): 5611-5622.
206. Pasqualucci L, Bhagat G, Jankovic M, Compagno M, Smith P, Muramatsu M *et al.* AID is required for germinal center-derived lymphomagenesis. *Nat Genet* 2008; **40**(1): 108-12.
207. Marculescu R, Le T, Simon P, Jaeger U, Nadel B. V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. *J Exp Med* 2002; **195**(1): 85-98.
208. Aplan PD, Lombardi DP, Ginsberg AM, Cossman J, Bertness VL, Kirsch IR. Disruption of the human SCL locus by "illegitimate" V-(D)-J recombinase activity. *Science* 1990; **250**(4986): 1426-1429.

209. Taub R, Kirsch I, Morton C, Lenoir G, Swan D, Tronick S *et al.* Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. *Proc.Natl.Acad.Sci.U.S.A* 1982; **79**(24): 7837-7841.
210. Ramiro AR, Jankovic M, Eisenreich T, Difilippantonio S, Chen-Kiang S, Muramatsu M *et al.* AID is required for c-myc/IgH chromosome translocations in vivo. *Cell* 2004; **118**(4): 431-8.
211. Bose S, Deininger M, Gora-Tybor J, Goldman JM, Melo JV. The presence of typical and atypical BCR-ABL fusion genes in leukocytes of normal individuals: biologic significance and implications for the assessment of minimal residual disease. *Blood* 1998; **92**(9): 3362-7.
212. Surralles J, Puerto S, Ramirez MJ, Creus A, Marcos R, Mullenders LH *et al.* Links between chromatin structure, DNA repair and chromosome fragility. *Mutat.Res* 1998; **404**(1-2): 39-44.
213. Jeffs AR, Benjes SM, Smith TL, Sowerby SJ, Morris CM. The BCR gene recombines preferentially with Alu elements in complex BCR-ABL translocations of chronic myeloid leukaemia. *Hum Mol Genet* 1998; **7**(5): 767-76.
214. Chen SJ, Chen Z, Font MP, d'Auriol L, Larsen CJ, Berger R. Structural alterations of the BCR and ABL genes in Ph1 positive acute leukemias with rearrangements in the BCR gene first intron: further evidence implicating Alu sequences in the chromosome translocation. *Nucleic Acids Res* 1989; **17**(19): 7631-7642.
215. Toth G, Jurka J. Repetitive DNA in and around translocation breakpoints of the Philadelphia chromosome. *Gene* 1994; **140**(2): 285-288.
216. van der Feltz MJ, Shivji MK, Allen PB, Heisterkamp N, Groffen J, Wiedemann LM. Nucleotide sequence of both reciprocal translocation junction regions in a patient with Ph positive acute lymphoblastic leukaemia, with a breakpoint within the first intron of the BCR gene. *Nucleic Acids Res* 1989; **17**(1): 1-10.

217. Aplan PD, Raimondi SC, Kirsch IR. Disruption of the SCL gene by a t(1;3) translocation in a patient with T cell acute lymphoblastic leukemia. *J Exp.Med* 1992; **176**(5): 1303-1310.
218. Thandla SP, Ploski JE, Raza-Egilmez SZ, Chhalliyil PP, Block AW, de Jong PJ *et al.* ETV6-AML1 translocation breakpoints cluster near a purine/pyrimidine repeat region in the ETV6 gene. *Blood* 1999; **93**(1): 293-299.
219. Strick R, Zhang Y, Emmanuel N, Strissel PL. Common chromatin structures at breakpoint cluster regions may lead to chromosomal translocations found in chronic and acute leukemias. *Hum Genet* 2006; **119**(5): 479-95.
220. Ploski JE, Aplan PD. Characterization of DNA fragmentation events caused by genotoxic and non-genotoxic agents. *Mutat.Res* 2001; **473**(2): 169-180.
221. Stanulla M, Wang J, Chervinsky DS, Thandla S, Aplan PD. DNA cleavage within the MLL breakpoint cluster region is a specific event which occurs as part of higher-order chromatin fragmentation during the initial stages of apoptosis. *Mol.Cell Biol.* 1997; **17**(7): 4070-4079.
222. Khodarev NN, Sokolova IA, Vaughan AT. Abortive apoptosis as an initiator of chromosomal translocations. *Med Hypotheses* 1999; **52**(5): 373-376.
223. Reichel M, Gillert E, Nilson I, Siegler G, Greil J, Fey GH *et al.* Fine structure of translocation breakpoints in leukemic blasts with chromosomal translocation t(4;11): the DNA damage-repair model of translocation. *Oncogene* 1998; **17**(23): 3035-3044.
224. Reiter A, Saussele S, Grimwade D, Wiemels JL, Segal MR, Lafage-Pochitaloff M *et al.* Genomic anatomy of the specific reciprocal translocation t(15;17) in acute promyelocytic leukemia. *Genes Chromosomes Cancer* 2003; **36**(2): 175-188.

225. Mattarucchi E, Guerini V, Rambaldi A, Campiotti L, Venco A, Pasquali F *et al.* Microhomologies and interspersed repeat elements at genomic breakpoints in chronic myeloid leukemia. *Genes Chromosomes Cancer* 2008; **47**(7): 625-32.
226. Vilenchik MM, Knudson AG. Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proc.Natl.Acad.Sci.U.S.A* 2003; **100**(22): 12871-12876.
227. Preston DL, Kusumi S, Tomonaga M, Izumi S, Ron E, Kuramoto A *et al.* Cancer incidence in atomic bomb survivors. Part III. Leukemia, lymphoma and multiple myeloma, 1950-1987. *Radiat Res* 1994; **137**(2 Suppl): S68-97.
228. Deininger MW, Bose S, Gora-Tybor J, Yan XH, Goldman JM, Melo JV. Selective induction of leukemia-associated fusion genes by high-dose ionizing radiation. *Cancer Res* 1998; **58**(3): 421-5.
229. Digweed M, Sperling K. Nijmegen breakage syndrome: clinical manifestation of defective response to DNA double-strand breaks. *DNA Repair (Amst)* 2004; **3**(8-9): 1207-1217.
230. Mills KD, Ferguson DO, Alt FW. The role of DNA breaks in genomic instability and tumorigenesis. *Immunol.Rev.* 2003; **194**: 77-95.
231. Slany RK, Lavau C, Cleary ML. The oncogenic capacity of HRX-ENL requires the transcriptional transactivation activity of ENL and the DNA binding motifs of HRX. *Mol.Cell Biol.* 1998; **18**(1): 122-129.
232. Zhang JG, Goldman JM, Cross NC. Characterization of genomic BCR-ABL breakpoints in chronic myeloid leukaemia by PCR. *Br.J Haematol.* 1995; **90**(1): 138-146.
233. Kantarjian HM, Walters RS, Keating MJ, Smith TL, O'Brien S, Estey EH *et al.* Results of the vincristine, doxorubicin, and dexamethasone regimen in adults with

- standard- and high-risk acute lymphocytic leukemia. *J Clin Oncol* 1990; **8**(6): 994-1004.
234. Secker-Walker LM, Cooke HM, Browett PJ, Shippey CA, Norton JD, Coustan-Smith E *et al.* Variable Philadelphia breakpoints and potential lineage restriction of bcr rearrangement in acute lymphoblastic leukemia. *Blood* 1988; **72**(2): 784-791.
 235. Chan LC, Karhi KK, Rayter SI, Heisterkamp N, Eridani S, Powles R *et al.* A novel abl protein expressed in Philadelphia chromosome positive acute lymphoblastic leukaemia. *Nature* 1987; **325**(6105): 635-637.
 236. Fialkow PJ, Denman AM, Jacobson RJ, Lowenthal MN. Chronic myelocytic leukemia. Origin of some lymphocytes from leukemic stem cells. *J Clin Invest* 1978; **62**(4): 815-823.
 237. Hirsch-Ginsberg C, Childs C, Chang KS, Beran M, Cork A, Reuben J *et al.* Phenotypic and molecular heterogeneity in Philadelphia chromosome-positive acute leukemia. *Blood* 1988; **71**(1): 186-195.
 238. Kurzrock R, Gutterman JU, Talpaz M. The molecular genetics of Philadelphia chromosome-positive leukemias. *N Engl J Med* 1988; **319**(15): 990-998.
 239. Cobaleda C, Gutierrez-Cianca N, Perez-Losada J, Flores T, Garcia-Sanz R, Gonzalez M *et al.* A primitive hematopoietic cell is the target for the leukemic transformation in human philadelphia-positive acute lymphoblastic leukemia. *Blood* 2000; **95**(3): 1007-1013.
 240. Schenk TM, Keyhani A, Bottcher S, Kliche KO, Goodacre A, Guo JQ *et al.* Multilineage involvement of Philadelphia chromosome positive acute lymphoblastic leukemia. *Leukemia* 1998; **12**(5): 666-674.
 241. Bacher U, Haferlach T, Hiddemann W, Schnittger S, Kern W, Schoch C. Additional clonal abnormalities in Philadelphia-positive ALL and CML

demonstrate a different cytogenetic pattern at diagnosis and follow different pathways at progression. *Cancer Genet Cytogenet.* 2005; **157**(1): 53-61.

242. Turhan AG, Eaves CJ, Kalousek DK, Eaves AC, Humphries RK. Molecular analysis of clonality and bcr rearrangements in Philadelphia chromosome-positive acute lymphoblastic leukemia. *Blood* 1988; **71**(5): 1495-1498.
243. Hotfilder M, Rottgers S, Rosemann A, Schrauder A, Schrappe M, Pieters R *et al.* Leukemic stem cells in childhood high-risk ALL/t(9;22) and t(4;11) are present in primitive lymphoid-restricted CD34+. *Cancer Res* 2005; **65**(4): 1442-1449.
244. Castor A, Nilsson L, Astrand-Grundstrom I, Buitenhuis M, Ramirez C, Anderson K *et al.* Distinct patterns of hematopoietic stem cell involvement in acute lymphoblastic leukemia. *Nat Med* 2005; **11**(6): 630-7.
245. Li S, Ilaria RL, Jr., Million RP, Daley GQ, Van Etten RA. The P190, P210, and P230 forms of the BCR/ABL oncogene induce a similar chronic myeloid leukemia-like syndrome in mice but have different lymphoid leukemogenic activity. *J Exp.Med* 1999; **189**(9): 1399-1412.
246. McLaughlin J, Chianese E, Witte ON. Alternative forms of the BCR-ABL oncogene have quantitatively different potencies for stimulation of immature lymphoid cells. *Mol.Cell Biol.* 1989; **9**(5): 1866-1874.
247. Gillert E, Leis T, Repp R, Reichel M, Hosch A, Breitenlohner I *et al.* A DNA damage repair mechanism is involved in the origin of chromosomal translocations t(4;11) in primary leukemic cells. *Oncogene* 1999; **18**(33): 4663-4671.
248. Haluska FG, Tsujimoto Y, Croce CM. The t(8;14) chromosome translocation of the Burkitt lymphoma cell line Daudi occurred during immunoglobulin gene rearrangement and involved the heavy chain diversity region. *Proc.Natl.Acad.Sci.U.S.A* 1987; **84**(19): 6835-6839.

249. Engler P, Weng A, Storb U. Influence of CpG methylation and target spacing on V(D)J recombination in a transgenic substrate. *Mol. Cell Biol.* 1993; **13**(1): 571-577.
250. Hsieh CL, Lieber MR. CpG methylated minichromosomes become inaccessible for V(D)J recombination after undergoing replication. *EMBO J* 1992; **11**(1): 315-325.
251. Storb U, Arp B. Methylation patterns of immunoglobulin genes in lymphoid cells: correlation of expression and differentiation with undermethylation. *Proc. Natl. Acad. Sci. U.S.A* 1983; **80**(21): 6642-6646.
252. Gauss GH, Domain I, Hsieh CL, Lieber MR. V(D)J recombination activity in human hematopoietic cells: correlation with developmental stage and genome stability. *Eur J Immunol* 1998; **28**(1): 351-8.
253. Shteper PJ, Siegfried Z, Asimakopoulos FA, Palumbo GA, Rachmilewitz EA, Ben-Neriah Y *et al.* ABL1 methylation in Ph-positive ALL is exclusively associated with the P210 form of BCR-ABL. *Leukemia* 2001; **15**(4): 575-582.
254. Raghavan SC, Swanson PC, Wu X, Hsieh CL, Lieber MR. A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. *Nature* 2004; **428**(6978): 88-93.
255. Cross NC, Melo JV, Feng L, Goldman JM. An optimized multiplex polymerase chain reaction (PCR) for detection of BCR-ABL fusion mRNAs in haematological disorders. *Leukemia* 1994; **8**(1): 186-189.
256. Willis TG, Jadayel DM, Coignet LJ, Abdul-Rauf M, Treleaven JG, Catovsky D *et al.* Rapid molecular cloning of rearrangements of the IGHJ locus using long-distance inverse polymerase chain reaction. *Blood* 1997; **90**(6): 2456-64.
257. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* 2002; **30**(12): e57.

258. Gellert M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* 2002; **71**: 101-32.
259. Wiemels JL, Leonard BC, Wang Y, Segal MR, Hunger SP, Smith MT *et al.* Site-specific translocation and evidence of postnatal origin of the t(1;19) E2A-PBX1 fusion in childhood acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* 2002; **99**(23): 15101-6.
260. Myers S, Freeman C, Auton A, Donnelly P, McVean G. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat Genet* 2008.
261. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* 2005; **310**(5746): 321-4.
262. Hesse JE, Lieber MR, Gellert M, Mizuuchi K. Extrachromosomal DNA substrates in pre-B cells undergo inversion or deletion at immunoglobulin V-(D)-J joining signals. *Cell* 1987; **49**(6): 775-783.
263. Raghavan SC, Kirsch IR, Lieber MR. Analysis of the V(D)J recombination efficiency at lymphoid chromosomal translocation breakpoints. *J Biol.Chem* 2001; **276**(31): 29126-29133.
264. Lewis SM, Agard E, Suh S, Czyzyk L. Cryptic signals and the fidelity of V(D)J joining. *Mol Cell Biol* 1997; **17**(6): 3125-36.
265. Hesse JE, Lieber MR, Mizuuchi K, Gellert M. V(D)J recombination: a functional definition of the joining signals. *Genes Dev* 1989; **3**(7): 1053-61.
266. Akamatsu Y, Tsurushita N, Nagawa F, Matsuoka M, Okazaki K, Imai M *et al.* Essential residues in V(D)J recombination signals. *J Immunol* 1994; **153**(10): 4520-9.

267. Raghavan SC, Swanson PC, Ma Y, Lieber MR. Double-strand break formation by the RAG complex at the bcl-2 major breakpoint region and at other non-B DNA structures in vitro. *Mol Cell Biol* 2005; **25**(14): 5904-19.
268. Jeffs AR, Wells E, Morris CM. Nonrandom distribution of interspersed repeat elements in the BCR and ABL1 genes and its relation to breakpoint cluster regions. *Genes Chromosomes Cancer* 2001; **32**(2): 144-154.
269. Denny CT, Shah NP, Ogden S, Willman C, McConnell T, Crist W *et al.* Localization of preferential sites of rearrangement within the BCR gene in Philadelphia chromosome-positive acute lymphoblastic leukemia. *Proc.Natl.Acad.Sci.U.S.A* 1989; **86**(11): 4254-4258.
270. Jager U, Bocskor S, Le T, Mitterbauer G, Bolz I, Chott A *et al.* Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. *Blood* 2000; **95**(11): 3520-9.
271. Melo JV, Gordon DE, Tuszynski A, Dhut S, Young BD, Goldman JM. Expression of the ABL-BCR fusion gene in Philadelphia-positive acute lymphoblastic leukemia. *Blood* 1993; **81**(10): 2488-91.
272. Papadopoulos PC, Greenstein AM, Gaffney RA, Westbrook CA, Wiedemann LM. Characterization of the translocation breakpoint sequences in Philadelphia-positive acute lymphoblastic leukemia. *Genes Chromosomes Cancer* 1990; **1**(3): 233-9.
273. Gauss GH, Lieber MR. Mechanistic constraints on diversity in human V(D)J recombination. *Mol Cell Biol* 1996; **16**(1): 258-69.
274. Litz CE, McClure JS, Copenhaver CM, Brunning RD. Duplication of small segments within the major breakpoint cluster region in chronic myelogenous leukemia. *Blood* 1993; **81**(6): 1567-72.

275. Reid AG, Huntly BJ, Hennig E, Niederwieser D, Campbell LJ, Bown N *et al.* Deletions of the derivative chromosome 9 do not account for the poor prognosis associated with Philadelphia-positive acute lymphoblastic leukemia. *Blood* 2002; **99**(6): 2274-5.
276. Specchia G, Albano F, Anelli L, Storlazzi CT, Zagaria A, Mancini M *et al.* Deletions on der(9) chromosome in adult Ph-positive acute lymphoblastic leukemia occur with a frequency similar to that observed in chronic myeloid leukemia. *Leukemia* 2003; **17**(3): 528-31.
277. Agrawal A, Eastman QM, Schatz DG. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 1998; **394**(6695): 744-51.
278. Hiom K, Melek M, Gellert M. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* 1998; **94**(4): 463-70.
279. Mills KI, Sproul AM, Leibowitz D, Burnett AK. Mapping of breakpoints, and relationship to BCR-ABL RNA expression, in Philadelphia-chromosome-positive chronic myeloid leukaemia patients with a breakpoint around exon 14 (b3) of the BCR gene. *Leukemia* 1991; **5**(11): 937-41.
280. Chen SJ, Chen Z, d'Auriol L, Le Coniat M, Grausz D, Berger R. Ph1+bcr- acute leukemias: implication of Alu sequences in a chromosomal translocation occurring in the new cluster region within the BCR gene. *Oncogene* 1989; **4**(2): 195-202.
281. Lieber MR, Hesse JE, Mizuuchi K, Gellert M. Lymphoid V(D)J recombination: nucleotide insertion at signal joints as well as coding joints. *Proc Natl Acad Sci U S A* 1988; **85**(22): 8588-92.
282. Gauss GH, Lieber MR. Unequal signal and coding joint formation in human V(D)J recombination. *Mol Cell Biol* 1993; **13**(7): 3900-6.

283. Rubin H, Yao A, Chow M. Heritable, population-wide damage to cells as the driving force of neoplastic transformation. *Proc Natl Acad Sci U S A* 1995; **92**(11): 4843-7.
284. Messier TL, O'Neill JP, Hou SM, Nicklas JA, Finette BA. In vivo transposition mediated by V(D)J recombinase in human T lymphocytes. *EMBO J* 2003; **22**(6): 1381-8.
285. Chatterji M, Tsai CL, Schatz DG. Mobilization of RAG-generated signal ends by transposition and insertion in vivo. *Mol Cell Biol* 2006; **26**(4): 1558-68.
286. Reddy YV, Perkins EJ, Ramsden DA. Genomic instability due to V(D)J recombination-associated transposition. *Genes Dev* 2006; **20**(12): 1575-82.
287. Romero P, Colombatti A. The first intron of the BCR gene contains two minor alternative exons. *Leuk Res* 1995; **19**(12): 963-70.
288. Chung KF, Hew M, Score J, Jones AV, Reiter A, Cross NC *et al.* Cough and hypereosinophilia due to FIP1L1-PDGFR α fusion gene with tyrosine kinase activity. *Eur. Respir. J* 2006; **27**(1): 230-232.
289. Pardanani A, Brockman SR, Paternoster SF, Flynn HC, Ketterling RP, Lasho TL *et al.* FIP1L1-PDGFR α fusion: prevalence and clinicopathologic correlates in 89 consecutive patients with moderate to severe eosinophilia. *Blood* 2004; **104**(10): 3038-45.
290. La SR, Specchia G, Cuneo A, Beacci D, Nozzoli C, Luciano L *et al.* The hypereosinophilic syndrome: fluorescence in situ hybridization detects the del(4)(q12)-FIP1L1/PDGFR α but not genomic rearrangements of other tyrosine kinases. *Haematologica* 2005; **90**(5): 596-601.
291. Gotlib J, Cools J, Malone JM, III, Schrier SL, Gilliland DG, Coutre SE. The FIP1L1-PDGFR α fusion tyrosine kinase in hypereosinophilic syndrome and

chronic eosinophilic leukemia: implications for diagnosis, classification, and management. *Blood* 2004; **103**(8): 2879-2891.

292. Score J, Curtis C, Waghorn K, Stalder M, Jotterand M, Grand FH *et al.* Identification of a novel imatinib responsive KIF5B-PDGFR α fusion gene following screening for PDGFR α overexpression in patients with hypereosinophilia. *Leukemia* 2006; **20**(5): 827-832.
293. Klion AD, Noel P, Akin C, Law MA, Gilliland DG, Cools J *et al.* Elevated serum tryptase levels identify a subset of patients with a myeloproliferative variant of idiopathic hypereosinophilic syndrome associated with tissue fibrosis, poor prognosis, and imatinib responsiveness. *Blood* 2003; **101**(12): 4660-4666.
294. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻($\Delta\Delta C_T$) Method. *Methods* 2001; **25**(4): 402-408.
295. Gabert J, Beillard E, van dVW, Bi W, Grimwade D, Pallisgaard N *et al.* Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia - a Europe Against Cancer program. *Leukemia* 2003; **17**(12): 2318-2357.
296. Pongers-Willemse MJ, Verhagen OJ, Tibbe GJ, Wijkhuijs AJ, de HV, Roovers E *et al.* Real-time quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia using junctional region specific TaqMan probes. *Leukemia* 1998; **12**(12): 2006-2014.
297. Jeffreys AJ, Wilson V, Neumann R, Keyte J. Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Res* 1988; **16**(23): 10953-10971.

298. Brisco MJ, Condon J, Sykes PJ, Neoh SH, Morley AA. Detection and quantitation of neoplastic cells in acute lymphoblastic leukaemia, by use of the polymerase chain reaction. *Br.J Haematol.* 1991; **79**(2): 211-217.
299. Hughes T, Deininger M, Hochhaus A, Branford S, Radich J, Kaeda J *et al.* Monitoring CML patients responding to treatment with tyrosine kinase inhibitors: review and recommendations for harmonizing current methodology for detecting BCR-ABL transcripts and kinase domain mutations and for expressing results. *Blood* 2006; **108**(1): 28-37.
300. Costa JM, Ernault P, Olivi M, Gaillon T, Arar K. Chimeric LNA/DNA probes as a detection system for real-time PCR. *Clin.Biochem.* 2004; **37**(10): 930-932.
301. Stover EH, Chen J, Folens C, Lee BH, Mentens N, Marynen P *et al.* Activation of FIP1L1-PDGFRalpha requires disruption of the juxtamembrane domain of PDGFRalpha and is FIP1L1-independent. *Proc Natl Acad Sci U S A* 2006; **103**(21): 8078-83.
302. Mosselman S, Claesson-Welsh L, Kamphuis JS, van Zoelen EJ. Developmentally regulated expression of two novel platelet-derived growth factor alpha-receptor transcripts in human teratocarcinoma cells. *Cancer Res* 1994; **54**(1): 220-5.
303. Metzgeroth G, Walz C, Erben P, Schmitt-Graeff A, Haferlach C, Grimwade D *et al.* Safety and efficacy of imatinib in chronic eosinophilic leukemia and hypereosinophilic syndrome - a phase-II study. *Br.J Haematol.* 2008.
304. Klion AD, Robyn J, Maric I, Fu W, Schmid L, Lemery S *et al.* Relapse following discontinuation of imatinib mesylate therapy for FIP1L1/PDGFRalpha-positive chronic eosinophilic leukemia: implications for optimal dosing. *Blood* 2007; **110**(10): 3552-3556.
305. von Bubnoff N, Sandherr M, Schlimok G, Andreesen R, Peschel C, Duyster J. Myeloid blast crisis evolving during imatinib treatment of an FIP1L1-PDGFR

alpha-positive chronic myeloproliferative disease with prominent eosinophilia. *Leukemia* 2005; **19**(2): 286-7.

306. Heinrich MC, Corless CL, Blanke CD, Demetri GD, Joensuu H, Roberts PJ *et al.* Molecular correlates of imatinib resistance in gastrointestinal stromal tumors. *J.Clin.Oncol.* 2006; **24**(29): 4764-4774.
307. Shah NP, Lee FY, Luo R, Jiang Y, Donker M, Akin C. Dasatinib (BMS-354825) inhibits KITD816V, an imatinib-resistant activating mutation that triggers neoplastic growth in most patients with systemic mastocytosis. *Blood* 2006; **108**(1): 286-91.
308. Kamal A, menar-Queralt A, LeBlanc JF, Roberts EA, Goldstein LS. Kinesin-mediated axonal transport of a membrane compartment containing beta-secretase and presenilin-1 requires APP. *Nature* 2001; **414**(6864): 643-648.
309. Tanaka Y, Kanai Y, Okada Y, Nonaka S, Takeda S, Harada A *et al.* Targeted disruption of mouse conventional kinesin heavy chain, kif5B, results in abnormal perinuclear clustering of mitochondria. *Cell* 1998; **93**(7): 1147-1158.
310. Safley AM, Sebastian S, Collins TS, Tirado CA, Stenzel TT, Gong JZ *et al.* Molecular and cytogenetic characterization of a novel translocation t(4;22) involving the breakpoint cluster region and platelet-derived growth factor receptor-alpha genes in a patient with atypical chronic myeloid leukemia. *Genes Chromosomes Cancer* 2004; **40**(1): 44-50.
311. Mitelman F. Recurrent chromosome aberrations in cancer. *Mutat Res* 2000; **462**(2-3): 247-53.
312. Pierce AJ, Stark JM, Araujo FD, Moynahan ME, Berwick M, Jasin M. Double-strand breaks and tumorigenesis. *Trends Cell Biol* 2001; **11**(11): S52-9.
313. van Gent DC, Hoeijmakers JH, Kanaar R. Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet* 2001; **2**(3): 196-206.

314. Weinstock DM, Richardson CA, Elliott B, Jasin M. Modeling oncogenic translocations: distinct roles for double-strand break repair pathways in translocation formation in mammalian cells. *DNA Repair (Amst)* 2006; **5**(9-10): 1065-74.
315. Bacolla A, Wojciechowska M, Kosmider B, Larson JE, Wells RD. The involvement of non-B DNA structures in gross chromosomal rearrangements. *DNA Repair (Amst)* 2006; **5**(9-10): 1161-70.
316. Harrison CJ. Acute lymphoblastic leukaemia. *Best Pract Res Clin Haematol* 2001; **14**(3): 593-607.
317. Ross DM, Hughes TP. Current and emerging tests for the laboratory monitoring of chronic myeloid leukaemia and related disorders. *Pathology* 2008; **40**(3): 231-46.
318. Radich J, Gehly G, Lee A, Avery R, Bryant E, Edmands S *et al.* Detection of bcr-abl transcripts in Philadelphia chromosome-positive acute lymphoblastic leukemia after marrow transplantation. *Blood* 1997; **89**(7): 2602-9.
319. Lu CP, Posey JE, Roth DB. Understanding how the V(D)J recombinase catalyzes transesterification: distinctions between DNA cleavage and transposition. *Nucleic Acids Res* 2008; **36**(9): 2864-73.
320. Huye LE, Purugganan MM, Jiang MM, Roth DB. Mutational analysis of all conserved basic amino acids in RAG-1 reveals catalytic, step arrest, and joining-deficient mutants in the V(D)J recombinase. *Mol Cell Biol* 2002; **22**(10): 3460-73.
321. Kriatchko AN, Anderson DK, Swanson PC. Identification and characterization of a gain-of-function RAG-1 mutant. *Mol Cell Biol* 2006; **26**(12): 4712-28.

Project: Analysis of clustering and sequence feature association with breakpoints in BCR-ABL translocations
PI: Nick Cross, Joe Wiemels
Analysis: Ru-Fang Yeh <rufang@biostat.ucsf.edu>

I. Data: Three sets of breakpoint data involving BCR/ABL:

	p190	p210 ALL	p210 CML	Total
# subjects	25	25	50	100

II. Analysis Aims:

- Assess breakpoint clustering patterns and identify 'hotspots' if any.
- Search for sequence feature associated with breakpoints.
- Compare the distribution/motif features
 - in ABL between ALL (p190 and p210) vs CML
 - in ABL between p190 vs p210 (ALL+CML)
 - in ABL between p190 ALL vs p210 ALL
 - in ABL between p190 ALL vs p210 CML
 - in BCR between p210 ALL and p210 CML

III. Results:

1. Compare the breakpoint distributions of various subgroups.

[No differences in all comparisons of the subgroup breakpoint distributions.](#)

Table 1 Komolgorov-Smirnov test p-values for the comparisons of breakpoint distribution in different subtypes of leukemia.

Comparison	Komolgorov-Smirnov test p-value
ABL.ALL.vs.CML	0.967
ABL.p190.vs.p210	0.6
ABL.p190ALL.vs.p210ALL	0.915
ABL.p190ALL.vs.p210CML	0.637
BCR.p210ALL.vs.p210CML	0.915

2. Test for existence of clusters using scan statistics, using bandwidth determined by various selection methods. The p-values of scan statistics were calculated as described in Segal & Wiemels 2002 using the endpoint correction method (Loader 1991).

Table 2 The p-value of the scan statistics for all breakpoints and within each subtypes using corresponding bandwidth as determined by different kernel smoothing bandwidth selection methods: nrd0: Silverman's rule-of-thumb for choosing the bandwidth of a Gaussian kernel density estimator (Silverman 1986), $0.9 * \min(sd, IQR/(1.34 * n^{(-1/5)}))$; nrd: Scotts (1992) variation of the rule-of-thumb with factor 1.06; ucv: unbiased cross-validation; bcv: biased cross-validation; ste - the solve-the-

equation method in Sheather & Jones (1991); dpi - the direct plug-in method in Sheather & Jones (1991).

BCR breakpoints					
	p190	p210ALL	p210CML	p210	
nrd0	4996.9, p=1	359.8, p=0.053	315.4, p=0.27	301.2, p=0.015	
nrd	5885.3, p=1	423.7, p=0.043	371.4, p=0.14	354.8, p=0.0049	
ucv	6329.8, p=1	151.6, p=0.2	283.7, p=0.14	238, p=0.079	
bcv	6329.6, p=1	455.5, p=0.069	417.2, p=0.13	380.7, p=0.012	
SJste	5968.8, p=1	170.2, p=0.07	258.9, p=0.51	219.5, p=0.11	
SJdpi	5997.7, p=1	294.4, p=0.061	283.5, p=0.14	247, p=0.041	

ABL breakpoints					
	p190	p210ALL	p210CML	p210	All breaks
nrd0	21059.5, p=0.067	19385.4, p=1	17167.2, p=1	15656.2, p=1	15006.8, p=0.16
nrd	24803.3, p=0.17	22831.7, p=1	20219.2, p=1	18439.6, p=0.7	17674.7, p=0.19
ucv	2776.8, p=0.0019	16646.3, p=1	21710.1, p=1	12074.3, p=0.87	8816.3, p=0.1
bcv	26662.2, p=0.24	24542.8, p=1	21677.3, p=1	19821.5, p=0.98	19002, p=0.19
SJste	11803.7, p=0.041	13808, p=1	17956.9, p=1	12084.9, p=0.87	10395.8, p=0.067
SJdpi	16217.5, p=0.054	16777.3, p=1	18992.3, p=1	14189.8, p=0.97	12033.7, p=0.1

3. Number of cluster.

From the Silverman's method and gap statistics assessment, there is one or less cluster for all the subgroups of BCR breakpoints, and 2 clusters for the subgroups of ABL breakpoints.

Table 3 Assessing number of modes using Silverman's smoothed bootstrap method as described in Segal and Wiemels (2002). bw: critical bandwidth.

BCR breakpoints									
nModes	p190		p210ALL		p210CML		p210		
	bw	p-value	bw	p-value	bw	p-value	bw	p-value	
1	4700	0.4775	362	0.396	340	0.4015	293	0.488	
2	3030	0.46	302	0.0765	205	0.575	167	0.7955	
3	3000	0.0735	146	0.42	192	0.194	153	0.4275	

ABL breakpoints									
nModes	p190		p210ALL		p210CML		p210		all breakpoints
	bw	p-value	bw	p-value	bw	p-value	bw	p-value	bw p-value
1	29620	0.0545	28530	0.043	22720	0.103	25130	0.014	26060 0.012
2	11650	0.443	8790	0.725	9030	0.801	8950	0.594	6950 0.8365
3	6660	0.6795	5540	0.873	7540	0.676	5710	0.904	5840 0.754

Figure 1 Assessing number of clusters using gap statistics with k-means clustering algorithm.

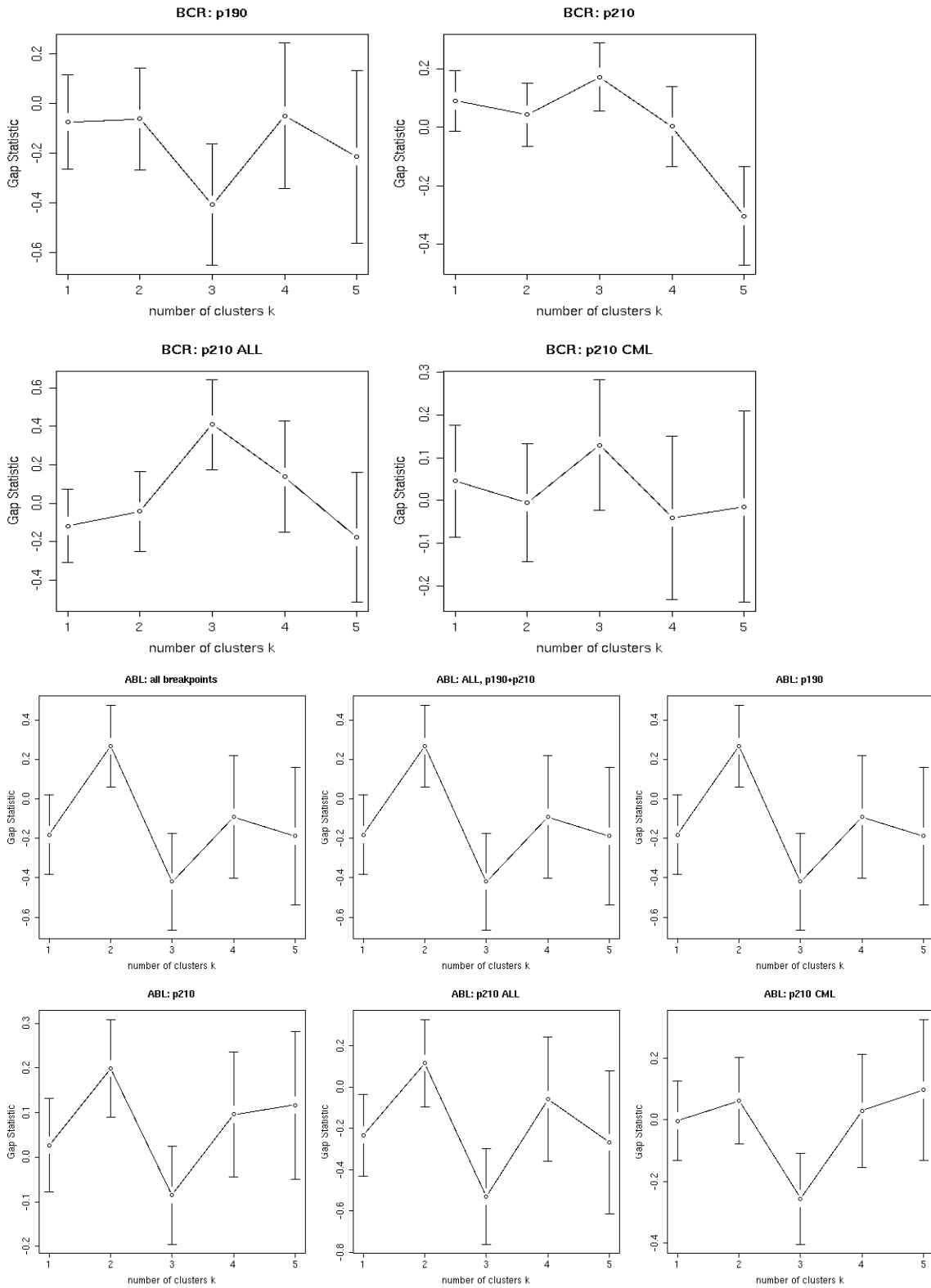


Figure 2 BCR breakpoints. (A) p190 breakpoints. (B) p210 breakpoints. Density curves were determined using the Sheather-Jones solve-the-equation bandwidth. Color boxes beneath the breakpoints indicate the locations of different classes of repetitive elements as determined by RepeatMasker, and known motif matches (exact match) using fuzznuc or PWM.

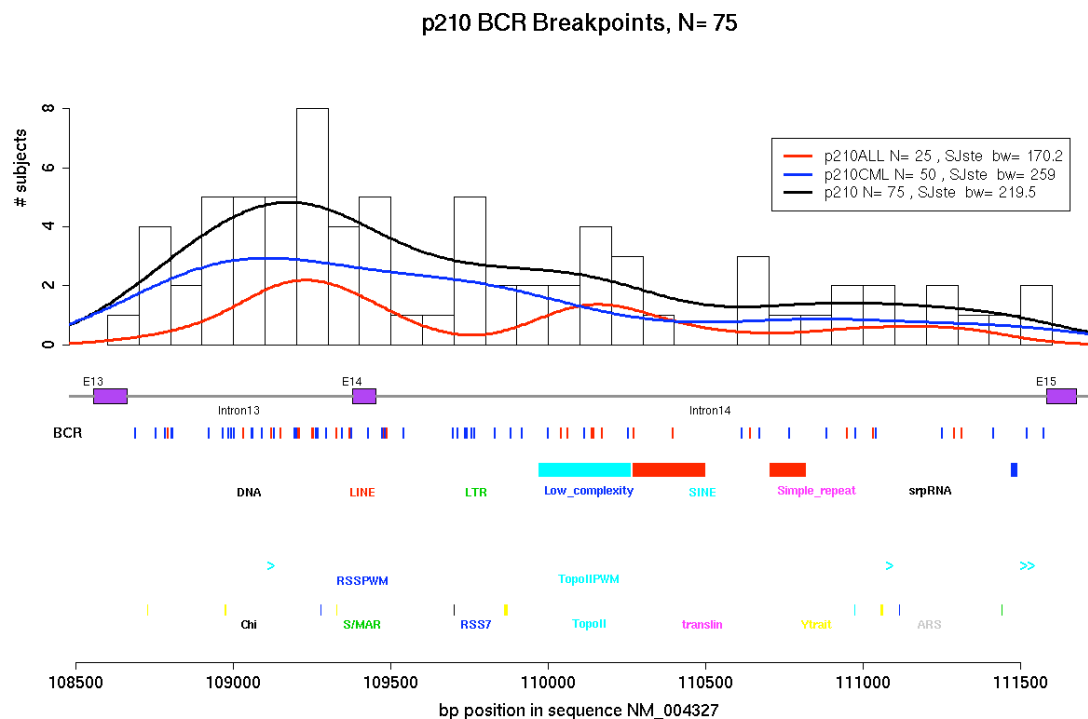
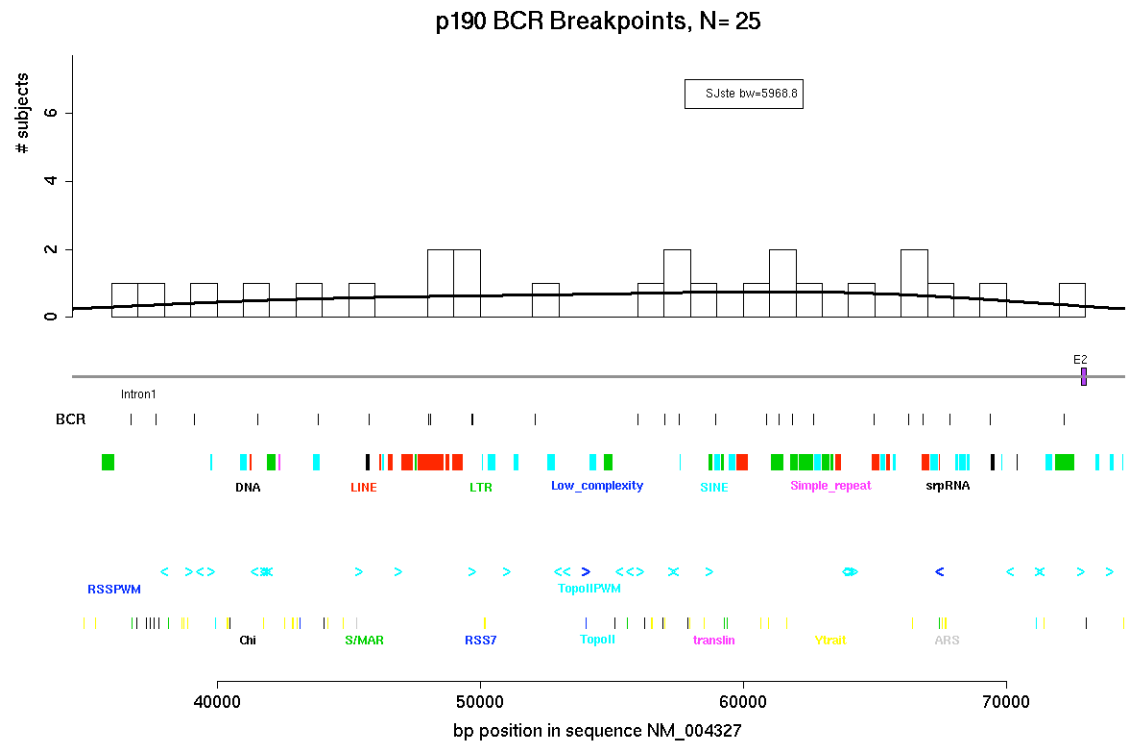


Figure 3 ABL breakpoints. Density curves were determined using the Sheather-Jones solve-the-equation bandwidth. Color boxes beneath the breakpoints indicate the locations of different classes of repetitive elements as determined by RepeatMasker, and known motif matches (exact match) using fuzznuc or PWM.

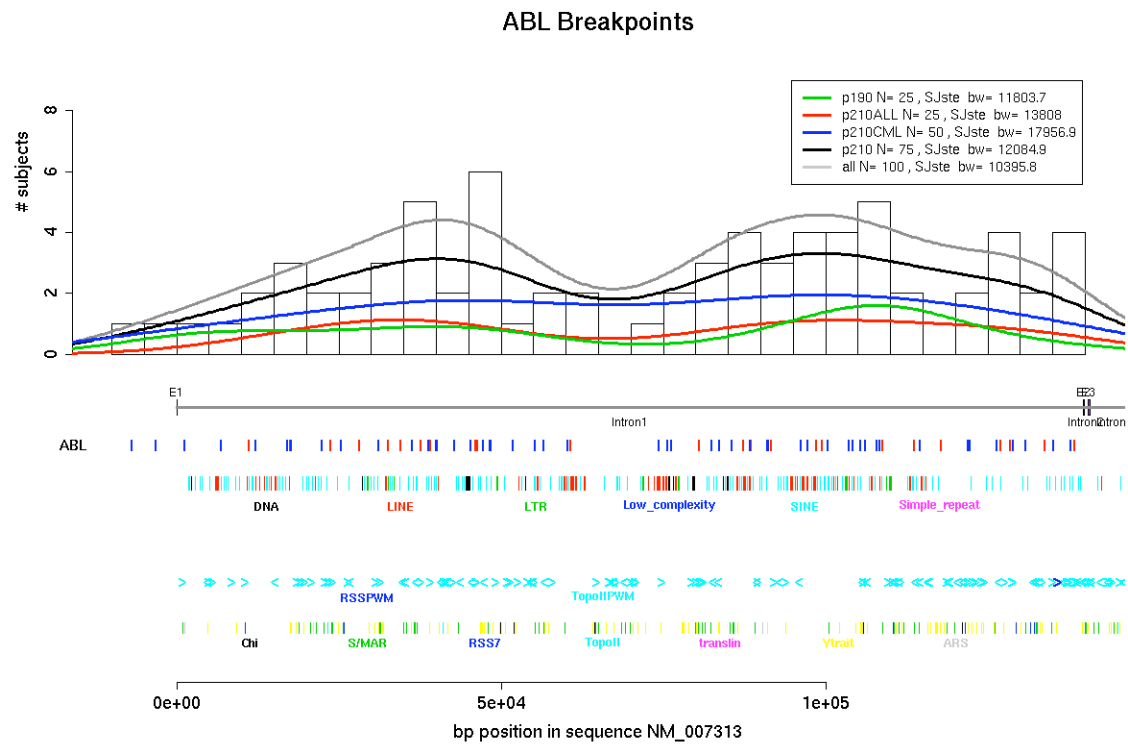
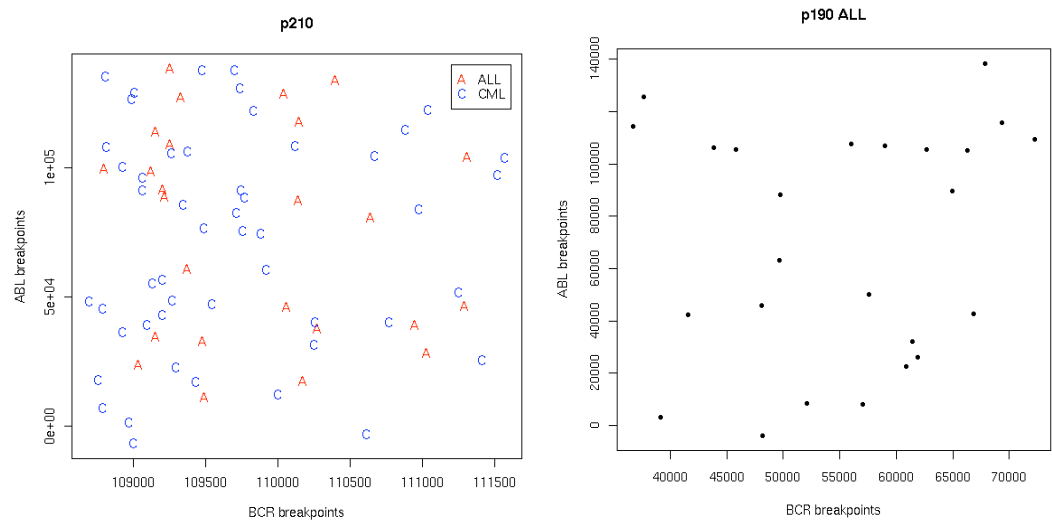


Figure 4 BCR vs ABL breakpoints. (a) p210; (b) p190 ALL. No apparent association.

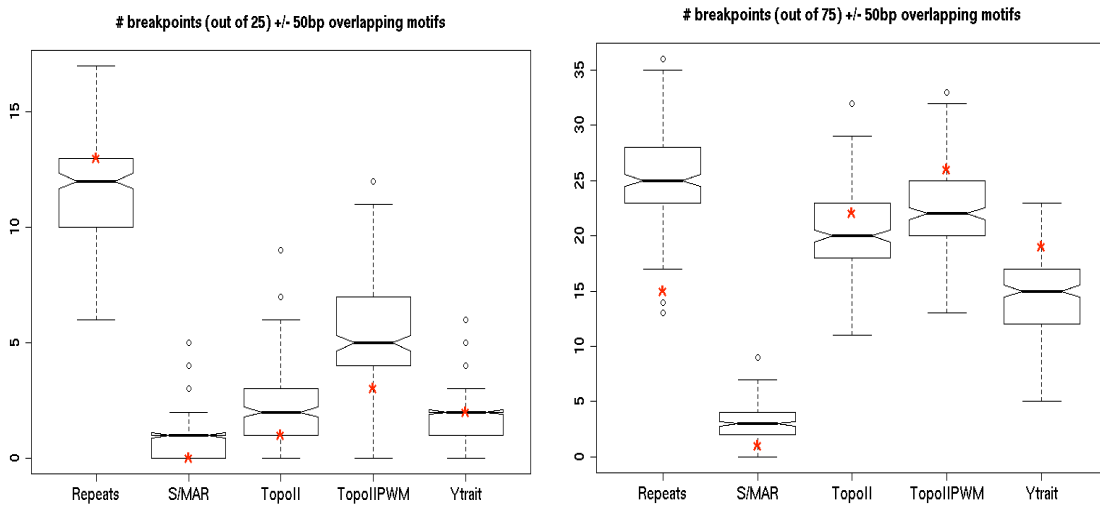


4. Search for sequence features/motifs associated with breakpoints.

(i) Map **known motifs** (listed in Table 3) to the sequence and test for enrichment of each known motif in the proximity of breakpoints using simulations. Matches were plotted in Figure 2 and 3. And Figure 5 compares % of BCR breakpoints overlapping motifs in their proximity (within 50bp or 100bp of either side of breakpoints), compared to the motif co-localization of simulated, randomly distributed breakpoints. No apparent enrichment found.

Figure 5. % of breakpoints associated with motifs within 50bp of either side of breakpoints). The red points are the observed from the breakpoint data, and the box plot gives the results based on 200 simulations. Each simulation draws N breakpoints uniformly from the sequence region and calculate % of N breakpoints overlapping the given type of motif within (-50, +50). No apparent motif enrichment found. Repeats: repetitive elements identified by RepeatMasker; S/MAR: regular expression matches of the motifs; TopoII: regular expression matches of the TopoII consensus; TopoIIPWM: sites identified by PWM scoring; Ytrait: regular expression matches of the Y trait.

(a) p190 (b) p210 (c) p210ALL (d) p210CML



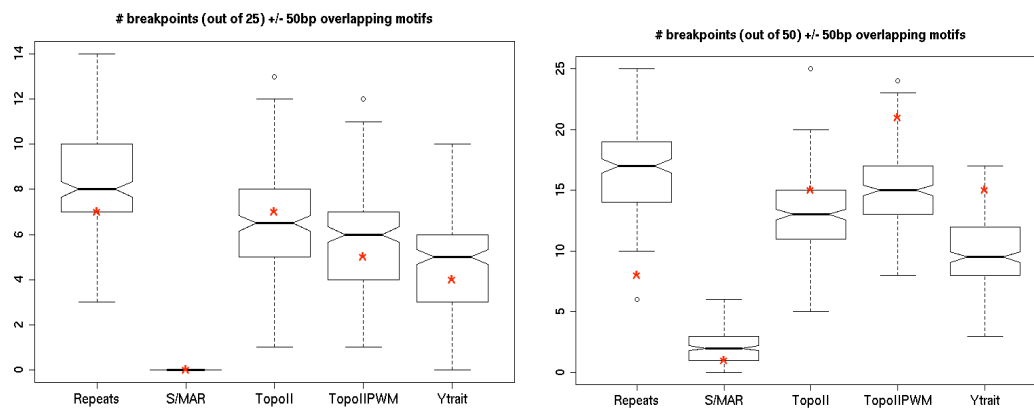


Figure 6. % of breakpoints associated with motifs within 50bp of either side of breakpoints) in ABL. (a) p190 (b) p210 (c) p210ALL (d) p210CML

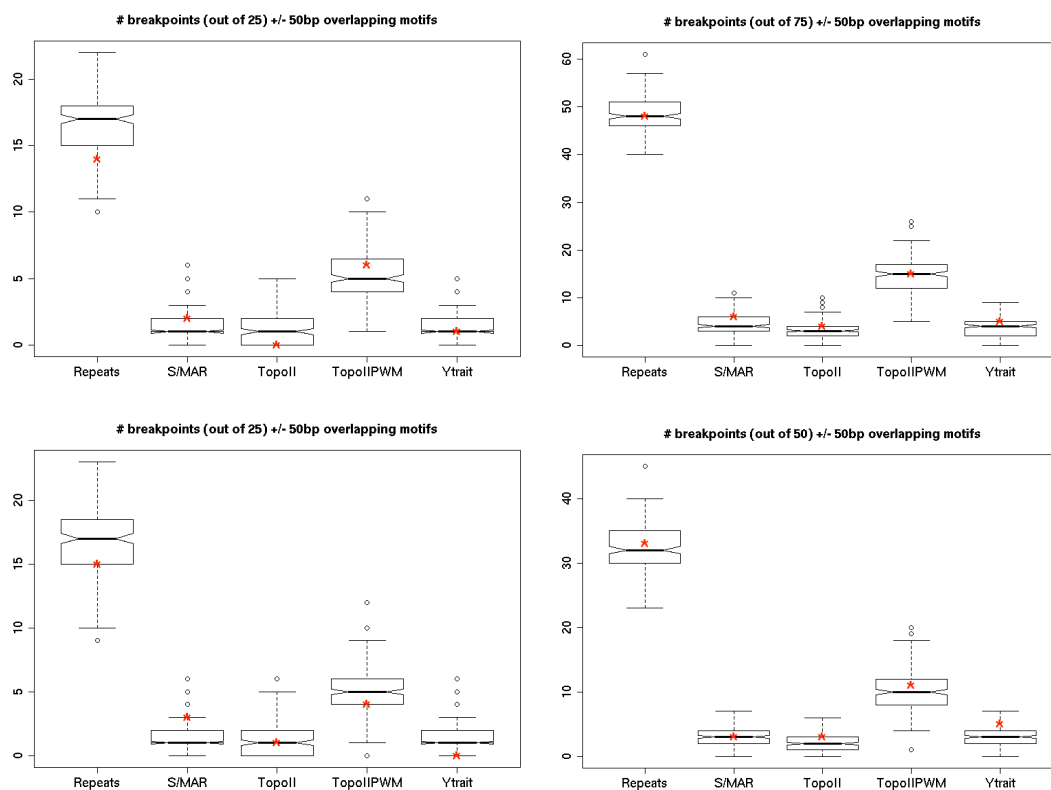
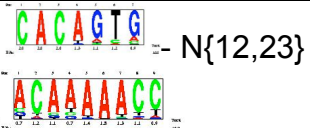



Table 4 Known motifs searched in the BCR and ABL breakpoint regions. fuzznuc: use EMBOSS/fuzznuc to search for regular expression pattern match of the motif; PWM: use custom perl script and the position-specific weight matrix to find high-scoring matches.

What	Motif	Search Method	References
Repetitive elements (SINE, LINE, LTR, transposon fossils, simple repeats...)		Repeat Masker http://www.repeatmasker.org	Smit 1996.
V(D)J RSS (Recombination Site Sequence)	 - N{12,23}	Fuzznuc, PWM	Lewis et al 1997
Topoisomerase II binding site	RNYNNCNNGYNGKTNINY; GTNWAYAYYNATNNG 	Fuzznuc, PWM	Spitzner & Muller 1988
S/MAR (Scaffold/matrix attachment regions)	WADAWAYAWW; TWWTDTTWWW; AATATTTTT	Fuzznuc, SMARTest [#]	Gasser et al 1989; Liebich et al 2002
Translin binding site	Major: GCNCWSSW-N ₀₋₂ - GCCCWSSW Minor: MTGCAG- N ₀₋₄ - GCCCWSSW	Fuzznuc	Aoki et al 1995
Chi-like sequence	GCWGGWGG	Fuzznuc	Krowczynska et al 1990
Purl binding site	GGNNGAGGGAGARRRR	Fuzznuc	Bergemann & Johnson 1992
Eukaryotes replication origin sequence	WAWTTDDWWWDHWGW HMAWTT	Fuzznuc	Dobbs et al 1994
ARS- <i>S.cerevisiae</i>	WTTTAYRTTTW	Fuzznuc	Campbell & Newlon 1991; Stinchcomb et al 1981
Putative triple helices	(GA) ₆	Fuzznuc	Caddle et al 1990; Lyamichev et al 1986; Wells 1988
Pyrimidine tract	Y ₁₂	Fuzznuc	Yamaguchi et al















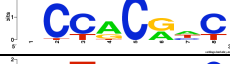



			1985; Suzuki et al 1993
Human minisatellite core sequence	GGGCAGGARG	Fuzznuc	Wahls et al 1990
Satellite III core	YMCNTACSYCCKYYTNN C	Fuzznuc	Kas & Laemmli 1992

- # **MAR/SAR** (matrix/scaffold attachment region): There are several program tools available to search for candidate S/MAR in a sequence, including
- **EMBOSS/marscan**, search for bipartite sequence element within 200bp. (AATAAYAA and AWWRTAANNWWGNNNC, van Drunen et al 1999)
<http://emboss.sourceforge.net/apps/release/4.0/emboss/apps/marscan.html>
 - Genomatix/**SMARTest** (Frisch et al 2002; Liebich et al 2002), use a proprietary library of currently 97 S/MAR-associated weight matrices to find good sequence matches and identify candidates by thresholding # bp covered by S/Mar matrices in a sliding window of 300bp.
<http://www.genomatix.de>
 - **MAR-finder**, utilizes the pattern-density of a number of known sequence features, including origin of replication motifs (ATTA, ATTTA and ATTTTA), TG-richness, curved DNA (AAAA_n7AAAA_n7AAAA as well as the motif TTATAA), kinked DNA (copies of dinucleotides TG, CA or TA separated by 2-4 or 9-12 bp), topoisomerase II binding site consensus sequence, AT-richness.
<http://www.futuresoft.org/MAR-Wiz/>
 - **ChrClass** (Glazko et al 2001),
<http://ftp.bionet.nsc.ru/pub/biology/chrclass/chrclass2.zip>

(ii) **De novo motif finding** to search for novel sequence motifs enriched in the proximity of breakpoints. Take (-50bp, +50bp) of each breakpoint, and run motif finder Bioprosppector (Liu et al 2003; <http://ai.stanford.edu/~xsliu/cgi-bin/BPsearch.cgi>) for enriched motif.

	Motif 1	Motif 2	Motif 3
BCR: p190			
BCR: p210			
BCR: p210ALL			
BCR: p210CML			

	Motif 1	Motif 2	Motif 3
--	---------	---------	---------

ABL: all			
ABL: p190			
ABL: p210			
ABL: p210ALL			
ABL: p210CML			
ABL: p190+p210 ALL			

IV. References:

van Drunen CM., Sewalt RGAB., Oosterling RW., Weisbeek PJ., Smeekens SCM. and van Driel R. "A bipartite sequence element associated with matrix/scaffold attachment regions" *Nucleic Acids Research*. 1999. Vol 27, No. 14, pp. 2924-2930

Glazko GV, Rogozin IB, Glazkov MV. Comparative study and prediction of DNA fragments associated with various elements of the nuclear matrix. *Biochim Biophys Acta*. 2001 Feb 16;1517(3):351-64.

Frisch, M., Frech, K., Klingenhoff, A., Cartharius, K., Liebich, I., Werner, T. (2002). A new tool for the in silico prediction of matrix attachment regions in large genomic sequences. *Genome Research* 12, 349-354.

Liebich, I., Bode, J., Frisch, M., Wingender, E. (2002). S/MARt DB: a database on scaffold/matrix attached regions. *Nucleic Acids Res.* 30, 372-374.