Abnormalities affecting tyrosine kinase signalling in atypical myeloproliferative disorders

by

Claire Hidalgo-Curtis

Thesis for the degree of Doctor of Philosophy

May 2009
UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH & LIFE SCIENCES
SCHOOL OF MEDICINE

Doctor of Philosophy

ABNORMALITIES AFFECTING TYROSINE KINASE SIGNALLING IN ATYPICAL MYELOPROLIFERATIVE DISORDERS

By Claire Hidalgo-Curtis

The myeloproliferative disorders (MPDs) are a group of haematopoietic stem cell diseases, characterised by proliferation of one or more cells of the myeloid lineage. Several lines of evidence have highlighted the importance of aberrant tyrosine kinase signalling in the pathogenesis of these disorders. Cloning of rare chromosomal translocations and point mutation analysis in the MPDs has identified diverse deregulated tyrosine kinase genes, notably PDGFRA, PDGFRB, FGFR1 and JAK2. However the majority of atypical MPDs still remain to be characterised and identification of patients harbouring fusions, particularly those involving the PDGF receptors is of increasing importance, as they are likely to be responsive to targeted therapy with imatinib.

I am investigating MPD patients for abnormalities affecting tyrosine kinase signalling, and have used three approaches, translocation cloning, expression analysis and SNP array analysis to detect regions of loss of heterozygosity (LOH). Thus far, by translocation cloning I have identified a previously undescribed partner gene fused to PDGFRB and two new PDGFRA fusion genes. I have also designed two reverse transcriptase PCR (RT-PCR) assays and a cDNA MLPA assay to detect over-expression of specific tyrosine kinases screening approximately 200 patients. Each assay identified all patients previously diagnosed with known fusions. Additionally, two patients identified with overexpression of PDGFRB have been found to have cryptic ETV6-PDGFRB fusions and overexpression of PDGFRA in one patient lead to the discovery of a previously undescribed fusion involving a novel partner gene (KIF5B).
Recent evidence has indicated that acquired isodisomy is a novel mechanism by which mutations in cancer may be reduced to homozygosity. Typically, acquired isodisomy is associated with oncogenic changes rather than tumour suppressor genes, eg. the activating \textit{JAK2 V617F} mutation and 9p aUPD. I have undertaken a screen using Affymetrix 50K SNP arrays for regions of acquired isodisomy as a means to identify genomic regions that may harbour novel oncogenes in different subgroups of MPD patients. Large tracts of homozygosity (defined as $>$20Mb running to a telomere), strongly suggesting acquired isodisomy, were seen in 40% aMPD patients. The homozygous tracts encompassed diverse genomic regions in aMPD, but two common regions (3 cases for each region) were identified at 7q and 11q. Mutations in the CBL ubiquitin ligase gene were discovered in all three aCML patients with 11q aUPD as well as in an additional 23 MPD patients following further screening.
CONTENTS

Contents........................................................................................................................................i
List of Figures...................................................................................................................................viii
List of Tables......................................................................................................................................xiii
Declaration of Authorship..................................................................................................................xv
Acknowledgments............................................................................................................................xvii
List of Abbreviations.........................................................................................................................xviii

1 INTRODUCTION 1
1.1 Normal Haematopoiesis 1
1.2 Leukaemia 3
  1.2.1 Clonality 4
  1.2.2 Classification 4
1.3 The chronic myeloproliferative disorders (CMPD and the myelodysplastic/myeloproliferative disorders (MDS/MPD) 5
1.4 The paradigm for deregulated tyrosine kinases –
  Chronic myelogenous leukaemia (CML) 7
  1.4.1 Molecular pathogenesis of CML 7
  1.4.2 Therapy in CML and the development of Imatinib 9
1.5 The tyrosine kinases 11
  1.5.1 Non-receptor tyrosine kinases (NTRKs) 12
  1.5.2 Receptor tyrosine kinases (RTKs) 13
1.6 Molecular pathogenesis and deregulated tyrosine kinases 15
1.7 PDGFRA and PDGFRB rearrangements in MPDs 17
  1.7.1 Fusions involving PDGFRA 17
  1.7.2 Chronic eosinophilic leukaemia (CEL) 18
  1.7.3 Fusions involving PDGFRB 19
  1.7.4 Chronic myelomonocytic leukaemia (CMML) 21
  1.7.5 JAK2 and the myeloproliferative disorders 22
1.8 Summary 26
1.9 Statement of aims 26

2 METHODS 28

2.1 Sample processing 28
2.1.1 Preparation of blood and bone marrow samples 28
2.1.2 Enriching cultures for T-cells 29
2.1.3 RNA extraction 30
2.1.4 cDNA synthesis 31
2.1.5 DNA extraction 31
2.1.5.1 DNA extraction from fixed cell suspension 31
2.1.5.2 Extraction from cell pellets 32
2.1.5.3 DNA extraction from Guanidinium thiocyanate 33
2.1.5.4 DNA extraction from mouthbrush samples 34
2.1.5.5 DNA extraction from haematological stained slides 34
2.1.5.6 GenomiPhi V2 DNA amplification kit 35

2.2 Amplification and analysis 35
2.2.1 Reverse transcriptase PCR (RT-PCR) 35
2.2.2 Microsatellite PCR 36
2.2.3 Control gene PCR 36
2.2.4 Powerplex® 16 system 37
2.2.5 Long-range PCR 38
2.2.6 Bubble-PCR 38
2.2.7 Gel Electrophoresis 39
2.2.8 Denaturing high-performance liquid chromatography (dHPLC) 40
2.2.9 Multiplex ligation-dependent probe amplification (MLPA) 41
2.2.10 Exo-SAP and sequencing 42
2.2.10.1 Sequencing of PCR products 43
2.2.10.2 Sequencing following cloning 43
2.2.11 Cloning PCR products 44
2.2.11.1 Ligations 44
2.2.11.2 Transformations 44
2.2.11.3 Qiagen™ Miniprep kit 45
2.2.11.4 Digests and PCR 46
2.2.12 5'Rapid amplification of cDNA ends 46
2.2.13 Pyrosequencing 48

2.3 Fluorescence in situ hybridisation (FISH) 49
2.3.1 Maxiprep DNA isolation from BAC clones 49
2.3.2 Labelling 50
2.3.3 Slide making 51
2.3.4 Probe preparation 51
2.3.5 Slide preparation 51
2.3.6 Post hybridisation and visualisation 52

2.4 Tissue Culture 53
2.4.1 Imatinib sensivity assay 53
2.4.1.1 Thawing of primary haematopoietic cells 53
2.4.1.2 Colony assay 53
2.4.2 Midi prep & nucleotide removal kit 54
2.4.2.1 Midi-prep 54
2.4.2.2 QIAquick nucleotide removal kit 56
2.4.2.3 Transfection by electroporation 56
2.4.2.4 Electroporation 57
2.4.2.5 Antibiotic selection 57
2.4.2.6 IL-3 independent growth 58

2.5 Solutions 58
2.6 List of probes used for FISH 61

3 IDENTIFICATION OF NOVEL FUSION GENES ASSOCIATED WITH VISIBLE CHROMOSOME ABNORMALITIES 62
3.1 Introduction 62
3.2 A novel PDGFRB fusion gene 62
3.2.1 Patient details 62
3.2.2 Two-colour fluorescence in situ hybridisation (FISH) 63
3.2.3 5’Rapid amplification of cDNA ends (5’RACE) 64
3.2.4 Confirmation of the GOLGA4-PDGFRB fusion and cloning of the genomic breakpoints 69
3.2.5 Consequence of the t(3;5) 75
3.2.6 Identification of a second MPD patient with a GOLGA4-PDGFRB fusion 76
3.2.7 Two-colour fluorescence in situ hybridisation (FISH) 78
3.2.8 Imatinib sensitivity assay 79
3.2.9 Discussion 81

3.3 Identification of two novel PDGFRA fusion genes 83

3.3.1 Patient details 83
3.3.2 Bubble PCR 84
3.3.3 Confirming the STRN-PDGFRA fusion 86
3.3.4 Consequence of the STRN-PDGFRA fusion 88
3.3.5 Confirming the ETV6-PDGFRA fusion 89
3.3.6 Two-colour fluorescence in situ hybridisation 91
3.3.7 Consequence of the ETV6-PDGFRA fusion 92
3.3.8 Detection of MRD and the response to imatinib in cases 1 & 2 93
3.3.9 Discussion: PDGFRA fusions 93

3.4 A novel FGFR1 fusion gene 96

3.4.1 Patient details 97
3.4.2 Fluorescence in situ hybridisation 97
3.4.3 5’Rapid amplification of cDNA ends (5’RACE) PCR 98
3.4.4 Confirming the CPSF6-FGFR1 fusion 99
3.4.5 Consequence of the t(8;12) 99
3.4.6 Discussion 100

3.5 Discussion 102

4 IDENTIFICATION OF DEREGULATED EXPRESSION OF TYROSINE KINASE GENES 107
5.1.2 aUPD and MPDs
5.1.3 aUPD & the discovery of the JAK2 V617F acquired mutation
5.1.4 The association between JAK2 V617F and 9p aUPD
5.1.5 Acquired UPD is not limited to MPDs
5.1.6 Aims of my analysis

5.2 SNP chips to identify regions of UPD
5.2.1 SNP chips
5.2.2 Plan of analysis and methods
5.2.3 Patient samples
5.2.4 SNP array results
  5.2.4.1 Overview of results
  5.2.4.2 JAK2 V617F negative polycythaemia vera
  5.2.4.3 JAK2 V617F negative myelofibrosis
  5.2.4.4 JAK2 V617F positive myelofibrosis
  5.2.4.5 CML in transformation
  5.2.4.6 aCML/aMPDs

5.3 Recurrent aUPD of chromosome 7q
5.3.1 7q abnormalities in myeloid disorders
5.3.2 Recurrent 7q aUPD identified by SNP chip analysis
5.3.3 Screening for mutations in 7q aUPD patients
  5.3.3.1 Candidate tyrosine kinases located within the region of 7q UPD
  5.3.3.2 Expression analysis
  5.3.3.3 Other candidate genes

5.4 Recurrent aUPD of chromosome 11q
5.4.1 11q abnormalities in myeloid disorders
5.4.2 Recurrent 11q aUPD identified by SNP chip analysis
5.4.3 Screening 11q aUPD patients
  5.4.3.1 Candidate genes located within the region of 11q aUPD
  5.4.3.2 Screening the Mixed lineage leukaemia gene
LIST OF FIGURES

Figure 1. 1: A hierarchical model of haematopoiesis..............................2
Figure 1. 2: non-receptor tyrosine kinase signalling.................................13
Figure 1. 3: RTK activation and signalling .............................................14
Figure 2. 1: 5’ Rapid Amplification of cDNA Ends (5’RACE).......................47
Figure 2. 2: A schematic representation of pyrosequencing........................48
Figure 3. 1: FISH analysis of patient sample P1157..................................64
Figure 3. 2: Agarose gel picture of nested 5’RACE PCR .............................66
Figure 3. 3: PDGFRB BLAST results..........................................................67
Figure 3. 4: GOLGA4 BLAST results..........................................................68
Figure 3. 5: Characterisation of the GOLGA4-PDGFRB mRNA fusion..........70
Figure 3. 6: Agarose gel of the GOLGA4-PDGFRB genomic fusion..............72
Figure 3. 7: Agarose gel of the amplification products of the GOLGA4-PDGFRB genomic fusion amplified from plasmid DNA .....................73
Figure 3. 8: GOLGA4-PDGFRB genomic fusion in patient 1 with t(3;5)......74
Figure 3. 9: Schematic representation of the normal PDGFRβ, GOLGA4 and GOLGA4-PDGFRβ protein.................................................................76
Figure 3. 10: Characterisation of the GOLGA4-PDGFRB mRNA fusion in a second patient ....................................................................................77
Figure 3. 11: Fluorescent in situ hybridisation for PDGFRB and GOLGA4......79
Figure 3. 12: Graph showing CFU-GM growth of blood mononuclear cells ......80
Figure 3. 13: Schematic of bubble PCR strategy ........................................85
Figure 3. 14: Agarose gel picture of nested bubble PCR ............................86
Figure 3. 15: Characterisation of the STRN-PDGFRα fusion gene................87
Figure 3. 16: Structure of Pdgfrα, Striatin and the predicted fusion protein ......89
Figure 3. 17: Characterisation of the ETV6-PDGFRα fusion ......................90
Figure 3. 18: t(4;12) FISH analysis.............................................................91
Figure 3. 19: Structure of PDGFRα, ETV6 and the predicted fusion protein ......92
Figure 3. 20: Sequence of the CPSF6-FGFR1 mRNA fusion .....................98
Figure 3.21: RT-PCR Amplification of the CPSF6-FGFR1 fusion .................99
Figure 3.22: Domain structure of FGFR1, CPSF6, and the predicted fusion protein.................................................................100
Figure 4.1: PDGFRB RT-PCR multiplex primer design.......................110
Figure 4.2: PDGFRB RT-PCR multiplex schematic and agarose gel picture...111
Figure 4.3: Agarose gel picture of patients screened by the PDGFRB multiplex.........................................................................112
Figure 4.4: PDGFRA RT-PCR multiplex primer design.......................114
Figure 4.5: PDGFRA RT-PCR assay and agarose gel picture of control samples.............................................................................115
Figure 4.6: Structure of the ETV6-PDGRβ fusion...............................117
Figure 4.7: MLPA traces of patient E508........................................119
Figure 4.8: Schematic representation of the normal PDGRα, KIF5B and the KIF5B-PDGRα fusion.........................................................125
Figure 4.9: A schematic representation of Multiplex Ligation-dependent Probe Amplification (MLPA)......................................................127
Figure 4.10: Schematic representing the MLPA FGFR1 intracellular domain probe..............................................................................129
Figure 4.11: First MLPA run validated with normal control cDNA and cell lines..............................................................................132
Figure 4.12: Third MLPA run.............................................................135
Figure 4.13: Fourth MLPA run – patients with known fusions................136
Figure 4.14: The cDNA MLPA assay, over-expression of PDGFRB.........141
Figure 4.15: Agarose gel of nested 5’RACE PCR for patient E1026 ........143
Figure 4.16: Novel ETV6-PDGFRB fusion .......................................144
Figure 4.17: RT-PCR Amplification of the ETV6-PDGFRB fusion by single step RT-PCR.................................................................145
Figure 4.18: Amplification of the genomic ETV6-PDGFRB novel fusion.........................................................................................146
Figure 4.19: Schematic of PDGRβ and ETV6 proteins ......................148
Figure 4.20: Pie chart showing the molecular pathogenesis of atypical MPD patients.

Figure 5.1: Schematic illustrating the principal classical mutational events that lead to LOH.

Figure 5.2: Mechanisms and consequences of 9p LOH in PV.

Figure 5.3: Schematic representing how the Affymetrix SNP chips work.

Figure 5.4: Example ideograms showing UPD.

Figure 5.5: aUPD of chromosome 1 and the corresponding pyrosequencing trace of the identified MPL mutation in one patient.

Figure 5.6: Figure showing 9p aUPD and chromosome 5q aUPD in Patient 00/1180.

Figure 5.7: Comparison of MNC and granulocyte DNA for patient E1191.

Figure 5.8: Ideograms depicting E632 SNP array results.

Figure 5.9: Analysis of aMPD patient E180 exhibiting whole chromosome 13 aUPD.

Figure 5.10: Copy number and LOH graphs produced for chromosome 7 for three aCML/aMPD patients.

Figure 5.11: Schematic representation of chromosome 7 and the regions of aUPD observed in the three aCML/aMPD patients.

Figure 5.12: Mutations in PIK3CA in various human cancers.

Figure 5.13: 50K SNP array analysis of chromosome 11 in three aCML patients.

Figure 5.14: Schematic representation of chromosome 11 and the regions of aUPD observed in all three aCML/aMPD patients.

Figure 5.15: Exon structure of the 5’ region of the MLL gene illustrating the common e9/e3 PTD.

Figure 5.16: Receptor tyrosine kinase (RTK) negative regulation and the role of CBL.

Figure 5.17: Agarose gels showing amplification of CBL exon 8 and 9 for all three aMPD patients with 11qUPD.
Figure 5.18: Electropherograms showing part of CBL exon 8 amplified from 11q aUPD patients E484 and E1191. 

Figure 5.19: Sequence analysis of patient E632 showing CBL exon 9 mutation P417A. 

Figure 5.20: Schematic showing the CBL protein domain. 

Figure 5.21: The 5bp CBL deletion in patient E110. 

Figure 5.22: Schematic showing the locations and sequencing electropherograms of all four intronic CBL mutations. 

Figure 5.23: Electropherogram for patient E1191 CBL exon 8 mutation. 

Figure 5.24: Electropherogram depicting the CBL mutation in patient E879. 

Figure 5.25: Agarose gel picture showing a patient with exonic CBL deletion. 

Figure 5.26: Schematic detailing chromosome 7 and the regions of aUPD reported. 

Figure 5.27: Schematic detailing a theoretical analysis of chromosome 7. 

Figure 5.28: Location of CBL missense mutations on a schematic representing the protein structure of CBL. 

Figure 6.1: Spider diagram illustrating translocations associated with MPDs and related disorders. 

Figure 6.2: Schematic showing the universal structure of tyrosine kinase fusions based on the BCR-ABL paradigm. 

Figure 6.3A: Schematic showing how the GOLGA4-PDGFRβ fusion might abrogate anterograde and endocytotic/retrograde transport through cellular organelles. 

Figure 6.3B: Schematic showing how the KIF5B-PDGFRα fusion might abrogate anterograde and endocytotic/retrograde transport through cellular organelles. 

Figure 6.3C: Schematic showing how the STRN-PDGFRα fusion might abrogate anterograde and endocytotic/retrograde transport through cellular organelles.
Figure 6.4: Pie chart illustrating the proportion of aMPD cases related to known genetic variants.
LIST OF TABLES

Table 1. 1:  5q33 translocations in leukaemia involving PDGFRB…………………20
Table 1. 2:  Percentages of patients found to have the V617F mutation categorised by disease entity…………………………………………………………22
Table 3. 1:  First set of primers used to amplify the genomic breakpoint in the t(3;5) patient………………………………………………………………………71
Table 3.2:  Components of endocytic machinery found as partners of oncogenic fusion proteins…………………………………………………………105
Table 4. 1:  Patients observed to have PDGFRB over-expression using the PDGFRB multiplex RT-PCR assay…………………………………………..116
Table 4. 2:  Validation of the PDGFRA multiplex. ………………………………..121
Table 4. 3:  Positive cases determined by the PDGFRA RT-PCR assay…….….122
Table 4. 4:  Table showing the MLPA results for the cell line Jurkat……………..133
Table 4. 5:  A table showing the variability between the first MLPA run and the repeat MLPA run using cell lines as controls……………………………..134
Table 4. 6:  Table with rules for analysing unknown patients to be screened by MLPA for possible over-expression of a tyrosine kinase domain….137
Table 5. 1:  Microsatellite marker name, location on chromosome 9p in relation to the JAK2 gene, and the relative heterozygosity of each microsatellite marker. ……………………………………………………………….161
Table 5. 2:  9p LOH in V617F homozygous patients compared to V617F negative cases……………………………………………………………..163
Table 5. 3:  Number of samples sent to the RZPD for 50K SNP chip analysis from each leukaemic subgroup………………………………………169
Table 5. 4:  Summary of SNP array results. ………………………………………172
Table 5.5: Results of microsatellite analysis on the MF case with 9p homozygosity..............................175
Table 5.6: Mutation screening of EPHA1 and EPHB6.........................186
Table 5.7: Mutational analysis of the MET gene .........................188
Table 5.8: Table showing the gene name, accession number and chromosomal location of some potential candidate genes on 7q................191
Table 5.9: Initial list of candidate genes on 11q..........................195
Table 5.10: Summary of CBL mutations according to disease subtype........205
Table 5.11: Summary of CBL mutations identified in this study..............206
Table 5.12: 11q Microsatellite markers..................................................210
Table 5.13: 11q LOH in CBL mutation positive patients.......................211
Table 5.14: MLPA copy number analysis.............................................213
Table 5.15: Results of JAK2 V617F, NRAS & FLT3 ITD mutation analysis in CBL mutated cases. ..........................................................219
Table 5.16: Mutational time course of disease in patient E879................221
Table 5.17: Nine patients with 11q aUPD and a CBL mutation............234
DECLARATION OF AUTHORSHIP

I, Claire Elizabeth Hidalgo-Curtis, declare that the thesis entitled:

Abnormalities affecting tyrosine kinase signalling in atypical myeloproliferative disorders

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 or mainly while in candidature for a research degree at this University;
• where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
• where I have consulted the published work of others, this is always clearly attributed;
• where I have quoted from 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 with 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:


- A novel ETV6-PDGFRB fusion transcript missed by standard screening in a patient with an imatinib responsive chronic myeloproliferative disease.

- The t(1;9)(p34;q34) and t(8;12)(p11;q15) fuse pre-mRNA processing proteins SFPQ (PSF) and CPSF6 to ABL and FGFR1. Hidalgo-Curtis C, Chase A, Drachenberg M, Roberts MW, Finkelstein JZ, Mould S, Oscier D, Cross NC, Grand FH. Genes Chromosomes Cancer. 2008 May;47(5):379-85.


Signed: ..........................................................

Date: ..........................................................
Acknowledgements

I would like to gratefully acknowledge the invaluable supervision, advice and support of Professor Nick Cross and Dr Francis Grand whilst completing this work. I would also like to acknowledge the help of Dr Andy Chase for all his additional technical advice and guidance. I am also thankful to have such tolerant and supportive laboratory colleagues at the Wessex Regional Genetics Laboratory, to name but a few; Amy Jones, Joannah Score, Kathy Waghorn, Thomas Ernst, Victoria Hall and Gemma Watkins. Their encouragement and friendship in all areas has been a great source of strength and comfort. Finally, I want to thank all those behind the scenes. My parents have been an endless source of love and support, believing that I can achieve anything I set my mind to. My brother, Dr Tim Curtis, whose footsteps I have followed (rather more slowly), has always been a great role model and someone to whom I can only aspire. A special thanks for understanding what a PhD entails, and for all your additional advice. Thanks must also go to Sue and Damian for their never-ending support. Last, but most definitely not least, my husband Marcos deserves exceptional gratitude for encouraging me throughout the whole course of my work whilst also putting up with my many crisis and mood swings! He is an endless source of joy and love, and has given me the most precious gift of all, our beautiful baby son Rafë.
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABI</td>
<td>Applied Biosystems</td>
</tr>
<tr>
<td>ABL</td>
<td>Abelson gene</td>
</tr>
<tr>
<td>aCML</td>
<td>atypical CML</td>
</tr>
<tr>
<td>AML</td>
<td>Acute Myeloid Leukaemia</td>
</tr>
<tr>
<td>ARG</td>
<td>Abelson related gene</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine Triphosphate</td>
</tr>
<tr>
<td>BAC</td>
<td>Bacterial Artificial Chromosome</td>
</tr>
<tr>
<td>BAR domains</td>
<td>Bin/Amphiphysin/Rvs domains</td>
</tr>
<tr>
<td>BCR</td>
<td>Breakpoint Cluster Region</td>
</tr>
<tr>
<td>BIN2</td>
<td>Bridging integrator 2</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic local alignment search tool</td>
</tr>
<tr>
<td>BLAT</td>
<td>The BLAST-like alignment tool</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>53BP1</td>
<td>53-Binding Protein 1</td>
</tr>
<tr>
<td>BRLMM</td>
<td>Bayesian Robust Linear Model with Mahalanobis</td>
</tr>
<tr>
<td>BTK</td>
<td>Bruton agammaglobulinemia tyrosine kinase</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CEL</td>
<td>Chronic Eosinophilic Leukaemia</td>
</tr>
<tr>
<td>CEP110</td>
<td>Centriole associated protein 110</td>
</tr>
<tr>
<td>c-FMS</td>
<td>Colony stimulating factor 1 receptor</td>
</tr>
<tr>
<td>CFU-GM</td>
<td>Colony forming unit – granulocyte macrophage</td>
</tr>
<tr>
<td>CIP</td>
<td>Calf Intestinal phosphatase</td>
</tr>
<tr>
<td>CML</td>
<td>Chronic Myeloid Leukaemia</td>
</tr>
<tr>
<td>CML-CP</td>
<td>Chronic Myeloid Leukaemia – Chronic Phase</td>
</tr>
<tr>
<td>CMML</td>
<td>Chronic Myelomonocytic Leukaemia</td>
</tr>
<tr>
<td>CMPD</td>
<td>Chronic Myeloproliferative disorder</td>
</tr>
<tr>
<td>CNL</td>
<td>Chronic Neutrophilic Leukaemia</td>
</tr>
<tr>
<td>CNAT</td>
<td>GeneChip® Copy Number Analysis Tool</td>
</tr>
<tr>
<td>dATP</td>
<td>Deoxyadenosine 5'-triphosphate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model</td>
</tr>
<tr>
<td>IFNα</td>
<td>Interferon alpha</td>
</tr>
<tr>
<td>IL-3</td>
<td>Interleukin 3</td>
</tr>
<tr>
<td>IMF</td>
<td>Idiopathic Myelofibrosis</td>
</tr>
<tr>
<td>JAK2</td>
<td>Janus kinase 2</td>
</tr>
<tr>
<td>JMML</td>
<td>Juvenile Myelomonocytic Leukaemia</td>
</tr>
<tr>
<td>KI</td>
<td>Kinase Insert</td>
</tr>
<tr>
<td>KIF5B</td>
<td>Kinesin family member 5B</td>
</tr>
<tr>
<td>LOH</td>
<td>Loss of heterozygosity</td>
</tr>
<tr>
<td>LREC</td>
<td>Local Research Ethics Committee</td>
</tr>
<tr>
<td>Mb</td>
<td>Megabase</td>
</tr>
<tr>
<td>MDS</td>
<td>Myelodysplastic</td>
</tr>
<tr>
<td>MEK</td>
<td>MAPK [mitogen-activated protein kinase]/ERK kinase</td>
</tr>
<tr>
<td>MET</td>
<td>Hepatocyte growth factor receptor</td>
</tr>
<tr>
<td>MLL</td>
<td>Myeloid/lymphoid or mixed-lineage leukemia</td>
</tr>
<tr>
<td>MLPA</td>
<td>Multiplex Ligation-dependent Probe Amplification</td>
</tr>
<tr>
<td>MMLV</td>
<td>Moloney Murine Leukaemia Virus</td>
</tr>
<tr>
<td>MNCs</td>
<td>Mononuclear Cells</td>
</tr>
<tr>
<td>MPDs</td>
<td>Myeloproliferative disorders</td>
</tr>
<tr>
<td>MPL</td>
<td>Myeloproliferative leukemia virus oncogene</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>MYO18A</td>
<td>Myosin 18A</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Centre for Biotechnology Information</td>
</tr>
<tr>
<td>NIN</td>
<td>Ninein</td>
</tr>
<tr>
<td>NTRKs</td>
<td>Non-receptor tyrosine kinases</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate Buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PDE4DIP</td>
<td>Phosphodiesterase 4D interacting protein (myomegalin),</td>
</tr>
<tr>
<td>PDGF</td>
<td>Platelet derived growth factor</td>
</tr>
<tr>
<td>PDGFR</td>
<td>Platelet derived growth factor receptor</td>
</tr>
<tr>
<td>PDGFRA</td>
<td>Platelet derived growth factor receptor alpha</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PDGFRB</td>
<td>Platelet derived growth factor receptor beta</td>
</tr>
<tr>
<td>PE</td>
<td>Perkin Elmer</td>
</tr>
<tr>
<td>PI3K</td>
<td>Phosphatidylinositol-3-kinases</td>
</tr>
<tr>
<td>PLCγ</td>
<td>Phospholipase C gamma</td>
</tr>
<tr>
<td>PV</td>
<td>Polycythaemia Vera</td>
</tr>
<tr>
<td>5’RACE</td>
<td>5’Rapid Amplification of cDNA Ends</td>
</tr>
<tr>
<td>RCLB</td>
<td>Red Cell Lysis Buffer</td>
</tr>
<tr>
<td>RET</td>
<td>Ret proto-oncogene</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RTKs</td>
<td>Receptor tyrosine kinases</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse Transcriptase – PCR</td>
</tr>
<tr>
<td>SCF</td>
<td>Stem Cell Factor</td>
</tr>
<tr>
<td>SCLL</td>
<td>Stem cell leukaemia-lymphoma</td>
</tr>
<tr>
<td>siRNA</td>
<td>Short interfering RNA</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>STATs</td>
<td>Signal transducer and activator of transcription</td>
</tr>
<tr>
<td>SYK</td>
<td>Spleen tyrosine kinase</td>
</tr>
<tr>
<td>TAP</td>
<td>Tobacco alkaline phosphatase</td>
</tr>
<tr>
<td>TGFβ</td>
<td>Transforming Growth Factor beta</td>
</tr>
<tr>
<td>TGN</td>
<td>Trans Golgi Network</td>
</tr>
<tr>
<td>TK</td>
<td>Tyrosine Kinase</td>
</tr>
<tr>
<td>TRKB</td>
<td>Neurotrophic tyrosine kinase receptor, type 2</td>
</tr>
<tr>
<td>UPD</td>
<td>Uniparental disomy</td>
</tr>
<tr>
<td>WBC</td>
<td>White Blood cells</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
<tr>
<td>ZNF198</td>
<td>Zinc Finger protein 198</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

The focus of the Leukaemia Research group based at the Wessex Regional Genetics Laboratory in Salisbury is the study of a rare subset of neoplastic conditions known as the myeloproliferative disorders (MPDs) and in particular, the association between MPDs and deregulated tyrosine kinases. This introduction discusses normal haematopoiesis and its subversion leading to leukaemia, and specifically the involvement of mutations targeting tyrosine kinase genes.

1.1 Normal Haematopoiesis

Haematopoiesis is the dynamic process of blood cell formation from a common haematopoietic progenitor or stem cell, resulting in the production of a diverse range of cell types with very different functions, ranging from oxygen transport and clotting to the production of antibodies.

Haematopoietic cells first appear in the yolk sac in the second week of embryogenesis. By 8 weeks, haematopoiesis has become established in the liver and spleen of the embryo and by 12-16 weeks these are the major sites of haematopoietic cell formation and they remain active until birth. At 20 weeks gestation the bone marrow takes over as the active site for haematopoiesis, gradually increasing its activity until it then becomes the major site of production. In infants, haematopoiesis occurs in the bone marrow of all bones, whilst in adults haematopoiesis is confined to the flat bones of the sternum, ribs, cranium, pelvis, and the proximal ends of the femur and humerus.

The haematopoietic system itself consists of the haematopoietic stem cell and its progeny, and the supporting bone marrow stroma composed of endothelial cells, fat cells, fibroblasts, macrophages and an extracellular matrix. The stem cells are a self-
renewing population that have the potential to differentiate and become committed to a particular blood cell lineage. Initial differentiation of haematopoietic stem cells is along one of two major pathways, the myeloid (erythrocytes, neutrophils, monocytes/macrophages, eosinophils, basophils and megakaryocytes/platelets) or lymphoid (T-cells, B-cells and natural killer-cells) lineages (Figure 1.1). At this stage the stem cells have become progenitor cells for each type of blood cell, and have lost the capacity for self-renewal and are thus committed to a given cell lineage. These progenitor cells then divide and generate mature blood cells. Commitment of a progenitor cell to a given cell lineage depends upon the acquisition of responsiveness to certain growth factors. The particular microenvironment within which the progenitor cell resides controls differentiation. The bone marrow stroma is where haematopoiesis is controlled providing an inductive microenvironment consisting of an extracellular matrix and either membrane bound or diffusible growth factors (Dexter and Spooncer 1987). Evidence suggests that the dynamics of transcription factor networks within a cell, in addition to the combination of growth factors and cytokines in the environment are associated with commitment of a progenitor cell and hence cell diversity (Hoang 2004).

Haematopoiesis is a continuous process throughout adulthood, and indeed the numbers and distribution of the different cell types in the peripheral blood exhibit little variation under steady-state conditions. This process is regulated by the appropriate balance of cell division, differentiation and apoptosis to maintain a steady state. Apoptosis (also known as programmed cell death) is the process in which unwanted cells are removed. A great number of cells produced by haematopoiesis are destined to die by apoptosis before becoming fully differentiated (Williams, et al 1990). An imbalance between cell proliferation and cell death as a result of genetic mutations are observed in a number of haematological malignancies (Adjei 2001, Cucuianu 2001, Warmuth, et al 1999).
The term leukaemia comes from the Greek word meaning literally ‘white blood’ and was first described by Velpeau in 1827 (Velpeau 1827) but then took over a decade before the disease became recognised by other physicians (Bennett 1845, Cragie 1845, Virchow 1845). Leukaemia is defined as the clonal proliferation and accumulation of blood cells as a result of disruption to normal control mechanisms. The outcome of this is the bone marrow produces large numbers of abnormal cells, assumed to be derived from a single haematopoietic stem cell, thus disrupting the normal production of blood cells. The resultant release of immature cells into the peripheral blood and infiltration of organs such as the liver and spleen causing enlargement are common consequences of the disruption of normal haematopoiesis. Frequently other co-morbidities such as anaemia, thrombocytopenia and immunological dysfunction are also observed.

1.2 Leukaemia
1.2.1 Clonality

There is a large amount of evidence to suggest that haematological malignancies arise from a single ancestral cell that has acquired a selective growth advantage due to acquired somatic mutations, and are thus clonal. In normal individuals haematopoiesis is thought to be polyclonal and this has been shown by clonality assays based on Lyon’s hypothesis of X-inactivation (Lyon 1988). This states that there is random inactivation in females of one X-chromosome and the genes it carries during early embryogenesis. Therefore normal female tissue will be mosaic of cells expressing genes from one or other X-chromosomes whereas clonal tissue will express genes from one X-chromosome only. The assessment of clonality is now an essential element of the characterisation and definition of haematological malignancy. There are a number of different assays used to detect clonal derivation, most dependent on differential methylation patterns between active and inactive X-chromosomes at various polymorphic loci.

Clonality in leukaemia is now well established and is a basic concept for the definition of myeloid leukaemia (Fialkow, et al 1967, Raskind and Fialkow 1987, Wainscoat and Fey 1990). Most forms of leukaemia have demonstrated one variant of a polymorphic allele in all their leukaemic cells (Raskind, et al 1998). Another important step in defining clonality and leukaemia was demonstrated by the Philadelphia chromosome in CML patients, which is found in multiple cell types (neutrophils, basophils, monocytes and B-cells) suggesting that the disease derives from a stem cell rather than a myeloid progenitor (Fialkow, et al 1977).

1.2.2 Classification

Leukaemia is broadly classified as either acute or chronic, referring to the speed at which the leukaemia progresses. In chronic leukaemia the disease is both slow to
develop and progress and is characterised by the over-accumulation of mature cells. In acute leukaemia an amassing of immature haematopoietic blast or progenitor cells is seen due to the rapid progression of the disease. If left untreated, acute leukaemia is almost always fatal. Both classifications can be further subdivided into myeloid and lymphoid which refers to the phenotype of the affected cells.

Classification of the myeloid neoplasms has been defined by the World Health Organisation (WHO) (Vardiman, et al 2002, Vardiman, et al 2009). The focus of my work encompasses two of these categories: the chronic myeloproliferative disorders (CMPD) and the myelodysplastic/myeloproliferative disorders (MDS/MPD). Although these disorders are classified separately, the entities comprising each category are similar in that they are clonal stem cell proliferations with excess myeloproliferation. They exhibit increased marrow cellularity and organomegaly, and are susceptible to undergo genetic evolution associated with bone marrow failure or transformation to blast crisis (acute leukaemia). Despite these similarities, there are differences in the pathogenesis of each specific disease subtype that lead to clinical and morphological disparity.

1.3 The chronic myeloproliferative disorders (CMPD) and the myelodysplastic/myeloproliferative disorders (MDS/MPD)

The CMPD are haematological diseases that arise in a primitive bone marrow progenitor cell (Adamson, et al 1976, Fialkow, et al 1981, Fialkow, et al 1980, Jacobson, et al 1978). They are characterised by proliferation of one or more cells of the myeloid lineages (i.e., granulocytic, erythroid and megakaryocytic), believed to derive from a single neoplastic clone. The neoplastic haematopoiesis shows maturation followed by a corresponding increase in the number of one or more of the mature myeloid lineages in the peripheral blood (Figure 1.1). Splenomegaly and hepatomegaly are common features. These disorders are also prone to undergo
genetic evolution associated with bone marrow failure or transformation to blast phase. A finding of 10-19% blasts in the blood or bone marrow signifies disease acceleration and 20% or above indicates blast phase. There are seven entities identified within the CMPD classification: Chronic myelogenous leukaemia (CML), polycythaemia vera (PV), chronic idiopathic myelofibrosis (IMF), essential thrombocythaemia (ET), chronic neutrophilic leukaemia (CNL), chronic eosinophilic leukaemia (CEL) and unclassifiable CMPD.

The MDS/MPD group was created to accommodate disorders that have both MDS-like and MPD-like features, and as a consequence are difficult to assign to separate categories. They are also clonal haematopoietic neoplasms, however these disorders exhibit dysplastic and proliferative clinical, laboratory and morphological features at the time of initial presentation. They are characterised by hypercellular bone marrow with fibrosis, organomegaly and leukocytosis, commonly neutrophilic proliferation with or without monocytosis, similar to the CMPD. However, the cells are often morphologically dysplastic. Occasionally, simultaneous cytopenia of one or more of the other cell lineages is also observed due to ineffective proliferation or a decrease in the number of precursors. There are three main entities acknowledged in this classification: chronic myelomonocytic leukaemia (CMML), atypical chronic myeloid leukaemia (aCML) juvenile myelomonocytic leukaemia (JMML), in addition to unclassified MDS/MPD.

With the overlap of clinical, laboratory and morphological findings among the individual CMPD and MDS/MPD entities, distinction between the diseases in each category is made more difficult. For this reason, and the lack of knowledge surrounding the underlying molecular pathogenesis, the diagnosis of these disorders is based mainly upon the lineage of the predominant proliferating cells and the prominence of marrow fibrosis, combined with clinical, morphological and laboratory features. Due to the considerable overlap in the disease spectrum of these disorders I shall refer broadly to both groups together as myeloproliferative disorders (MPDs).
To date, there is no single molecular abnormality associated with this subset of disorders, although a pathognomonic molecular marker is associated with CML, and the V617F JAK2 mutation is seen in the great majority of cases of PV as well as other MPDs.

1.4  The paradigm for deregulated tyrosine kinases - Chronic Myelogenous Leukaemia (CML)

CML is perhaps the best characterised MPD it has an incidence of 1-1.5/100,000 population annually (Faderl, et al 1999, Gunz 1977, Young, et al 1981). It is a bi- or tri-phasic disorder with an indolent chronic phase (CML-CP), followed by one or both of the aggressive transformed stages; accelerated phase and/or blast phase. CML usually presents in chronic phase, which is broadly defined by the absence of features of accelerated or blast phase. The haematological indications of CML-CP, as observed on blood smears, show leukocytosis due to the large number and variable maturation stages of neutrophils. Marrow blasts usually account for less than 2% of the white blood cells in the blood and absolute basophilia, a defining feature, is commonly present. The natural course of the disease progression from CML-CP to a more aggressive phase is generally within 3-5 years, and is thought to be due to genetic instability (Mitelman 1993).

1.4.1  Molecular pathogenesis of CML

A molecular marker was first described in 1960, when Nowell and Hungerford discovered a ‘minute chromosome’ associated with the condition, a shortened chromosome 22 (Nowell and Hungerford 1960). It was not until much later in 1985 that the underlying molecular lesion was defined (Shtivelman, et al 1985). The hallmark of CML is the Philadelphia (Ph) chromosome, a reciprocal translocation,
t(9;22)(q34;q11), that fuses BCR on chromosome 22 with the ABL tyrosine kinase on chromosome 9 (de Klein, et al 1982, Faderl, et al 1999, Gunz 1977, Rowley 1973, Shtivelman, et al 1985, Young, et al 1981). The ABL gene on chromosome 9 has homology with a known viral oncogene v-abl, which was identified as the cause of some murine leukaemias. The normal ABL gene encodes a non-receptor tyrosine kinase that is active in signal transduction. The chimaeric BCR-ABL gene however encodes an aberrant tyrosine kinase of higher molecular weight and depending on the location of the breakpoint within BCR (Quackenbush, et al 2000), three oncogenic BCR-ABL fusion proteins: p190, p210 and p230 can be formed. They are all found to be sufficient to generate leukaemia in murine models (Daley, et al 1990, Huettner, et al 2000). Fusion to BCR constitutively activates the ABL tyrosine kinase moiety, and this activation depends on the BCR N-terminal coiled-coil domain (McWhirter and Wang 1997). The effect of this is autophosphorylation of the protein and activation of a number of cellular pathways important for transducing proliferative and anti-apoptotic signals (McGahon, et al 1994, Puil, et al 1994). It has been suggested that the strong resistance to apoptosis along with DNA repair deficiency induced by BCR-ABL allows the accumulation of secondary genetic abnormalities leading to clonal evolution toward blast crisis. Cytogenetically, such abnormalities typically include +8, +Ph or i(17q),(Barnes and Melo 2002, Faderl, et al 1999, Mitelman 1993, Mitelman, et al 1976, Sokal, et al 1988). Abrogation of apoptosis suggests that the expansion of the neoplastic clone may be due not only to a proliferative advantage but also to prolonged survival.

The Ph chromosome is now detected by routine cytogenetic techniques in 85-90% of CML patients at diagnosis. The remaining patients have variant translocations or a submicroscopic translocation of BCR-ABL that can only be detected by molecular techniques or fluorescence in situ hybridisation (FISH) (Hooberman, et al 1989, Sinclair, et al 1997, Werner, et al 1997, Westbrook and Keinanen 1992).
1.4.2 Therapy in CML and the development of Imatinib

Allogeneic stem cell transplantation is currently believed to be the only curative therapy for CML, however only a minority of patients is eligible for this procedure due to a relatively limited number of HLA-matched donors. In the past, the therapy most commonly used to achieve haematological remission in CML-CP was interferon-alpha (IFNα). However, an important advance in the therapy of CML came with the development of imatinib mesylate, a small molecule inhibitor that competes with ATP for occupancy of its specific binding site within the kinase domain of BCR-ABL. Inhibition of tyrosine kinase activity occurs, as imatinib lacks the essential phosphate groups usually provided by ATP. Thus, phosphorylation of substrate proteins cannot occur and they therefore fail to adopt the conformational changes necessary for binding of effector proteins. Ultimately, downstream signalling is prevented (Druker, et al 1996). The initial clinical trials of imatinib included 83 Ph(+) CML patients who had failed IFNα-based treatment and were in CML-CP. A complete haematological response was seen in 98% of patients, with complete cytogenetic responses in 13%. These initial studies were further expanded to include 260 CML patients in myeloid blast crisis. The overall response rate was 64%, with 11% achieving complete remission (<5% marrow blasts). However, 38% of patients either returned to CML-CP or had partial responses.

The results from blast crisis CML suggest partial dependence of the BCR-ABL clone and the likely acquisition of further molecular abnormalities that are not targeted by imatinib (Druker, et al 2001). O’Brien and colleagues later went on to report a comparative study of imatinib mesylate versus combination therapy with IFNα and cytarabine in newly diagnosed CP-CML patients, and found that 97% of these patients achieved a complete haematological remission by 18 months after imatinib mesylate was employed. In addition, a complete cytogenetic response was achieved in greater than 76% of CML patients, and more than 90% had no evidence of disease progression after 18 months (O’Brien, et al 2003). However, to properly assess the
overall survival increases after treatment with imatinib, long-term studies are required.

As previously mentioned, although patients with accelerated or blast phase CML show similar morphological and clinical changes with imatinib therapy, it is not effective in advanced disease (Druker, et al 2001, Druker, et al 1996). Furthermore, a drawback of imatinib is the emergence of acquired resistance, such as point mutations in the kinase domain or amplification of the BCR-ABL gene (Barthe, et al 2001, Branford, et al 2002, Gorre, et al 2001, Hochhaus, et al 2001, Mahon, et al 2000, von, et al 2002). The development of second generation ATP-competitive ABL kinase inhibitors are attempting to overcome some of the problems associated with imatinib resistance and advanced phase disease. These compounds include dasatinib (Sprycel®), nilotinib (Tasigna®), bosutinib (SKI-606) and INNO-406 (NS-187), all of which display increased potency against BCR-ABL. Nilotinib is a novel aminopyrimidine inhibitor (Weisberg, et al 2005) and has been developed as a selective potent BCR-ABL inhibitor. It has shown a good outcome in patients with CP-CML following imatinib resistance and intolerance (Kantarjian, et al 2007), in fact it has been shown in pre-clinical studies to be 30-fold more potent than imatinib. Dasatinib is a potent dual BCR-ABL and Src kinase inhibitor that is structurally unrelated to imatinib (Shah, et al 2004). Its greatest benefits have been observed in patients with CP-CML with 90% achieving a complete haematological response (Steinberg 2007). Much like imatinib, advanced phase-, blast crisis-CML or Ph positive ALL, all showed lower responses. A clinical trial comparing high dose imatinib (800mg/day) and dasatinib showed both produced complete haematological responses however, more achieved a major haematological response with dasatinib (32% versus 7%). Dasatinib has also been found to be effective against KIT and PDGFR, in addition to overcoming resistance exhibited by 18/19 imatinib resistant mutations reported (Shah, et al 2006). Neither of the second generation targeted compounds mentioned here are able to overcome the resistance displayed by the T135I mutant but nevertheless they offer the potential for combined therapy and additional hope in cases where imatinib resistance or intolerance has been observed.
Other therapies are also continuing to be developed, such as those that target
downstream signalling molecules using farnesyl transferase inhibitors (Peters, et al
2001). It is unclear exactly how these compounds work however they are thought to
inhibit the farnesylation of a wide range of target proteins, including RAS, ultimately
resulting in cell growth arrest. In preclinical models these inhibitors showed great
potency against tumour cells, however in clinical trials their activity was far less than
expected. It is possible that inhibition of farnesylation is sufficient for deactivation of
H-Ras, yet insufficient for K-Ras which may be activated by geranylgeranylation
thereby suppressing the effect of the farnesyl transferase inhibitors (Appels, et al
2005). Another area attracting the development of many haematological based
inhibitors is apoptosis-centered therapies. Many agents have been developed that
generally act by stimulating pro-apoptotic proteins and which are at various stages of
preclinical and clinical development,. In particular, compounds have been devised
that act against Akt, over activity of which can enable leukaemic cells to proliferate
and survive. Akt is activated downstream of several oncogenes including BCR-ABL.
An Akt inhibitor has been described that is relatively selective but unfortunately in
preclinical trials proved weak (Mitsiades, et al 2002). Heat shock 90 (HSP90)
inhibitors such as 17-AAG have also been shown to have an effect on Akt by acting
to destabilise and cause overall reductions in Akt levels (Solit, et al 2003).

BCR-ABL was the first human fusion gene to be identified in human malignancy. The
leukaemogenic mechanism in CML now serves as a paradigm for other myeloid
leukaemias, as subsequent to the discovery of BCR-ABL, many other oncogenic
fusion genes have been reported, including several that involve tyrosine kinases.

1.5 The tyrosine kinases

Tyrosine kinases are enzymes that catalyse the transfer of the \( \gamma \)-phosphate of ATP to
protein tyrosine substrates. The phosphorylation of tyrosine and to a lesser extent
threonine and serine is the major mechanism for reversible regulation of protein activity and function. Phosphorylation by protein kinases is central to cellular signal transduction and is responsible for a number of cell functions, including cell-cycle progression, differentiation, apoptosis, cell metabolism and DNA-replication. There are thought to be 518 putative protein kinase genes within the human genome (Manning, et al 2002) of which 90 are tyrosine kinases. These tyrosine kinases are classified as either non-receptor or receptor types (Blume-Jensen and Hunter 2001).

The receptor tyrosine kinases (RTKs) contain a transmembrane domain and are involved in extracellular ligand binding. The non-receptor tyrosine kinases (NRTKs) are intracellular and do not span the plasma membrane. There are 58 RTKs and 32 NRTKs in the human genome.

1.5.1 Non-receptor tyrosine kinases (NRTKs)

The NRTKs include SRC, JAK and ABL which are intrinsic to the cytosol and are maintained in an inactive state. Despite performing a similar role in signal transduction, apoptosis and differentiation as the receptor tyrosine kinases, activation is usually dependent on interaction with surface receptors that lack intrinsic catalytic activity. For example, the JAK family of NRTKs is involved in signal transduction by the cytokines IL-3 and G-CSF (Figure 1.2). The NRTKs can also be activated by intracellular signalling by enabling inhibitor dissociation which is how they are usually maintained in an inactive form. Interestingly, while most NRTKs are cytoplasmic, ABL is localised to both the cytosol and nucleus (Taagepera, et al 1998).
1.5.2 Receptor tyrosine kinases (RTKs)

The RTKs are transmembrane enzymes that are involved in signal transduction at the cell surface and which have been sub-classified into 20 classes based on their kinase domain sequence (Robinson, et al 2000). The class III RTKs includes KIT, FLT3, PDGFRα, PDGFRβ and FMS. They are characterised by five immunoglobulin-like domains in the extracellular ligand-binding region, a single transmembrane domain and two intracellular tyrosine kinase domains (TK1 and TK2) separated by a kinase insert domain (KI) (Schlessinger 2000). They are all thought to play a crucial role in normal haematopoiesis and are thus normally tightly regulated due to the impact they
have on cellular signal transduction. In their inactive states the receptors are thought to exist as monomers embedded in the plasma membrane. Ligand binding to the receptor induces dimerisation, juxtaposing the two catalytic domains causing a conformational change resulting in partial activation of its enzymatic activity (Jiang and Hunter 1999). Full activity follows only after autophosphorylation of key tyrosine residues in the activation loop. These phosphorylated tyrosines then act as docking sites for adaptor proteins that further recruit other downstream cytoplasmic signalling molecules. Subsequently a cascade of signalling events through the JAK/STAT, PLCγ, PI3K and Ras-Raf-MEK/ERK occurs leading to changes in expression of genes involved in the control of cell proliferation, differentiation and cell survival (Kharas and Fruman 2005) (Figure 1.3). Activated receptors are rapidly endocytosed and either degraded or recycled to the plasma membrane.

Figure 1.3: A schematic representation of RTK activation and signalling.
1.6  Molecular pathogenesis and deregulated tyrosine kinases

Following the identification of BCR-ABL in CML, a number of other tyrosine kinases have been implicated in the pathogenesis of leukaemia. In fact, in the past two decades several mechanisms that deregulate tyrosine kinases, such as receptor amplification, chromosomal translocations and point mutations have been identified, with reports of more that 30 deregulated tyrosine kinases associated with cancer (Blume-Jensen and Hunter 2001). In MPD patients in particular, translocation cloning has lead to the identification of numerous fusion genes involving tyrosine kinase genes, for example the fibroblast growth factor receptor 1 (FGFR1) on chromosome 8p11. The resulting myeloproliferative disorder is loosely referred to as the 8p11 myeloproliferative syndrome (EMS), also known as stem cell leukaemia-lymphoma (SCLL) (Macdonald, et al 2002a).

constitutive enzymatic activity and deregulate cell growth in a similar fashion to that generated by BCR-ABL (Cross and Reiter 2002).

In addition to translocation cloning, loss of heterozygosity (LOH) has recently re-emerged as an extremely useful tool for identifying regions harbouring potential candidate oncogenes or tumour suppressor genes in leukaemia. This is exemplified by the recent discovery of the V617F JAK2 mutation in patients with PV, MF and ET. A study of PV patients in 2002 using microsatellite screening of the whole genome, showed significant loss of heterozygosity (LOH) of chromosome 9p in all PV patients tested (Kralovics, et al 2002). LOH was observed without change of copy number, indicating that it occurred by acquired uniparental disomy (UPD) as a consequence of mitotic recombination rather than heterozygous deletion. Within the minimal region of UPD they identified a number of possible candidate genes, including JAK2. UPD for other regions has also been associated with gain of function mutations, for example UPD 13q and FLT3 internal tandem duplications (Fitzgibbon, et al 2005, Griffiths, et al 2005). Although these results did not actually lead to the discovery of the JAK2 or FLT3 mutations, they suggest that screening for LOH/UPD may be a useful approach to identify candidate oncogenes.

At the current time it is important to identify patients with molecular lesions involving imatinib responsive genes. Although designed for use against the ABL tyrosine kinase, imatinib has also been shown to inhibit the tyrosine phosphorylation of ARG, KIT, FMS, PDGFRα and PDGFRβ. On identification of a lesion involving one of these genes in a MPD patient, imatinib therapy can be successfully initiated. Translocations involving imatinib responsive genes are commonly described in a minority of MPD patients, notably involving rearrangements of the PDGFR genes.
1.7 PDGFRA and PDGFRB rearrangements in MPDs

PDGFRα and PDGFRβ are 2 highly related RTKs with 85% and 75% homology between the two intracellular kinase domains (Matsui, et al 1989). PDGF ligand consists of a family of four dimeric isoforms linked by disulphide bonds, PDGF-AA, PDGF-BB, PDGF-CC and PDGF-DD. The N-terminal region of each ligand must be proteolytically removed to enable binding to the receptor (Bergsten, et al 2001). PDGFRβ binds the ligands PDGF-BB and PDGF-DD agonistically, whereas PDGFRα binds all the isoforms with the exception of PDGF-DD. While PDGFRs can form homo-and heterodimers between the α and β receptors, it has also been shown to have the ability to dimerise with the epidermal growth factor receptor (EGFR) and the hybrid can also be stimulated by PDGF (Saito, et al 2001). Both PDGFR receptors have been linked to haematological malignancies and in particular the MPDs.

1.7.1 Fusions involving PDGFRA

Over-expression and amplification of PDGFRA has been described in only a minority of human malignancies (Bozzi, et al 2007, Ebert, et al 1995, Hermanson, et al 1996, Holtkamp, et al 2007, MacDonald, et al 2001, Puputti, et al 2006, Zhang, et al 2005), but specific abnormalities have been associated with patients with a CML-like MPD. A rare variant translocation, t(4;22)(q12;q11), resulting in a constitutively active BCR-PDGFRα fusion protein was first detected in a patient with atypical CML (Baxter, et al 2002). Following this discovery, Cools et. al., (2003) discovered an interstitial cytogenetically invisible deletion of 4q12 resulting in the fusion of FIP1L1 to PDGFRA. The resulting oncoprotein was found to transform Ba/F3 cells to growth factor independence, similar to BCR-ABL. Although originally found in a patient described clinically with HES, patients with FIP1L1-PDGFRA are now classified as CEL since the presence of the fusion indicates clonality (Cools, et al 2003a). More recently, the link between PDGFRA fusion genes and CML-like MPDs has been

1.7.2 Chronic eosinophilic leukaemia (CEL)

CEL is the clonal proliferation of eosinophilic precursors that results in persistently increased numbers of eosinophils in the blood, bone marrow and peripheral tissues. Organ damage occurs because of leukaemic infiltration or the release of cytokines, enzymes or other proteins by the eosinophils. Unusually, eosinophilic leukaemia shows a marked male predominance, and in particular, HES has been found to have a 9:1 male to female ratio. Due to the difficulty in differentiating CEL and HES, the prevalence of these disorders is unknown, although it is clear that both are rare. However if there is no evidence of clonality of the eosinophils, a diagnosis of HES is preferred, whereas the finding of a clonal myeloid abnormality supports the diagnosis of CEL (Bain 1996, Weller and Bubley 1994). The FIP1L1-PDGFRA cytogenetically cryptic fusion gene was first identified in patients with HES/CEL. Patients with the fusion protein show an excellent clinical response to imatinib mesylate (Cools, et al 2003a).

A staurosporine derivative, PKC412, has also been developed to inhibit the PDGFRs and other kinases such as KIT and FGFR3 (Chen, et al 2005, Gotlib 2005, Growney, et al 2005, Seo, et al 1999), and although it has not been used in patients with PDGFR-rearrangements yet, it has been shown to overcome imatinib-resistance in a murine model of FIP1L1-PDGFRα induced MPD (Cools, et al 2003b). It has also now entered Phase II trials for FLT3 inhibitor therapy in AML with encouraging results however the emergence of primary and secondary resistance does occur in the majority of patients (Stone, et al 2005). Other inhibitors such as Sorafenib are also active against FIP1L1-PDGFRA and its most common imatinib-resistant mutant (Lierman, et al 2006).

A number of other clonal cytogenetic abnormalities have also been reported in rare cases of CEL, CMML with eosinophilia and aCML primarily involving PDGFRB,
FGFR1 and JAK2. Acquired reciprocal chromosomal translocations involving the chromosome band 5q31-33 (PDGFRB), in particular, are associated with a significant minority of MPD patients.

1.7.3 **Fusions involving PDGFRB**

Aberrant activation of PDGFRB resulting in cell transformation was first substantiated by Waterfield et. al., (1983) (Waterfield, et al 1983) who provided evidence that SIS, an oncogenic form of the PDGFR ligand, transformed cells through activation of its receptor. Nevertheless, demonstration of the direct association with PDGFRB in leukaemogenesis took another decade, when a rare but recurring cytogenetic abnormality was reported in patients with atypical CML. The abnormality was shown to be a t(5;12)(q33;p13) resulting in an ETV6-PDGFRB fusion gene (Golub, et al 1994). In an analogous manner to the BCR-ABL oncoprotein, pathogenicity is dependent on both the ETV6 dimerisation motif and the tyrosine kinase activity of PDGFRβ. Reports of additional chromosomal translocations involving PDGFRB in patients with MPDs, strengthens the association between myeloid leukaemias and deregulated tyrosine kinases. To date, 21 partner genes have been reported (Table 1.1).
<table>
<thead>
<tr>
<th>TRANSLOCATION</th>
<th>PARTNER GENE</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>t(5;7)(q33;q11)</td>
<td>HIP1</td>
<td>(Ross, et al 1998)</td>
</tr>
<tr>
<td>t(5;14)(q33;q32)</td>
<td>CEV14(TRIP11/GMAP210)</td>
<td>(Abe, et al 1997)</td>
</tr>
<tr>
<td>t(5;17)(q33;p13)</td>
<td>Rabaptin-5/RABEP1</td>
<td>(Magnusson, et al 2001)</td>
</tr>
<tr>
<td>t(1;5)(q23;q33)</td>
<td>PDE4DIP (myomegalin)</td>
<td>(Wilkinson, et al 2003)</td>
</tr>
<tr>
<td>t(5;14)(q33;q24)</td>
<td>NIN</td>
<td>(Vizmanos, et al 2004)</td>
</tr>
<tr>
<td>t(5;17)(q33;p11.2)</td>
<td>HCMOGT-1/SPECC1</td>
<td>(Morerio, et al 2004)</td>
</tr>
<tr>
<td>t(5;15)(q33;q15)</td>
<td>TP53BP1</td>
<td>(Grand, et al 2004b)</td>
</tr>
<tr>
<td>t(5;14)(q33;q32)</td>
<td>KIAA1509</td>
<td>(Levine, et al 2005b)</td>
</tr>
<tr>
<td>t(5;12)(q33;p13.3)</td>
<td>ERC1</td>
<td>(Greiello, et al 2008)</td>
</tr>
<tr>
<td>t(1;5)(q21;q33)</td>
<td>TPM3</td>
<td>(Rosati, et al 2006)</td>
</tr>
<tr>
<td>t(3;5)(q22;q33)</td>
<td>WDR48</td>
<td>SUBMITTED</td>
</tr>
<tr>
<td>t(5;12)(q33;q13.13)</td>
<td>Bin2</td>
<td>SUBMITTED</td>
</tr>
<tr>
<td>t(3;5)(q22;q33)</td>
<td>GOLGA4</td>
<td>SUBMITTED</td>
</tr>
<tr>
<td>t(5;12)(q31-33;q24)</td>
<td>GIT2</td>
<td>(Walz, et al 2007)</td>
</tr>
<tr>
<td>t(4;5)(q23;31;q33)</td>
<td>PRKG2</td>
<td>(Gallagher, et al 2008, Walz, et al 2007)</td>
</tr>
<tr>
<td>t(4;5)(q21;q33)</td>
<td>GPIAP1</td>
<td>(Walz, et al 2007)</td>
</tr>
<tr>
<td>Der(1)(1;5)(p34;q33),der(5)(1;5)(p34;15),</td>
<td>GPIAP1</td>
<td>(Walz, et al 2007)</td>
</tr>
<tr>
<td>der(11)ins(11;5)(p12;q15q33)[23];46XY[1]</td>
<td>NDEI</td>
<td>(La Starza, et al 2007)</td>
</tr>
<tr>
<td>t(5;16)(q33;p13)</td>
<td>NDEI</td>
<td>(Walz, et al 2009)</td>
</tr>
<tr>
<td>t(5;17)(q33-34;q11.2)</td>
<td>MYO18A</td>
<td>(Walz, et al 2009)</td>
</tr>
<tr>
<td>t(2;5)(p21;q33)</td>
<td>SPTBN1</td>
<td>(Gallagher, et al 2008)</td>
</tr>
</tbody>
</table>

Table 1.1: 5q33 translocations in leukaemia involving PDGFRB
1.7.4 *Chronic myelomonocytic leukaemia (CMML)*

CMML is a clonal disorder of a bone marrow stem cell, in which monocytosis is the hallmark. It is, however, a heterogeneous disorder; in some cases, the white blood cell (WBC) count is elevated due to neutrophilia and monocytosis with minimal dysplasia, but significant splenomegaly, and the disease is more reminiscent of a CMPD than MDS. In other patients, although monocytosis is present, neutrophil counts are reduced or normal and there is prominent dysplasia and no splenomegaly, so that the disease resembles MDS rather than CMPD. The median age at diagnosis of CMML is 65-75 years, with a male to female ratio reported to be between 1.5-3:1 (Germing, et al 2000, Onida, et al 2002). The most frequent recurring chromosomal abnormalities include trisomy 8, monosomy 7/deletion (7q) and structural abnormalities of 12p and i(17q); approximately 20-40% of CMML patients have detectable clonal abnormalities (Fenaux, et al 1996, Toyama, et al 1993).

A subset of CMML cases, generally those associated with marked eosinophilia, have also been associated with *PDGFRB* fusion genes (Golub, et al 1994, Kulkarni, et al 2000, Magnusson, et al 2001, Sawyers and Denny 1994, Schwaller, et al 2001). Mutations of *RAS* are also detected, in approximately a third of CMML cases (Willman 1998). Mutations of *FLT3* may also be present in this group, a recent study by Shih et. al, (2002) found 3 out of 7 CMML patients carried the internal tandem duplication (Shih, et al 2002).

Apperley et. al., (2002) has also provided evidence that patients with rearrangements of the *PDGFRB* gene respond well to treatment with imatinib mesylate (Apperley, et al 2002). These findings make it increasingly important to gain an accurate molecular diagnosis of *PDGFR*A and *PDGFRB* gene rearrangements, as it is likely they will play a role in the disease management of such patients.
1.7.5 JAK2 and the myeloproliferative disorders

Aberrant tyrosine kinase signalling has been implicated as the pathogenic mechanism of so many of the MPDs, it has not come as a surprise to find the presence of an acquired point mutation in the Janus Kinase 2 (JAK2) gene predominantly in patients with PV, IMF and ET. The JAK2 mutation itself has been the single most important discovery in the pathogenesis of MPDs in the last decade. The guanine to thymine substitution in exon 12 of JAK2 changes a highly conserved valine at position 617 to phenylalanine (V617F). This alteration is within the pseudokinase (JH2) domain and acts by disrupting the inhibitory regulation of JAK2, leaving it constitutively activated (Baxter, et al 2005, James, et al 2005, Jones, et al 2005, Kralovics, et al 2005a, Levine, et al 2005a). The activated kinase can then bind to a receptor and recruit STATs independent of haematopoietic growth factors. Five groups have independently confirmed the identification of the V617F mutation. The numbers of PV, IMF and ET patients found to have the mutation is remarkable, and will now act as an invaluable diagnostic tool to distinguish between the MPD and other reactive conditions such as secondary thrombocytosis and erythrocytosis (Table 1.2).

Interestingly, this mutation does not occur in non-haematological malignancies and is uncommon in myeloid disorders other than the classic BCR-ABL negative MPDs (Scott, et al 2005).

<table>
<thead>
<tr>
<th>Disease</th>
<th>No. of cases</th>
<th>No. of V617F positive</th>
<th>No. of cases</th>
<th>No. of V617F positive</th>
<th>No. of cases</th>
<th>No. of V617F positive</th>
<th>No. of cases</th>
<th>No. of V617F positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polycythaemia vera</td>
<td>73</td>
<td>71(97%)</td>
<td>164</td>
<td>121(74%)</td>
<td>128</td>
<td>83(65%)</td>
<td>72</td>
<td>58(81%)</td>
</tr>
<tr>
<td>Essential thrombocythaemia</td>
<td>51</td>
<td>29(57%)</td>
<td>115</td>
<td>37(32%)</td>
<td>93</td>
<td>21(23%)</td>
<td>59</td>
<td>24(41%)</td>
</tr>
<tr>
<td>Chronic idiopathic myelofibrosis</td>
<td>16</td>
<td>8(50%)</td>
<td>46</td>
<td>16(35%)</td>
<td>23</td>
<td>13(57%)</td>
<td>35</td>
<td>15(43%)</td>
</tr>
</tbody>
</table>

**Table 1.2:** Percentages of patients found to have the V617F mutation categorised by disease entity.
However, a number of questions regarding JAK2 V617F remain including: how an identical mutation can cause three separate clinical diseases, and why some patients are heterozygous for the mutation and others are homozygous. If those with homozygous mutations are the result of two separate steps, why is it that the first is enough to cause disease features alone?

Since the discovery of the JAK2 V617F mutation a considerable amount of investigation has been performed trying to answer these, among other questions, with no definitive conclusions. It is odd that only a single point mutation occurs in this gene at this position. If this codon is prone to mutation why is only one amino acid substitution observed? What makes this single mutation activate JAK2 in the majority of patients with these disorders given the fact that other activating mutations such as T875N have been found in acute megakaryoblastic leukemia (Mercher, et al 2006). Work has recently been published looking at this anomaly: Dusa et. al., 2008 have used random site-directed mutagenesis to mutate codon 617 to each of the 20 amino acid possibilities and tested their ability to induce constitutive signalling in vitro in Ba/F3 cells expressing the erythropoietin receptor. Only 4 mutations, V617W, V617M, V617I and V617L were able to induce cytokine independence and constitutive downstream signalling. Additionally, V617W was the only mutant that produced constitutive activity comparable to that of V617F, and could generate an MPD in mice that displayed predominant erythrocytosis and megakaryocytic proliferation (Dusa, et al 2008). Nevertheless, to achieve this amino acid substitution would require a three base pair change (GTC>TTG) and this is unlikely to occur naturally in patients.

Since this initial finding of V617F, other mutations have been identified in JAK2 exon 12 in PV patients negative for the JAK2 V617F mutation. In contrast, these mutations are varied in terms of their molecular basis, with substitutions, complex deletions and insertions identified (Butcher, et al 2008, Scott, et al 2007b). These mutations account for one third of all bona fide V617F negative PV patients although
this figure will probably turn out to be greatly reduced with screening of increased numbers (Scott, et al. 2007a).

In vitro and in vivo data suggests JAK2 V617F is sufficient for the development of PV, ET and IMF, however this does not account for how three phenotypically related, but clinically distinct MPDs can arise from an identical JAK2 mutation. Additional genetic and epigenetic events have been hypothesised to contribute to this phenotypic pleiotropy. It is possible there are still unidentified mutations that contribute to the pathogenesis of these disorders or there may be unknown initiating events which precede the acquisition of the JAK2 V617F mutation. Pardanani et. al., 2007 screened for genetic differences in allele frequencies of SNPs in JAK2, EPOR, GSCF and MPL and found a number of SNP and haplotype associations linked with PV and not ET or IMF in the JAK2 gene. This group concluded that genetic variation in MPDs may contribute to phenotypic diversity in the presence of specific mutations (Pardanani, et al. 2007). In addition, a study by Bellanne-Chantelot et. al., 2006 looked at familial MPDs. Somatic JAK2 V617F mutations were identified in some, but not all MPD patients. In fact some families showed both JAK2 V617F-positive and -negative affected members, and a small number of relatives without a diagnosis of PV, ET or MF had endogenous erythroid colony formation in the absence of the JAK2 V617F allele, suggesting perhaps that unidentified mutations may be contributing to the pathogenesis of PV, ET and MF, regardless of JAK2 mutational status (Bellanne-Chantelot, et al. 2006). In addition to phenotypic pleiotropy and inherited SNPs, epidemiological data and family studies have also suggested that inherited factors may also predispose to these specific MPDs (Landgren, et al. 2008, Rumi, et al. 2008). This work has been expanded and a strong association between JAK2 V617F disease and specific constitutional JAK2 haplotypes has recently been identified, providing the first evidence in MPDs of inherited genetic factors influencing the risk of acquiring a specific somatic mutation (Jones, et al. 2009, Kilpivaara, et al. 2009, Olcaydu, et al. 2009).

It has also been suggested that the JAK2 V617F mutation may arise in different progenitor cell populations thus accounting for the difference in clinical phenotype
between the three separate disorders found to harbour this mutation. Jaimeson et. al.,
2006 recently used flow cytometry to isolate different progenitor cell populations
from PV patients and showed the mutation was observed in all cell fractions
providing evidence that PV arises in haematopoietic stem cells (HSCs) and involves
myeloid, erythroid and megakaryocytic lineages (Jamieson, et al 2006) and thus, like
CML, V617F positive MPDs are believed to arise from an early haemopoietic
progenitor or stem cell.

A further interesting outcome has been the discovery that homozygous clones occur
in most patients with PV but are never found in patients with ET. In addition to this
observation is that ET JAK2 V617F positive patients have phenotypic similarities of
PV for instance, increased haemoglobin, increased numbers of neutrophils, bone
marrow erythropoiesis and granulopoiesis and a propensity to venous thromboses. ET
patients lacking the JAK2 V617F mutation also differ in that they have more
pronounced and isolated thrombocytosis, less active erythropoiesis and
granulopoiesis. From these in-depth studies it has been suggested that V617F positive
thrombocythaemia and polycythaemia could be viewed as a continuum of disease and
not as two separate entities, with depleted iron stores, genetic modifiers and
erthropoietin homeostasis tipping the disease in favour of ET and sex (mainly biased
toward men), genetic modifiers and homozygosity for V617F pushing the disease
balance toward PV (Campbell, et al 2005).

Whilst these questions still remain unresolved, work has at least progressed toward
possible therapeutic strategies that will hopefully become clinically effective.
Targeted potent inhibitors of JAK2 are currently in development for the treatment of
compounds that are undergoing preclinical and clinical testing inhibit both wild-type
JAK2 and JAK2 V617F. JAK2 signalling is important in many cellular functions
hence inhibition of wild-type may result in unwanted haematopoietic side effects.
Additionally JAK2 is one of a closely related family and wild-type JAK3 inhibition
has to be considered due to its association between loss of function mutations and

25
severe combined immunodeficiency (Macchi, et al 1995, Russell, et al 1995). Nevertheless many compounds are now in clinical trials and early results indicate objective improvements in many patients but, disappointingly, only minimal reductions in the size of the neoplastic clone.

1.8 Summary

The BCR-ABL paradigm in CML has proved a useful model for the study of other MPDs and aberrant tyrosine kinase signalling has been identified as a common factor driving myeloproliferation. Chromosomal translocations have been instrumental in the identification of numerous deregulated kinases, however it is apparent from the discovery of the V617F JAK2 mutation that LOH/UPD analysis is also a valuable tool for identifying chromosomal regions containing novel oncogenes.

The efficacy of targeted therapy of deregulated tyrosine kinases makes it critically important to identify patients harbouring these molecular lesions. This is especially true for mutations involving PDGFRα or PDGFRβ as these are imatinib sensitive.

1.9 Statement of aims

The general aim of this study is to further our knowledge about the molecular pathogenesis of the MPDs and to pursue the hypothesis that these diseases are caused primarily by deregulation of tyrosine kinases. My specific aims are:

i) To characterise acquired chromosomal translocations seen in association with MPDs
ii) To develop screening procedures capable of detecting overexpression of specific tyrosine kinase genes and to investigate the molecular basis for any observed overexpression.

iii) To investigate MPDs for regions of acquired uniparental disomy as an indicator towards possible oncogenes or tumour suppressor genes.
2 METHODS

2.1 Sample processing

2.1.1 Preparation of blood and bone marrow samples

Patient samples were either (i) referred from Consultant Haematologists throughout the UK for routine diagnostic analysis (usually the detection of specific acquired genetic abnormalities or determination of \textit{in vitro} sensitivity to imatinib), or (ii) provided by collaborators from abroad. For research purposes, UK samples were anonymised following diagnostic analysis and reporting of results. Samples from abroad were provided in coded form. Use of these samples for research was approved by the Salisbury and South Wiltshire LREC (LREC reference 05Q2008/6; mechanisms and consequences of tyrosine kinase activation in chronic myeloproliferative disorders and related conditions; chief investigator Professor NCP Cross).

Patient material usually ranged from 5-50ml of peripheral blood, or 1-10ml bone marrow aspirate. In some cases fixed cell suspensions for cytogenetics were also provided. Mononuclear cells (MNCs) and granulocytes were separated by centrifugation over Lymphoprep\textsuperscript{TM} (Axis Shield, Oslo, Norway). The cell counts of the peripheral blood and bone marrow samples were determined using a Neubauer type haemocytometer (BDH, Poole, UK). 2x10\textsuperscript{7} cells were diluted in Hanks’ balanced salt solution (HBSS) to a maximum volume of 25ml and this suspension was then layered over an equal volume of lymphoprep in a 50ml propylene tube (Sarstedt, UK). Centrifugation at 1,800rpm for 30 minutes without a brake created an interface comprised of MNCs. The MNCs were transferred to a new tube using a sterile pasteur pipette (BDH, Poole, UK). The recovered cells were then washed twice in RPMI 1640 supplemented with 10% foetal calf serum (FCS), L-glutamine and penicillin-streptomycin before counting. To obtain the granulocytes, the remainder of the lymphoprep mixture from the original 50ml tube was discarded,
leaving the red cell pellet. To the pellet, ice-cold red cell lysis buffer (RCLB) was added to lyse the red cells. The tubes were then placed on ice for 10 minutes (intermittent mixing), and then spun for 5 minutes at 1600rpm. Before counting, the white cell pellet obtained was washed twice in supplemented RPMI (as before). For RNA extraction 1x10⁷ cells were lysed in either 1ml guanidinium thiocyanate (GTC; the composition of all solutions are given in section 2.5) freshly mixed with β-mercaptoethanol at a ratio of 7μl of β-mercaptoethanol (Sigma) per 1ml GTC lysate or 1.5x10⁷ cells were lysed in 1.5ml TRIZOL® reagent (Invitrogen, UK) and stored at -70°C. For DNA extraction, cell pellets of 1-2.5x10⁷ cells were frozen at -70°C. Excess MNCs were cryopreserved in aliquots of between 2.5 and 6x10⁷ cells. Each cell aliquot was spun down and resuspended in 500μl of supplemented RPMI (10% FCS, L-glutamine and penicillin-streptomycin). An equal volume of freezing mix (composition detailed in section 2.5) is added dropwise to the resuspended cells. Finally, 1ml of this mixture was again added dropwise to a cryopreservation tube and frozen overnight at -70°C before being transferred to the liquid nitrogen stores.

2.1.2 Enriching cultures for T-cells

To investigate if a mutation has been acquired, DNA prepared from a granulocytic fraction can be compared to DNA from a T-cell fraction. T-cells are lymphoid in origin and are not thought to be involved in myeloproliferative disease. Phytohaemagglutinin (PHA) is a lectin extracted from white kidney beans and is used to stimulate mitotic division of lymphocytes maintained in cell culture.

Peripheral blood was first separated by centrifugation over Lymphoprep™ (Axis Shield, Oslo, Norway) into mononuclear cells (MNCs) and granulocytes as detailed in section 2.1.1. The cell count of the mononuclear cells was determined using a Neubauer type haemocytometer (BDH, Poole, UK). Two cultures of 5x10⁶ cells were then resuspended in 5ml of RPMI 1640 supplemented with 10% foetal calf serum (FCS), L-glutamine and penicillin-streptomycin in a Nunclon® polystyrene flat-sided cell culture tube (Scientific Laboratory Supplies Ltd, Nottingham). To one culture
500µl of myeloid cytokine mix was added (GM-mix, for components see 2.5) and to the other 50µl of PHA. After 3 days incubation at 37 °C the cell suspension was spun at 1500rpm for 5 minutes. The cell pellet was then resuspended in 2ml of supplemented RPMI 1640 (as above) and the cells counted (as previously). Two aliquots of 4 x10⁴ cells were used to produce two post-PHA statspins. The remaining cells were spun at 1500rpm for 5 minutes and DNA made from the cell pellet according to the protocol in section 2.1.5.2. The statspins were taken to our Pathology department where both the pre- and post-PHA slides were stained using May-Grunwald Giemsa stain in an automated machine. (This used to involve fixing the cells in methanol for 15 minutes, staining in May-Grunwald for 5 minutes, then staining in Giemsa for 10 minutes. The slide was then rinsed first in buffer pH 6.8, followed by a 50/50 buffer/acetone rinse, dehydration in acetone twice and cleared using three xylene washes). This stain enables different types of cell (erythrocytes, lymphocytes, etc.) and components within the cell, to be distinguishable from one another by their colour and shape when viewed under a microscope. Lymphocytes look like circular cells with large round blue nuclei that take up ~60% of the cytoplasm with no vesicles. Pre-and post-PHA statspins were used to assess the level of T cell enrichment. DNA made from post-PHA incubation was used as a control for the investigation of acquired mutations in the myeloid cell fraction.

2.1.3 RNA extraction

RNA was extracted from GTC samples using the Qiagen RNeasy kit (Qiagen, West Sussex). 350µl of GTC was mixed with an equal volume of 70% ethanol, mixed well, and applied to the Qiagen column then left to stand for 30 minutes. The column was spun for 10 seconds at 10000rpm in a Genofuge 16M centrifuge. The column was then washed with the buffers provided in the kit: 650µl of RW1 for 10 seconds at 10000rpm, and then twice with 500µl RPE, the first time for 10 seconds at 10000rpm and the second wash at 14000rpm for 2 minutes. Columns were then left in a class II tissue culture cabinet with the lids open for 10 minutes to allow evaporation of any
remaining ethanol. The bound RNA was then eluted from the column with 25μl distilled water. RNA was usually directly reverse transcribed into cDNA.

RNA was extracted from samples in TRIZOL® using chloroform followed by the Qiagen RNeasy kit. Briefly, to 350μl TRIZOL® sample 350μl of chloroform (Sigma Aldrich) was added and vortexed for 15 seconds. The sample was then spun at 14,000 rpm for 5 minutes at room temperature. Centrifugation separates the solution into an aqueous phase and an organic phase, the RNA remains in the aqueous phase. The aqueous phase was transferred to a clean eppendorf and washed again with 350μl chloroform as before. The final aqueous phase was carefully transferred to a clean eppendorf containing 350 μl of 70% ethanol, mixed well and then applied to an RNeasy column. The protocol continues as specified above once on the column.

2.1.4 cDNA synthesis

After elution from the column the RNA was placed on ice and then heated to 65°C for 5 minutes. The cDNA synthesis was performed at 37°C for 2 hours utilising 1-5μg RNA in 25μl (GTC samples) or 50μl (TRIZOL® samples) water. The RNA was combined with 21μl (GTC samples) or 42μl (TRIZOL® samples) of cDNA mix, which resulted in final concentrations of 50mM Tris pH8.3, 75mM KCl, 3mM MgCl₂, 1mM DTT, 1mM dATP, dCTP, dTTP, dGTP, as well as 100μg/μl of random pd(N)₆ hexamers (Amersham Pharmacia, Amersham, UK), 150 units of Moloney Murine Leukaemia Virus (MMLV) reverse transcriptase (Invitrogen, UK) and 15 units/ml of RNasin (Promega, UK). Synthesis was terminated by heating to 65°C for 10 minutes. cDNA was stored at -20°C.

2.1.5 DNA Extraction

2.1.5.1 DNA Extraction from fixed cell suspension
Fixed cell suspensions were centrifuged at 4000 rpm for 10 minutes in a Thermo IEC, Micromax centrifuge. The cell pellet was then resuspended in 1ml of sterile PBS before spinning the pellet as before and then the washing was repeated three times. After the third centrifugation step the pellet was resuspended in 200µl of sterile PBS (independent of the number of cells) and the Qiagen DNeasy Blood & Tissue kit protocol: ‘Purification of total DNA from animal blood or cells’ was followed starting from step 1c. To the resuspended pellet in 200µl of PBS, 20µl proteinase K and 200µl of Buffer AL (without ethanol added) was added. This was mixed thoroughly by vortexing before incubation at 56ºC for 10 minutes. 200µl ethanol (96-100%) was then added and the sample mixed thoroughly by vortexing, this mixture was then added to the DNeasy Mini spin column and centrifuged at 8000rpm for 1 minute. The collection tube containing the flow through was then discarded and the spin column containing the bound DNA placed in a new collection tube. 500µl of Buffer AW1 was added to the spin column and centrifuged at 8000 rpm for 1 minute. Again the collection tube was discarded and the spin column placed in a new collection tube, before the addition of 500µl Buffer AW2 and centrifugation at 14,000 rpm for 20 minutes to dry the DNeasy membrane. The collection tube was discarded again and the spin column placed in a sterile 1.5ml microcentrifuge tube. 50µl Buffer AE was added directly to the DNeasy membrane and incubated at room temperature for 1 minute before centrifugation for 1 minute at 8000rpm. To increase the DNA yield the elution step above was repeated.

2.1.5.2 Extraction from cell pellets

Cell pellets from peripheral blood and bone marrow samples were stored at -70ºC. Depending on the cell count the volume below was either doubled (>2x10^7 cells) or halved (<5x10^6 cells). A 1x10^7 pellet was thawed quickly by hand before the addition of 500µl resuspension buffer (RSB – see section 2.5 for components), 10µl 10% SDS and 3µl proteinase K (Qiagen, West Sussex). The cell pellet was then incubated at 37ºC overnight. On the second day, the pellet was first checked for proper digestion, if not digested properly a further 2µl of proteinase K was added and left for as long as
necessary. When properly digested, 150µl of 6M NaCl was added and the sample shaken vigorously by hand for 20 seconds. The digested cell pellet was then spun at room temperature in a Thermo IEC, Micromax centrifuge at 14,000rpm for 20 minutes. The supernatant was then removed to a clean sterile bijoux tube and 1.5 times the volume of the supernatant was added of 100% ethanol. The bijoux was inverted to obtain a ‘hairball’ of DNA, and hooked out using a sterile needle into a sterile eppendorf containing 70% ethanol. The size of the DNA ‘hairball’ was assessed and an appropriate volume (50-1000µl) of sterile Tris-EDTA buffer (Promega, UK) was aliquoted into a sterile 2ml screw cap tube, to this the washed ‘hairball’ of DNA was added and left to resuspend properly in the TE buffer before use.

2.1.5.3 DNA Extraction from Guanidinium Thiocyanate

This protocol uses the Qiagen DNA Mini Kit ‘Blood and Body Fluid spin protocol’ (Qiagen, West Sussex). To an equal volume of GTC, 200µl of Buffer AL was added and mixed by pulse-vortexing for 15 seconds. The sample was then incubated at 56ºC for 10 minutes and briefly centrifuged in a Thermo IEC, Micromax centrifuge to remove the drops from the inside lid. The mixture was then carefully applied to the QIAamp spin column (in a 2ml collection tube) without wetting the rim and the lid closed before centrifugation at 8000rpm for 1 minute. The collection tube containing the filtrate was discarded and the spin column placed in a new collection tube. 500µl of Buffer AW1 was added to the spin column and spun at 8000rpm for 1 minute, the filtrate discarded and 500µl of Buffer AW2 added to the column and spun at 14,000rpm for 3 minutes. To remove excess ethanol the filtrate was discarded and the column spun for a further minute at 14,000rpm. The spin column was then placed in a sterile microcentrifuge tube and 200µl Buffer AE added, incubated for 1 minute at room temperature and then spun at 8000rpm for 1 minute.
2.1.5.4 DNA Extraction from Mouthbrush samples

DNA obtained from mouthbrush samples was used as a control for the investigation of acquired mutations. In general, two mouthbrush specimens were received for each patient. To one mouthbrush tube 200µl of resuspension buffer (see section 2.5) was added. The lid of the other mouthbrush is snapped off and the mouthbrush added to the tube containing the first mouthbrush and 200µl of RSB. The tube was then vortexed for 30 seconds before using a sterile Pasteur pipette to remove the 200µl to a sterile 2ml tube. To this, 10µl of 10% SDS and 3µl of Proteinase K (Qiagen, West Sussex) were added, before incubation at 37°C overnight. The following day 150µl of 6M NaCl was added to the sample, vortexed to mix and then spun at 4°C at 14,000rpm for 20 minutes. The supernatant was then collected into a new sterile bijoux and 1.5 times the volume of 100% ethanol (Sigma Aldrich, UK) added. A hairball of DNA is rarely visible and so the sample was split between two sterile eppendorfs and frozen at -20°C for >2 hours. The sample was then spun again at 4°C at 14,000rpm for 20 minutes and each pellet resuspended in 25µl of 1 x Tris-EDTA buffer. These were left to resuspend for 1 hour at 37°C before combining in a labelled sterile 2ml screw cap tube.

2.1.5.5 DNA extraction from haematological stained slides

DNA extracted from haematological stained slides was used as a control for the investigation of acquired mutations and to investigate when the patient acquired the mutation. The dried material on the slide was first moistened with a drop of sterile PBS (Sigma Aldrich, UK). To a sterile eppendorf 200µl of sterile PBS was added. The cytological material was then scraped into this tube from the slide using a sterile scalpel blade. The resulting ‘sludge’ was mixed by pipetting up and down before adding 20µl of proteinase K (Qiagen, West Sussex) and 200µl of buffer AL and further mixing by pulse-vortexing for 15 seconds. The sample was then left overnight at 56°C to digest before continuing the following day with the Qiagen Blood and
Body Fluid Spin Protocol from step 5. As there is only a small amount of DNA to be obtained, a carrier DNA was always used. Qiagen recommend poly dA:dT carrier as this works best with the Qiagen columns, 10µg of polydA:dT (Sigma, Poole) was added after the overnight digestion and then continued as stated above.

2.1.5.6 GenomiPhi V2 DNA Amplification Kit

The Illuistra™ GenomiPhi™ V2 DNA Amplification Kit is part of the Phi29 DNA polymerase family of products (GE Healthcare, UK) for mini-scale whole genome DNA amplification. Our samples are often precious and whole genome amplification offers the possibility to greatly expand the amount of usable material. A typical DNA yield of 47 µg DNA can be achieved in less than two hours from an input of 1-10 ng genomic DNA with an average product length of >10 kb. The template must be at least 1ng/µl and requires an initial denaturation. On ice in thin-walled PCR tubes, 1µl (typically 10 ng) of template genomic DNA was combined with 9µl of sample buffer. The sample was then denatured at 95ºC for 3 minutes before cooling to 4ºC. Whilst the sample was denaturing a master mix combining 9µl of reaction buffer with 1µl of enzyme per sample was made up on ice. This was then added to the cooled denatured sample on ice prior to incubation at 30ºC for 90 minutes. The enzyme was then heat inactivated at 65ºC for 10 minutes before cooling to 4ºC. To test the quality of the genomic DNA, a control gene PCR was set up (See section 2.2.3).

2.2 Amplification and analysis

2.2.1 Reverse Transcriptase PCR (RT-PCR)

RT-PCRs were undertaken in a volume of 10-25µl. All amplifications were performed for between 30-32 cycles in a MJ Research Inc. thermocycler (single platform or Tetrad) in thin-walled PCR tubes. PCR mixes contained 0.5µM primers, 0.2µM each of dCTP, dATP, dGTP and dTTP, 1x Buffer (10mM Tris-Cl pH8.3, 50mM KCl, 0.1% gelatin), 1.75mM MgCl₂ and 0.05U Taq DNA polymerase
Typically multiplex PCRs were performed with 0.2µM each primer and MgCl₂ at a concentration of 2.5mM.

### 2.2.2 Microsatellite PCR

DNA samples were amplified with a series of fluorescently labelled primer pairs flanking highly polymorphic microsatellite markers on specific chromosomes of interest. All amplifications were performed in a MJ Research Inc. thermocycler (Tetrad) in thin walled 0.2ml 96-well plate format. The PCRs were undertaken in 10µl volumes with 25-50ng of genomic DNA. All primers were designed to anneal at 55°C and standard cycling parameters were used with 32 amplification cycles. The products were analysed on an ABI 3100 genetic analyser using the program Genotyper 2.0. The peak heights were compared to controls and scored as heterozygous for each marker if two peaks of similar intensity were clearly visible, or if the ratio of the two peaks was similar to normal controls. If only one peak was visible a patient was scored as homozygous. To allow for the presence of background normal cells, acquired homozygosity was also scored if one peak was one third or smaller than the expected size compared to normal controls.

### 2.2.3 Control gene PCR

The control gene PCR is a multiplex PCR capable of amplifying 5 different sized products (100bp, 200bp, 300bp, 400bp, 500bp & 600bp). This PCR is used to check the quality of genomic DNA, especially after GenomiPhi amplification. PCRs were undertaken in a volume of 50µl. All amplifications were performed for 35 cycles in a MJ Research Inc. thermocycler (Single platform or Tetrad) in thin-walled PCR tubes. The primers were combined to make two mixes: Mix A (25µl of 100bp.F, 100bp.R, 200bp.F, 200bp.R, 300bp.F, 300bp.R, 400bp.F & 400bp.R) and Mix B (50µl 600bp.F & 600bp.R) (van Dongen, et al 2003). PCR mixes contained 0.5µM Mix A primers, 1µM Mix B primers, 0.5µM of dNTPs, 1x Buffer, 2mM MgCl₂ and 1U AmpliTaq.
Gold DNA polymerase (Applied Biosystems, UK). The PCR was cycled as follows: 94°C for 7 minutes hotstart; 94°C for 1 minute; 60°C for 1 minute; 72°C for 1 minute for 35 cycles and an extension of 72°C for 7 minutes. The products were run on a 2% agarose gel to allow separation of the bands.

### 2.2.4 Powerplex® 16 System

The Powerplex® 16 kit (Promega, UK) allows the co-amplification and three colour detection of sixteen loci; fifteen short tandem repeat loci plus Amelogenin for sex determination. The system encompasses the following loci: Penta E & D, D18S51, D21S11, THO1, D3S1358, FGA, TPOX, D8S1179, vWA, CSF1PO, D16S539, D7S820, D13S317, D5S818 and Amelogenin as described in detail at [http://docs.appliedbiosystems.com/search.taf](http://docs.appliedbiosystems.com/search.taf). All sixteen loci are co-amplified in a single tube and analysed in a single injection on an Applied Biosystems 3100. This kit was designed to allow a high discriminatory power for cases of identity and paternity. I used this kit as a method of confirming that DNA samples at different time points were indeed from the same individual.

The Powerplex® 16 PCR was performed using the following mix per sample; 1.25µl Gold ST*R 10x Buffer, 1.25µl 10x Primer pair mix, 0.4µl AmpliTag Gold® DNA polymerase (Applied Biosystems, UK), 8.3µl sterile water. To 11.2µl of master mix 1.2µl of DNA at 0.9ng/µl was added. PCR was performed according to the following amplification protocol: 95°C for 11 minutes, 96°C for 1 minute, 94°C for 30 seconds, ramp 68 seconds to 60°C (hold for 30 seconds, ramp 50 seconds to 70°C (hold for 45 seconds), cycle the last two steps 10 times, 90°C for 30 seconds, ramp 60 seconds to 60°C (hold for 30 seconds), ramp 50 seconds to 70 °C (hold for 45 seconds), and cycle the last two steps 20 times, final extension at 60°C for 30 minutes and a 4°C soak.

After PCR the products were prepared for the ABI 3100. An additional lane was incorporated into the plate to include the Powerplex® allelic ladder which contains
the most common alleles for each locus. A master mix was made up of 9µl highly deionised formamide and 0.5µl Promega ILS600 size standard. The master mix was aliquoted into a 96 well plate (specific to the ABI 3100 machine) and 0.5 µl of each sample was then added. The samples were run under specific ABI 3100 settings for 45 minutes. Genotypes were then assigned by comparing the sizes obtained for the unknown samples with the sizes obtained for the alleles in the allelic ladder. Different samples can then be compared against each other based on allele specificity at each locus.

2.2.5 Long-Range PCR

Amplification of products larger than 2.5kb was performed using long-range PCR. The High Fidelity PCR Master kit (Roche Diagnostics, Lewes, UK) was used to amplify genomic DNA up to 10kb in length. All primers were designed with a melting temperature of 66°C with approximately 50% GC content. Typically 100ng of DNA was amplified in the PCR master kit with primers at a final concentration of 0.5µM. PCR was performed according to the following amplification protocol, allowing increased elongation times for the length of product: 94°C for 20 seconds, 66°C for 4 minutes, 68°C for 12 minutes, cycled 10 times, then 94°C for 20 seconds, 66°C for 40 seconds and 68°C for 14 minutes, cycled 22 times, final extension at 72°C for 10 minutes.

2.2.6 Bubble-PCR

Bubble PCR (Zhang, et al 1995) is a one sided PCR technique capable of amplifying an unknown sequence that lies adjacent to a known sequence. The bubble oligonucleotide is composed of two strands that are complementary at both ends but non-complimentary in the middle. For all primer sequences see appendix I, section I.
Sample DNA was first digested using two separate restriction enzymes (HaeIII and RsaI) designed to blunt end cut around the known sequence and regularly throughout the rest of the genome. The bubble oligonucleotide was then blunt end ligated to the digested genomic DNA. The ligated DNA was then denatured at 72°C for 5 minutes before removal of unligated bubble oligonucleotides using the QiaQuick PCR purification kit (Qiagen, West Sussex). A sample specific primer, in this case PDAI12-R3, was annealed to 1µl of the ligated DNA and extended using Amplitaq Gold (Applied Biosystems, UK) to the end of the bubble oligonucleotide (if there was no bubble sequence the rest of the genome to which bubbles had been ligated would also be amplified with the bubble primer). A single step Hotstart PCR using the same PDAI12-R3 primer and a forward NVAMP-1 primer, complementary to the bubble oligonucleotide, was then performed to amplify this strand. This results in a copy of the known, unknown (if present) and bubble sequence. The copied, non-complementary part of the bubble sequence is only present if the specific DNA was amplified. This strand was then subsequently amplified using Hotstart PCR with a bubble specific primer. A nested PCR was then performed to increase the specificity of the reaction using NVAMP-2 and PDAI12-R4. Rearranged bands were excised and gel purified with QiaQuick Gel Purification kit (Qiagen) and either sequenced directly or cloned and sequenced.

### 2.2.7 Gel Electrophoresis

The protocol was altered accordingly dependent on the size of the gel former and the percentage gel required. For a 1% agarose gel in a gel tank with a capacity of 100ml, 1g agarose (Bioline, UK) was added to 100ml 1x Tris-Borate EDTA (TBE) buffer (Sigma Aldrich, UK). This was then heated using a microwave to melt the agarose, before cooling slightly and adding 0.2mg/ml ethidium bromide to visualise the DNA. The melted gel was then poured into a gel former with combs and allowed to set for at least 20 minutes. Ensuring the gel was set the combs and formers were then removed and the gel and electrodes covered by 1 x TBE buffer. To 5µl of sample or PCR product 1-2µl of Orange-G sucrose loading buffer was mixed before loading the
whole amount into a well of the gel. One lane was always loaded with the 1kb plus DNA ladder (Invitrogen, UK) to enable sizing of products. The samples were run for 15 minutes or longer depending on the expected product size and percentage gel before visualising the bands under UV light.

### 2.2.8 Denaturing high-performance liquid chromatography (dHPLC)

Amplification of each exon of each gene to be analysed was performed from approximately 50-100ng of patient genomic DNA. A high fidelity DNA Taq polymerase was used to ensure that PCR did not introduce any sequence errors. PCRs were undertaken in a volume of 75μl. All amplifications were performed for 35 cycles in a MJ Research Inc. thermocycler (Tetrad) in thin-walled 12-strip PCR tubes. PCR mixes contained 0.5μM primers, 0.2μM each of dCTP, dATP, dGTP and dTTP, 1x GeneAmp PCR Buffer II (100mM Tris-HCl, pH8.3, 500mM KCl), 1.5mM MgCl$_2$ and 0.05U AmpliTaq Gold DNA polymerase (Applied Biosystems, Warrington, UK). The primers were designed at least 50 base pairs away from each end of the exon, with a maximum size of 500 base pairs for each amplicon. In some cases exons very close together were co-amplified with the intervening intron present. The primer sequences and the PCR conditions are in the appendix I, sections II & III.

Following amplification, the PCR product from each patient for each exon was mixed equally with amplified product from the same exon from human placental DNA (Sigma-Aldrich, Poole, UK). The mixed products were then denatured at 95°C for 5 minutes and then slowly reannealed to each other by reducing the temperature by 1.5°C every 22 seconds to room temperature allowing stable formation of heteroduplexes. Conditions for dHPLC analysis were calculated using the manufacturer’s software WAVEmaker v4.1. The computer predicts the temperature(s) over which the DNA fragment will be between 50 and 100% reannealed. It is estimated that more than 95% of heteroduplexes are detected. For longer products or those that have a high GC content, there may be several temperatures required to profile the whole fragment. The temperatures required for
heteroduplex analysis of each amplicon are given in the appendix, section III. Before each dHPLC run a plasmid DNA control supplied by the manufacturer (WAVE mutation standard; Transgenomics, Crewe, UK), with a known melting profile was tested to ensure consistency of results between runs. For each exon normal placental DNA PCR products were run as a normal control, again to ensure consistency of results for each exon between experiments. All results were analysed using the WAVEmaker v4.1 software. Any heteroduplexes observed, indicated by shifts in the traces or split peaks, were subsequently re-amplified and sequenced to look for the presence of mutations.

2.2.9 Multiplex Ligation-dependent Probe Amplification (MLPA)

MLPA is a method to establish the copy number of up to 45 nucleic acid sequences in one single reaction (Schouten, et al 2002). It uses genomic DNA to detect amongst other things copy number. Deletions and amplifications of a gene (or part of) will usually not be detected by sequence analysis of PCR amplified gene fragments as a normal copy is still present. Amplification products are separated by sequence type electrophoresis. As only one pair of PCR primers is used, MLPA reactions result in a very reproducible gel pattern with fragments ranging from 130 to 490 bp. Comparison of this gel pattern to that obtained with a control sample indicates which sequences show an aberrant copy number.

1µl (50ng) of DNA was added to 4µl of Tris-EDTA buffer in a 0.2ml PCR tube and overlaid with oil (this prevents evaporation overnight) before running on a MJ thermocycler tetrad ‘Denature’ programme (98°C 5 minutes; 25°C Hold). Whilst the DNA is denaturing the following mix was made up. For each sample, to 1.5µl relevant probe mix, 1.5µl MLPA buffer was added and mixed gently. 3µl of this mix was added to the denatured sample and the probes bound overnight at 60°C (95°C 1 minute; 60°C Hold). Following the overnight binding of the probes, on ice 3µl of ligase buffer A and B, 25µl water and 1 unit ligase were combined for each sample and mixed. The 60°C overnight run was stopped and the samples held at 54°C for at
least 2 minutes before adding 32µl of the above mix to each sample and running the
‘Ligate’ programme (54°C 15 minutes; 98°C 5 minutes; 4°C Hold). Whilst the probes
were ligating, the PCR mix was made up. For each sample 2µl SALSA buffer (MRC Holland, Amsterdam), 1µl SALSA primers, 1µl Enzyme buffer, 15.75µl water and
0.25µl (2.5 units) SALSA polymerase were mixed carefully and then aliquoted into
new 0.2ml PCR tubes on ice. To this PCR mix 5µl of the ligation mix was added and run on a ‘Cycles’ programme on a tetrad (95°C 30 seconds; 60°C 30 seconds; 72°C 1
minute; cycle 33 times; 72°C 20 minutes; 4°C hold). After PCR the samples were
analysed on an ABI 3100 by mixing 1µl PCR product to 9µl of formamide and 0.1µl
of ROX500 size standard (Applied Biosystems, UK).

Analysis of the samples was performed using Applied Biosystems Genescan™
software using a GS500 size standard. Following this Applied Biosystems
Genotyper™ software was used. Based on the number of probes in the MLPA kit the
software can be manipulated. The number of peaks was specified in addition to
selecting ‘highest’ peaks, ensuring only the correct number of highest peaks, within
the size range, are selected. All peaks were labelled with the peak height before
exporting the data into a table. MLPA analysis sheets for each MLPA probe kit have
been set up in Microsoft excel. The data from the genotyper table was imported into
the excel sheet and samples were compared to controls. The peak height of each
probe is compared against every other probe and against the control. This gives a data
sheet of numbers where 1 is normal, a horizontal row of 0.5 indicates a deletion and
horizontal row of 1.5 a duplication.

2.2.10 Exo-SAP and sequencing

Exonuclease I and shrimp alkaline phosphatase (SAP), two hydrolytic enzymes, are
combined at a ratio of 1:4 and used for the preparation of PCR products for
fluorescent-based sequencing. Exonuclease I (New England Biolabs, UK) degrades
any residual single stranded primers and extraneous single-stranded DNA, whilst
SAP (Roche Diagnostics Ltd, UK) hydrolyses the remaining dNTPs from the PCR
reaction. On ice, 1μl of exonuclease I/SAP mix was added to 2μl of PCR product and incubated at 37°C for 15 minutes, followed by 65°C for 15 minutes to denature the enzymes. The product from this reaction was diluted 10 fold and then used directly for sequencing.

2.2.10.1  Sequencing of PCR products

Exonuclease I/SAP purified PCR products were sequenced using a final concentration of 1μM of a single primer (either forward or reverse) and the BigDye terminator kit v3.1 or v1.1 (Applied Biosystems, Foster City CA, USA). Dependent on numbers of samples to be sequenced, reactions were either cleaned up using ABgene® Dye Terminator Removal Kit (Abgene Limited, UK) and dried using a MJ Research Inc. thermocycler (Tetrad), or by using the Montage SEQ96 Cleanup Kit (Millipore, Watford, UK). The first is a gel-based separation matrix and the latter a membrane-based size-exclusion filtration system and requires final resuspension in 15μl of washing buffer. DNA from both methods was resuspended in 10μl deionised formamide and run on an ABI 3100 or ABI3130xl sequencer.

2.2.10.2  Sequencing following cloning

Plasmid DNA was sequenced using 1μl of mini-prepped DNA (approximately 100-500ng) for each PCR (See section 2.2.11.3). Sequencing was performed using the forward and reverse M13 vector primers for each plasmid DNA (two reactions). The single stranded inserts were amplified with BigDye terminator kit v3.1 or v1.1 (Applied Biosystems, Foster City CA, USA), cleaned up using the Montage SEQ96 Cleanup Kit (Millipore, Watford, UK) and resuspended in 15μl wash buffer. To the resuspended DNA 10μl deionised formamide was added and the reactions run on an ABI 3100 or ABI3130xl sequencer.
2.2.11 Cloning PCR products

PCR products were cloned by ligation into a plasmid vector using the TA or TOP10 cloning kits according to the manufacturer’s instructions (Invitrogen, Paisley, UK).

2.2.11.1 Ligations

On ice, 1-2µl PCR product (depending on the intensity of the product observed on an agarose gel), 1µl salt solution (1.2M NaCl, 0.06M MgCl$_2$), 1µl pCR$^\text{®}$4 vector (10ng/µl), and sterile water to a final volume of 5µl were combined. The reaction was then mixed gently, and incubated at room temperature for 5 minutes. This reaction relies on *Taq* polymerase nontemplate-dependent terminal transferase activity which adds a single deoxyadenosine to the 3’end of PCR products. The vector supplied in the kit is linearised and has complementary single, overhanging 3’ deoxythymidine residues allowing PCR inserts to ligate efficiently. The ligation itself relies on the energy conserved by formation of a covalent bond between the 3’phosphate of the cleaved strand and the tyrosyl residue of the enzyme topoisomerase I which binds to and cleaves the phosphodiester backbone in certain regions of DNA.

2.2.11.2 Transformations

Chemically competent TOP10 cells provided with the Invitrogen kit were slowly thawed on ice. 2µl of the ligation reaction was added and stirred gently with a pipette tip to mix. The cells were then incubated on ice for 30 minutes. The remaining ligations were stored at -70°C. After 20 - 30 minutes, the cells were heat-shocked at 42°C for 30 seconds and immediately placed on ice for at least 2 minutes. 250µl of SOC medium (2% Tryptone, 0.5% yeast extract, 10mM NaCl, 2.5mM KCl, 10mM MgCl$_2$, 10mM MgSO$_4$, 20mM glucose) was then added to each vial, and the cells were then shaken horizontally at 37°C for 1 hour. LB/agar plates with the appropriate antibiotic added were spread with 25µl of 50mg/ml X-Gal (Promega, Madison, USA)
and allowed to dry. Varying amounts of each transformation mix were spread on individual plates. Once the liquid was absorbed the plates were inverted and placed in a 37°C incubator overnight. After incubation the plates were placed at 4°C to allow for colour development. White colonies were picked for rapid screening of inserts by PCR using the following miniprep protocol.

2.2.11.3  *Qiagen™ Miniprep kit*

Picked colonies were grown up overnight in single 10ml culture tubes with 2ml L-Broth (LB) and the appropriate antibiotic. The following day cells from 1.5ml culture medium were extracted using the *Qiagen™ Miniprep kit* (Qiagen Ltd, West Sussex) and the remaining 500µl is stored in LB/Glycerol (3:1) solution at -70°C. The 1.5ml of culture medium was spun down in an eppendorf at 10,000rpm for 10 minutes in a Thermo IEC, microcentrifuge. The supernatant was then carefully removed leaving only the pellet which was resuspended in buffer P1, ensuring no clumps remained before adding the P2 lysis buffer. The tubes were then inverted gently 4-6 times to ensure complete lysis before addition of the neutralisation buffer P3. On addition of buffer P3 each tube was immediately inverted 4-6 times; the solution became cloudy and white as the denatured proteins, cellular debris and chromosomal DNA precipitate out of solution. The tubes were then spun at 14,000rpm for 10 minutes, in a Thermo IEC microcentrifuge, and the supernatant carefully transferred to an appropriately labelled QIAprep spin column. The column was then spun at 14,000rpm for 1 minute allowing the DNA to bind to the column, and then washed twice for 1 minute 14,000rpm using firstly 500µl wash PB, followed by 750µl wash PE. The column was then spun for an additional 5 minutes at 14,000rpm to remove all residual PE buffer. To collect the DNA, 25µl of distilled water was added to the centre of the column, now in a new sterile labelled eppendorf, and spun for a further 1 minute at 14,000rpm.
2.2.11.4 **Digests and PCR**

5μl of each mini-prepped plasmid DNA was digested with 10 units EcoRI, (EcoRI sites flank the cloning site in pCR2.1 or pCR 4.0 vector), 2μl 5x EcoRI buffer and the appropriate amount of sterile water up to 10μl. Digests were performed at 37°C for 2-4 hours. The digested clones were run on a 1% agarose gel to determine the insert size and cloning efficiency. The plasmid DNA was then sequenced according to the protocol in section 2.2.10.2.

2.2.12 **5’ Rapid amplification of cDNA ends (5’RACE)**

Fusion genes were identified using the 5’RACE GeneRacer™Kit (Invitrogen, Paisley, UK). In brief, after RNA extraction using the Qiagen RNeasy kit (Qiagen Ltd, West Sussex, UK), approximately 5μg of total RNA was dephosphorylated, decapped and ligated to the GeneRacer™ RNA oligonucleotide (Figure 2.1). The ligated RNA was reverse transcribed using Superscript II™ reverse transcriptase (Invitrogen, Paisley, UK) and random primers. The initial 5’RACE PCR was performed using the 5’GeneRacer™ primer and a reverse primer from the relevant exon of the tyrosine kinase gene in question, using the High Fidelity PCR Master kit (Roche Diagnostics, Lewes, UK). The second step, nested PCR was performed using the 5’GeneRacer™ nested primer and a nested primer designed in the tyrosine kinase gene. After nested PCR, if a product was visible by gel electrophoresis, the product was cloned using the TOPO TA cloning kit for Sequencing (Invitrogen, Paisley, UK), and sequenced.
Figure 2.1: A schematic representation of the process of 5′ Rapid Amplification of cDNA Ends (5′RACE). (1) Calf intestinal phosphatase (CIP) is first used for dephosphorylation of truncated or non-mRNA. Full-length mRNA is protected from dephosphorylation due to its 7-methylguanosine cap structure. The truncated and non-mRNAs are thus eliminated; (2) Tobacco alkaline phosphatase (TAP) is then used to remove the cap structure of the full-length mRNA; (3) RNA ligase is then employed to ligate an RNA oligonucleotide of known sequence to the 5′ end of the full-length mRNA. Random primers are then used to reverse transcribe the ligated full-length RNA; (4) 5′RACE PCR is then performed using an oligonucleotide complementary to the ligated sequence combined with a gene-specific primer; (5) The product obtained can then be cloned and sequenced.
2.2.13 Pyrosequencing

Pyrosequencing™ is a technique used for accurate and quantitative analysis of DNA sequences that is based on the detection of released pyrophosphate (PPI) during DNA synthesis. A cascade of enzymatic reactions occurs resulting in the production of visible light that is proportional to the number of nucleotides incorporated. The enzymatic cascade begins with a nucleic acid polymerisation reaction as a result of nucleotide incorporation by polymerase, inorganic PPI is released. The released PPI is subsequently converted to ATP by ATP sulphurylase. ATP is the energy required by luciferase to oxidize luciferin and generate light. The amount of light is then detected by a photodiode, photomultiplier tube, or a charge-coupled device camera (CCD) camera (Figure 2.2).

**Figure 2.2:** Schematic representation of the progress of the enzyme reaction in pyrosequencing (Ronaghi 2001). A primed DNA template and four enzymes are placed in a well of a microtiter plate. The four different nucleotides are added stepwise to the reaction and incorporation is followed using the enzyme ATP sulphurylase and luciferase. The nucleotides are continuously degraded by nucleotide-degrading enzyme apyrase,
allowing addition of the subsequent nucleotide. \( \text{dXMP} \) indicates one of the four nucleotides.

Pyrosequencing was performed as follows: the amplicon in question was first generated in a 50 µL reaction volume containing 10 pmol each of the forward and reverse PCR primers, 0.2 mM deoxynucleoside triphosphates (Amersham), 2.5 mM \( \text{MgCl}_2 \), 1x Buffer II (Applied Biosystems), 1 U of AmpliTaq Gold (Applied Biosystems), and 1 µL (~20 ng) of DNA. PCR was performed in a Tetrad with the following conditions; 94 °C for 7 min, followed by 45 cycles of 94 °C for 30 s, 58 °C for 30 s, and 72 °C for 30 s; 1 cycle at 72 °C for 7 min; and a final hold at 15 °C. All samples were analysed in duplicate. Pyrosequencing reactions were performed according to the manufacturer’s instructions, using the Gold PSQ 96 SNP Reagent Kit (Biotage AB). The sequence was read and peaks heights quantified by the AQ Software.

2.3 Fluorescence in situ hybridisation (FISH)

2.3.1 Maxiprep DNA isolation from BAC clones

BACs were initially ordered from the Sanger Institute, Hinxton. The clones arrived as LB/Agar stabs supplemented with the appropriate antibiotic. BACs were also obtained from the 37K clone set which was subsequently acquired in house. BACs were colony purified before use by streaking out using a sterile loop onto a LB/Agar plate containing the relevant antibiotic. After overnight incubation, a single colony was picked and used to inoculate 5mls L-Broth, again containing the appropriate antibiotic. The cultures were incubated overnight in a shaking incubator (225rpm) at 37°C.

The starter culture was then used to inoculate 30ml of L-Broth medium with antibiotic in a 50ml fluted flask (Triple Red, Buckinghamshire, UK) and incubated as before. To harvest the bacteria, the overnight culture was transferred to a 50ml falcon
tube and spun at 3,000rpm in a RC5C Sorvall centrifuge. After centrifugation the supernatant was completely removed and the cell pellet was resuspended in 1.8ml of Solution I (see section 2.5 for all solution details). The resuspended pellet was then transferred to a plastic centrifuge tube. The cell suspension was then lysed by adding 1.8ml of freshly prepared Solution II and left to stand for 5 minutes at room temperature. Bacterial proteins and DNA were then precipitated out by adding 1.8ml of Solution III and leaving the tubes on ice for 5 minutes. This was then centrifuged at 10,000rpm for 10 minutes at 4°C in the RC5C Sorvall centrifuge (SS-34 rotor). Leaving the cell debris, the supernatant was then carefully removed by pipette, and aliquoted into a number of sterile eppendorf tubes containing 700µl chloroform (Sigma-Aldrich, Haverhill). After shaking to mix the chloroform-vector DNA mixture, they were spun at 13,000rpm for 5 minutes in a Thermo IEC, Micromax centrifuge. The supernatant was then removed and transferred to pre-prepared eppendorfs containing 800µl of ice-cold isopropanol (Sigma-Aldrich, Haverhill), inverted to mix and then placed at -20°C overnight to maximise DNA precipitation.

The following day, the tubes were removed from the freezer and centrifuged at 14,000rpm for 15 minutes in a Thermo IEC, Micromax centrifuge. The supernatant was then poured off and 500µl of 70% ethanol (BDH, Dorset) was added, the pellet was washed by inversion before centrifuging again at 14,000rpm for 5 minutes. The supernatant was then carefully removed by pipette, and the pellet allowed to air dry until it become translucent. Each pellet was then resuspended in 40µl of sterile water.

2.3.2 Labelling

1µg BAC DNA was labelled by nick-translation in the presence of 10U DNA polymerase I (Invitrogen, Paisley, UK) and 0.05U DNase 1 (Invitrogen, Paisley, UK). The labelling reaction was performed at 16°C for 1.75 hours using a standard labelling mix (see section 2.5). The reaction was stopped by denaturing the enzymes at 68°C for 10 minutes. 10µl of the product was then run on a 1% agarose gel to check the efficiency of the labelling reaction.
2.3.3  Slide making

Fixed cell material (approximately $5 \times 10^6$ cells) was spun down and washed with 3:1 methanol:acetic acid fixative and finally resuspended in 1-2mls of fixative. In a humidity-controlled room, one drop of suspension was dropped onto a microscope slide from a short height and the slide was allowed to dry. The slides were checked for spreading of nuclei, under phase contrast light microscopy. The slides were left at room temperature overnight, or exposed to UV for 10 seconds to age them before use.

2.3.4  Probe preparation

Labelled DNA was co-precipitated using an equal volume of $1.05\text{mg/ml COT-1 DNA}$ (EurX, Poland), one tenth the volume of $3\text{M sodium acetate pH5.2}$ and twice the total volume of $100\%$ ethanol and then frozen for 1 hour at $-20\text{°C}$. The labelled DNA was then spun down at 14,000rpm for 20 minutes, washed several times in $70\%$ ethanol, and resuspended in $12\mu\text{l hybridisation buffer}$ (see section 2.5).

2.3.5  Slide preparation

Slides were incubated for 2 hours at $37\text{°C}$ in $2 \times \text{SSC pH 7.0}$ (for all solution details see section 2.5), then treated with $100\text{ng/ml pepsin in 0.01M HCl}$ for 10 minutes. Following the pepsin digestion the slides were washed three times with phosphate buffered saline (PBS), before being fixed in a $1\%$ paraformaldehyde PBS for 10 minutes. The PBS wash was repeated (as described above), and then the slides were dehydrated through an ethanol series of 2 minutes in $70\%$ ethanol, 2 minutes $90\%$ ethanol and 2 minutes $100\%$ ethanol. The slides were then dried. The previously prepared probe was then aliquoted onto a cover slip, placed on the dried slide and sealed with vulcanising rubber solution (Weldtite, Barton-on-Humber). When the
rubber solution had dried the slides were denatured on a 75°C hotplate for 5 minutes, and hybridised in a moist chamber for 48 hours at 37°C.

2.3.6 Post hybridisation and visualisation

Following hybridisation the cover slips were removed and the slides washed briefly in 2 x SSC. Washes were then performed at 42°C: 3 x 5 minutes washes in 50% formamide/2 x SSC pH7.0, and 3 x 5 minutes washes in 2 x SSC. Post-hybridisation biotin and digoxygenin signals were detected using the following combined antibody layers: 1) 1/100 sheep anti-digoxigenin (1mg/ml, Roche Diagnostics, Germany), 2) 1/400 texas red avidin (2mg/ml, Vector Laboratories Inc, California, USA) and 1/200 rabbit anti-sheep FITC (1mg/ml, Vector Laboratories Inc, California, USA), 3) 1/100 goat anti-avidin biotinylated (1mg/ml, Vector Laboratories Inc, California, USA), and swine anti-rabbit FITC (1.1mg/ml, Dako, Denmark) and 4) 1/400 texas red avidin (Lichter and Cremer 1992). All dilutions were made with SSCTB (see section 2.5 for solution composition). The slides were incubated for 30 minutes to 1 hour at room temperature in a dark moist chamber, and then washed 3 x 5 minutes in SSCT (4 x SSC, 0.1% Triton-X, pH7.0) between each incubation. After application of all antibody layers the slides were rinsed with PBS and allowed to dry. The slides were then mounted, under a coverslip, with Vectorshield (Vector Laboratories, California, USA) containing 1µg/ml 4,6-diamidino-2-phenylindole (DAPI). Slides were examined using an Axioskop microscope (Zeiss, Germany). The images were captured using a CCD camera and image analysis software (MacProbe v4.3, Applied Imaging International, UK).
2.4  Tissue culture

2.4.1  Imatinib sensitivity assay

2.4.1.1  Thawing of primary haematopoietic cells

Peripheral blood mononuclear cells (MNCs) were separated using lymphoprep (Axis- Shield, Oslo, Norway) and grown either directly or following cryopreservation in methylcellulose supplemented with the recombinant cytokines; SCF (20ng/ml), GM-CSF (1ng/ml) and IL-3 (5ng/ml)+ (Stem Cell Technologies Ltd., Vancouver, British Columbia, Canada). Frozen vials typically contained between $2 \times 10^7$ and $6 \times 10^7$ cryopreserved cells. The cryopreserved MNCs were thawed quickly by incubating at $37^\circ C$, before transferring the cells to a 50ml tube. Dropwise, 20ml of pre-warmed thawing solution (section 2.5) was then added. The tube was topped up to 50ml with RPMI, supplemented with 10% foetal calf serum (FCS), L-glutamine and penicillin-streptomycin (as before), and then spun at 1,500rpm for 5 minutes. After centrifugation, the supernatant was discarded and 200,000U of DNase I was added directly to the cell pellet. The tube was gently shaken to disrupt the pellet and to prevent the formation of clumps. To the pellet, 10ml supplemented RPMI was then added and the cells centrifuged, as before. The DNaseI and RPMI wash was then repeated before finally discarding the supernatant and resuspending the cells in 10ml RPMI. To the resuspended cells, 200,000U DNaseI was added and the cells were left for one hour at room temperature. The cells were finally spun and resuspended in RPMI for counting and viability. The cells were counted using an equal volume of trypan blue (Sigma), to differentiate between live and dead cells.

2.4.1.2  Colony assay

A cell count of between $3 \times 10^5$ and $2 \times 10^6$ can be used to set up the in vitro assay that measures the effect of imatinib on colony growth (Crescenzi, et al 2007). After counting, the cells were spun down and resuspended in 1ml of RPMI supplemented
with 10% foetal calf serum (FCS), L-glutamine and penicillin-streptomycin (as before). The assay was set up using zero, 1μM or 5μM imatinib mesylate (Novartis, Basel, Switzerland), in triplicate. To each 3ml aliquot of methylcellulose (one for each concentration of imatinib) an equal volume of cell suspension was added to the top. The amount of imatinib required for each concentration was then calculated from a 1mM stock solution and also added to the top of the methylcellulose aliquot. After the addition of the cells and the imatinib, the methylcellulose was mixed thoroughly using a 1ml syringe.

From each aliquot of methylcellulose containing equal cell numbers, 3 x 1ml aliquots were dispensed into 3 separate 40 x 12mm tissue culture dishes (Fisher Scientific, Leicestershire). All the 40 x 12mm tissue culture dishes were then separated into 2 large tissue culture trays (144 x 21mm), ensuring at least one of each concentration were in separate trays in case of infection. A small 40 x 12mm dish was placed in the middle of the large tray with sterile water in it to prevent the methylcellulose drying out. The dishes were then left in a 37ºC incubator to grow for 7 days before the number of colonies was counted. The assay was counted again on day 14. An index of response was then calculated, as the mean response at 1 and 5µM at days 7 and 14 compared to the untreated control. This allows quantitative comparison of the response of different patients. For comparison, normal controls had been previously tested by Dr Andy Chase and the average response index was found to be 0.4. A number of CML patients, acting as positive controls were also tested by Dr Chase, these showed a much greater inhibition with a response index of 0.1. On the basis of these controls an index of below 0.2 was used to define a positive response as previously described (Crescenzi, et al 2007).

2.4.2 Midi prep & nucleotide removal kit

2.4.2.1 Midi-prep

The PureYield™ Plasmid Midiprep System (Promega, UK) is used for the isolation of large quantities of high-quality plasmid DNA that can be specifically used in
eukaryotic transfection and in vitro expression experiments. It contains an endotoxin removal wash that removes substantial amounts of protein, RNA and endotoxin contaminants from purified plasmid DNA.

A subculture was grown up overnight in 5ml L-Broth plus the relevant antibiotic before spiking a larger 25 or 45ml volume of L-broth plus antibiotic and leaving this to grow again overnight. In brief, preparation and lysis of the bacterial cell cultures was performed by pelleting the cells at 10,000 x g for 10 minutes in a Sorvall centrifuge. The supernatant was then discarded and the cells resuspended in 3ml Cell Resuspension Solution. To this 3ml of Cell Lysis Solution is added and mixed by inversion 3-5 times before leaving at room temperature for 3 minutes to incubate. Following this 5ml Neutralisation Solution was added to the lysed cells, mixed by inversion 3-5 times before allowing the lysate to settle in an upright position for 2-3 minutes. The DNA was then purified. The lysate was poured into a PureYield™ clearing column in a clean 50ml falcon (Sarstedt, UK) incubated for 2 minutes to allow the cellular debris to rise to the surface before centrifuging at 3,000 x g for 5 minutes (repeating if the not all the lysate has passed through the column). The filtered lysate was then poured into a PureYield™ binding column in a clean 50ml falcon and centrifuged for 3 minutes at 3,000 x g. The column-bound DNA was first washed with 5ml endotoxin removal wash solution and centrifuged as above, the supernatant was discarded and the column washed with 20ml column wash solution for 5 minutes at 3,000 x g. To ensure complete removal of ethanol the supernatant was discarded and the column spun for an additional 10 minutes at 3,000 x g. Elution of the DNA was performed by placing the binding column in a new falcon and adding 300μl nuclease-free water, centrifuging at 1,500 x g for 5 minutes and then transferring the filtrate to a 1.5ml tube.
2.4.2.2  *QIA*quick nucleotide removal kit

To ensure the midi-prepped DNA is of high enough quality for transfection protocols it is further cleaned up using a QIAquick nucleotide removal kit (Qiagen, West Sussex) according to the manufacturer’s instructions for use with a microcentrifuge.

2.4.2.3  Transfection by electroporation

Transfection is a method of introducing nucleic acids into eukaryotic cells by non-viral means. There are various transfection protocols published, all of which are based on neutralising or obviating the issue of introducing negatively charged molecules such as DNA into cells with negatively charged membranes. Chemicals such as calcium phosphate or cationic lipid-based reagents are used to coat the DNA resulting in neutralising or creating an overall positive charge improving the ability of the DNA to cross the membrane. Physical methods, such as electroporation, are based on simply punching a hole in the membrane allowing the DNA to be introduced directly to the cytoplasm. Electroporation was first used for gene transfer studies into mouse cells (Wong and Neumann 1982). It works by creating a transient pore that allows the DNA to be introduced into the cell by disturbing the cell membrane using an electrical pulse (Shigekawa and Dower 1988). Electroporation often requires more cells than chemical methods due to substantial cell death, thus initial optimisation is critical to ensure good transfection efficiency against cell viability.

The murine haemopoietic cell line Ba/F3 was used to transfect in the newly identified fusions involving the tyrosine kinase gene PDGFRβ. The Ba/F3 mouse pro B cells are dependent on IL-3 for growth and have an extremely low rate of spontaneous reversion to factor independence (Palacios and Steinmetz 1985). In addition, Ba/F3 cells express no detectable endogenous PDGFRβ (Drummond-Barbosa, *et al* 1995, Petti, *et al* 1997).
2.4.2.4  Electroporation

The day prior to electroporation, the Ba/F3 cells were split. The following day the cells were washed twice with sterile PBS and then counted using a Neubauer type haemocytometer (BDH, Poole, UK). The cells were then resuspended in pre-warmed RPMI at a concentration of $1 \times 10^7$ cell per ml. 400µl of these cells ($4 \times 10^6$) was then put into a 0.4cm Gene-pulser Cuvette (BioRad, Hertfordshire) and 16µg of the DNA to be integrated with the cells added. A mock transfection of the cells with no DNA and a vector only transfection (pcDNA 3.1) were carried out in the same way. The Gene Pulser was set to 140kV (Programme 10, mammalian, HL60) and the cuvette inserted into the electroporating pod (only one possible orientation). The cells are pulsed once by pressing the red button on the front of the Gene-Pulser machine before placing the cells in the cuvette on ice for 10 minutes. After this time the cells were transferred into a sterile 25cm$^2$ flask (Fisher Scientific, Leicestershire) containing 10mls of RPMI 1640 medium with 10% FCS and 10% WEHI (supernatant for IL-3) and left for 2-3 days to allow the cells to recover.

2.4.2.5  Antibiotic selection

After the recovery period the cells were plated out at two different concentrations ($2 \times 10^3$ and $5 \times 10^3$ cells) in 96 well plates (VWR International, Leicestershire) with selective media, in this case Geneticin® (Invitrogen, Paisley) at a concentration of 1mg/ml. The cells were counted as described previously and plated out using a multichannel pipette with 100µl per well (80 wells total). The two outer rows of each 96 well plate were filled with 100µl sterile water to prevent evaporation. The plates were left at least two weeks to allow selection. Wells with Geneticin® resistant clones were observed under phase microscopy as healthy clumps of cells. The clone positions were noted and then picked into 24 well plates (Fisher Scientific, Leicestershire) with 2 ml of RPMI 1640 medium containing 10% FCS and 10% WEHI supplemented with 1mg/ml Geneticin®. 96 well plates with control cells were expected to be completely dead, whereas 96 well plates with pcDNA 3.1 vector
control (containing Geneticin® resistance) were expected to show a significant proportion of resistant clones.

2.4.2.6 **IL-3 independent growth**

To confirm if the gene of interest confers IL-3 independent growth, the clones were replated over time into media with 1mg/ml Geneticin® without IL-3. The cells were subjected to a withdrawal of IL-3 over a two week period, beginning with a 50% reduction in IL-3, followed by 25%, 10%, 5%, 2% and finally no IL-3. Clones can only survive without IL-3 if the gene of interest has conferred this ability upon the cells.

2.5 **Solutions**

**Freezing mix**
4ml RPMI supplemented with 10% foetal calf serum, L-glutamine and penicillin-streptomycin, 4ml DMSO and 12ml foetal calf serum.

**Guanidinium Thiocyanate (GTC)**
4M Guanidinium thiocyanate, 5mM EDTA, 25mM citrate pH7.0, 0.5% sarcosyl. 7μl β-mercaptoethanol per 1ml of GTC was added fresh prior to use.

**GM-mix**
Granulocyte Macrophage-Colony Stimulating factor (GM-CSF) 1ng/ml final concentration, Stem Cell Factor (SCF) 20ng/ml final concentration, Granulocyte- Colony Stimulating factor (G-CSF) 100ng/ml final concentration & IL-3 5ng/ml final concentration (All from First...
<table>
<thead>
<tr>
<th><strong>Link Ltd, Birmingham, UK</strong></th>
<th>Link Ltd, Birmingham, UK) in a volume of 50ml dH₂O.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hybridisation buffer (HB)</strong></td>
<td>50% formamide, 2 x SSC, 20% dextran sulphate, at pH 7.0.</td>
</tr>
<tr>
<td><strong>L-Broth</strong></td>
<td>25g L-Broth per litre dH₂O. Autoclave before use.</td>
</tr>
<tr>
<td><strong>LB/Agar</strong></td>
<td>3g Agar per 200ml L-Broth. Autoclave before use.</td>
</tr>
<tr>
<td><strong>6M NaCl</strong></td>
<td>350.64g NaCl dissolved in 1L dH₂O.</td>
</tr>
<tr>
<td><strong>1% Parafomaldehyde PBS</strong></td>
<td>PBS was prepared according to manufacturers instructions. One tablet was dissolved in 100ml of dH₂O, plus 1g Paraformaldehyde. Concentrated sodium hydroxide is used to aid dissolving.</td>
</tr>
<tr>
<td><strong>Phosphate Buffered Saline (PBS)</strong></td>
<td>PBS was prepared according to manufacturers instructions (Oxoid, Basingstoke, UK). One tablet was dissolved in 100ml of dH₂O.</td>
</tr>
<tr>
<td><strong>1 x PBS/Tween</strong></td>
<td>PBS was prepared as above with the addition of 4ml 25% Tween.</td>
</tr>
<tr>
<td><strong>Proteinase K</strong></td>
<td>50mg/ml in dH₂O.</td>
</tr>
<tr>
<td><strong>Red Cell Lysis Buffer (RCLB)</strong></td>
<td>155mM KHCO₃, 0.1mM EDTA, pH 7.4 at 4°C.</td>
</tr>
</tbody>
</table>
**Resuspension Buffer (RSB)**  
0.075M NaCl, 0.024M EDTA, pH8.0.

**10% SDS**  
10g SDS dissolved in 100ml dH₂O.

**Solution I**  
5mM glucose, 25mM Tris-Cl pH8.0, 10mM EDTA pH8.0

**Solution II**  
0.2M NaOH, 1% SDS

**Solution III**  
5M potassium acetate, 3M glacial acetic acid

**20 x SSC**  
175.3g NaCl and 88.2g sodium citrate were dissolved in 800mls water. The pH was adjusted to 7.0 using concentrated HCl and the volume brought up to 1 litre.

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

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

**Standard labelling mix**  
0.5M Tris-Cl pH7.8, 0.05M MgCl₂, 0.1M β-mercaptoethanol, 20µM dATP, 20µM dCTP, 20µM dGTP, 20µM dTTP, 0.5µg BSA and 25µg of dUTP conjugated with biotin or digoxygenin.

**Thawing Solution**  
20mls HBSS (Invitrogen, Paisley, UK), 100,000U DNase I (Invitrogen, Paisley, UK) and 200U sodium heparin (CP Pharmaceuticals, Wrexham, UK).
2.6 List of probes used for FISH

RP11-100O5  centromeric of *PDGFRB*, 5q33
RP11-2303L4  telomeric of *PDGFRB*, 5q33

RP11-467K9  centromeric of *GOLGA4*, 3p22
RP11-158E7  telomeric of *GOLGA4*, 3p22

RP11-434C1  telomeric of *ETV6*, 12p13
RP11-525I3  centromeric of *ETV6*, 12p13.
3 IDENTIFICATION OF NOVEL FUSION GENES ASSOCIATED WITH VISIBLE CHROMOSOME ABNORMALITIES

3.1 Introduction

As described in the introduction, activation of tyrosine kinases is a recurrent theme in the pathogenesis of myeloproliferative disorders (MPDs). However, the molecular basis of disease in many individuals remains obscure. In this chapter I describe the identification of new tyrosine kinase fusion genes that arise as a consequence of chromosomial translocations.

3.2 A novel PDGFRB fusion gene

3.2.1 Patient details

A patient (from here on referred to as Patient 1) was referred to us with a diagnosis of atypical CML, and pronounced peripheral blood eosinophilia (27%). The bone marrow aspirate showed markedly increased cellularity with granulocytic hyperplasia (90%, with left shift, but no excess of blasts) with marked increase in eosinophils and their precursors. Erythropoiesis was reduced but normoblastic. He was also described as having persistent breathlessness. Cytogenetics was performed after the initial referral, and a t(3;5)(p21;q33) was identified. The clinical phenotype of the patient combined with the karyotype suggested that the PDGFRB gene at 5q33 might be rearranged. Fluorescence in situ hybridisation (FISH) was first employed to determine if this was indeed the case.
3.2.2 Two-colour fluorescence in situ hybridisation (FISH)

FISH is commonly used to detect translocation breakpoints (Baxter, et al 2003). The PDGFRB probes were labelled with differently coloured fluorochromes, and signals appear as two red/green (yellow) adjacent signals in the nucleus when hybridised to normal interphase cells (Figure 3.1). In cells with a translocation and breakpoint in PDGFRB the signals hybridise to the normal PDGFRB allele creating a yellow fusion signal, with two separate signals indicating the ‘split’ in the translocated PDGFRB allele. FISH with the BAC clones RP11-100O5 and CTD-2303L4 confirmed the rearrangement of the PDGFRB gene. The BAC clone RP11-100O5 lies centromeric to PDGFRB, and was labelled with Biotin-16-2’-deoxy-uridine-5’-triphosphate (Biotin-16-dUTP) (Roche Mannheim, Germany). CTD-2303L4 is telomeric to PDGFRB and was labelled with alkali stable digoxigenin-dUTP. Interphase FISH showed a separation of the red and green signals in some cells (Figure 3.1). Some separation of signals may occur when the chromosomes are uncondensed at interphase, therefore I counted 100 cells and found 56% of interphase cells to have a separation in signals, compared to <10% of cells with split signals seen with these probes in normal controls (Baxter, et al 2003). These FISH results confirmed rearrangement of the PDGFRB gene.
Figure 3.1: FISH analysis of patient sample P1157. *PDGFRB* BACs RP11-100O5 (green) and CTD-2303L4 (red) hybridised to two interphase cells: (A) Normal cell with two pairs of signals fused, (B) A cell carrying a *PDGFRB* rearrangement, one fused pair and one separated pair of signals is observed. The schematic to the right indicates the position of the probes relative to *PDGFRB* and the labelling for each clone.

### 3.2.3 5’Rapid Amplification of cDNA Ends (5’RACE)

Having established the disruption of *PDGFRB* by FISH I set out to determine if the t(3;5) resulted in a novel fusion gene. The majority of tyrosine kinase fusions involve the joining of a partner gene to the kinase in a 5’ to 3’ orientation. To detect the ‘unknown,’ 5’ partner gene sequence I performed 5’Rapid Amplification of cDNA Ends (5’RACE) using the Invitrogen GeneRacer™ Kit. The basic principle of this kit is to ligate a known sequence onto the 5’ end of the unknown mRNA, enabling subsequent PCR to be performed using an oligonucleotide that is complementary to the ligated sequence.
Three independent RNA samples for the t(3;5) patient were available, designated P1157, E38 and E112. All three samples were analysed in duplicate by 5’RACE PCR. I performed the 5’RACE PCRs with the High Fidelity PCR Master kit (Roche Diagnostics, Lewes, UK) according to manufacturer instructions, using the 5’GeneRacer™ primer and a reverse primer in PDGRFB exon 15.1R (for all primer sequences see appendix I, section IV). The amplification was performed under long-range PCR conditions (according to that specified in section 2.2.5) to enable fragments up to 10 kb to be amplified. No products were observed on a 1 % agarose gel after the first round PCR. I then performed a nested PCR using the same PCR conditions. For the nested PCR, I employed the 5’GeneRacer™ nested primer and a nested PDGFRB primer, exon 14.1R. As shown in Figure 3.2, several different size bands were obtained after nested PCR, two of which were very strong on the gel (indicated by the red arrows) and were sized at approximately 1 kb and 2.5 kb respectively. A third reaction yielded two medium intensity bands at approximately 0.5 kb plus a slightly larger weaker band (indicated by the white arrow).

Despite the fact that the amplified products were not consistent in size between samples, I cloned all three products using the TOPO TA cloning kit. Although visible colonies were observed on all recombinant plates, none of the picked colonies grew overnight, implying contamination, possibly via the SOC medium. Due to the strong sharp band achieved from nested 5’RACE PCR I went on to directly sequence the product from P1157 (Figure 3.2 - indicated by a star) in duplicate using the 5’GeneRacer™ nested primer (forward) and the PDGFRB exon 14.1R primer (reverse).
Figure 3. 2: 1% Agarose gel showing nested 5’RACE PCR for all three samples, in duplicate, of patient 1. Three different size bands were observed one in each of the samples; P1157 (L1), E112 (L3) and E38 (L5). The negative control shows no product (L7).

The sequences were first analysed by BLAT analysis (http://genome.ucsc.edu) against the whole human genome. The sequence was 630 bp in length and found to have 140 bp homology to chromosome 3, and 305 bp homology to chromosome 5. The chromosome 5 sequence was as expected, PDGFRB and the chromosome 3 sequence was 98.6% homologous to the gene GOLGA4. GOLGA4 is a member of the Golgin subfamily A and is located at chromosome 3p22.3 and is close to the cytogenetic breakpoint in the t(3;5). Using the accession numbers retrieved from the UCSC genome browser I then went on to BLAST the sequence (www.ncbi.nlm.nih.gov/) first against PDGFRB (NM_002609), to check the fragment was homologous to the region expected from the primers used and to identify the breakpoint. The sequence was found to diverge at nucleotide 1936 of PDGFRB (Figure 3. 3).
Figure 3.3: BLAST results – comparison of PDGFRB cDNA (NM_002609) against sequenced product from nested 5’RACE PCR (P1157). Note that position 267 and subsequent sequence matches PDGFRB.

BLAST analysis was then performed on the same sequence against GOLGA4 (NM_002078) and this also matched, ending at nucleotide 1513 (Figure 3.4). There were a number of bases not assigned by the sequencing programme due to dye spots in the original sequencing, I manually filled these in after BLAST analysis and determined the breakpoint to be at nucleotide 1520 in GOLGA4.
**Figure 3.4:** BLAST results – comparison of GOLGA4 (NM_002078) against sequenced product from nested 5' RACE PCR (P1157). Note sequence ends at nucleotide 1513.

The individual full-length mRNA sequences for PDGFRB and GOLGA4 were then copied into DNA Strider™ (a software program used to analyse DNA sequences). The open reading frame for each sequence was determined and the sequence translated, using both the DNA Strider™ software compared with the reported protein
sequence on the NCBI website. The results from the NCBI alignment between the RACE product sequence and the mRNA sequence were then used to establish the location of the breakpoint in relation to the exons of each gene. A fusion of exon 10 \textit{GOLGA4} to exon 11 \textit{PDGFRB} was revealed. Subsequent DNA Strider™ analysis of the fused mRNA sequence showed the fusion is in-frame (Figure 3.5A). To confirm the mRNA fusion I used reverse transcriptase PCR (RT-PCR), before going on to clone the genomic breakpoint. This confirmation is important as 5’RACE-PCR can give rise to a variety of artifacts.

\subsection{Confirmation of the GOLGA4-PDGFRB fusion and cloning of the genomic breakpoints.}

To confirm the \textit{GOLGA4-PDGFRB} fusion, I employed RT-PCR using the primers \textit{GOLGA4} exon 8.1F and \textit{PDGFRB} exon 12.1R with the High Fidelity PCR master kit according to manufacturer’s conditions (see appendix I, section VI for all primer sequences). The PCR was performed on a generic 66°C annealing temperature PCR protocol. The in-frame fusion was detectable by single step RT-PCR, as seen in Lane 5, Figure 3. 5B. Using the same High Fidelity PCR kit and amplification protocol I used RT-PCR with \textit{PDGFRB} exon 9.1F and \textit{GOLGA4} exon 10.1R to try and amplify the reciprocal translocation. \textit{PDGFRB-GOLGA4} reciprocal products were not detectable by single step, Lane 6, Figure 3. 5B or nested PCR.

Having confirmed the in-frame mRNA, I went on to amplify the genomic breakpoint to further substantiate the presence of this fusion. To do this I retrieved the sequences for the introns where I presumed the breakpoint to be in each gene (intron 10 of \textit{PDGFRB} and intron 10 of \textit{GOLGA4}, the intron directly before or after the exon in which the mRNA breakpoint was observed), focusing on these before considering any complexities, such as alternative splicing. The \textit{PDGFRB} intron is about 3.2 kb and I designed eight reverse primers approximately every 250 bp. In some cases the primers were separated by up to 800 bp due to areas of repetitive sequence (for all primer sequences see appendix I, section VII). \textit{GOLGA4} intron 10 is approximately
13kb and for this intron I designed a primer approximately every 1kb, again avoiding areas of repeat sequence (for all primer sequences see appendix I, section VIII). After checking the primers on the NCBI nucleotide BLAST site to ensure they only annealed to the gene in question, I set up a number of amplification reactions for the patient.

**Figure 3.5**: Characterisation of the *GOLGA4-PDGFRB* mRNA fusion. (A) Schematic showing sequence and amino acid translation of the in-frame fusion, (B) Amplification of the *GOLGA4-PDGFRB* fusion by single step RT-PCR: Lanes 1-4 represent normal controls, L5 represent the patient (sample P1157), L6 represents the reciprocal breakpoint PCR (sample P1157) and L7-8 represent no DNA controls, (C) Electropherogram showing the sequencing of the single step confirmatory PCR (L5) and the mRNA junction sequence of the in-frame fusion.

All amplifications were carried out on a long range PCR program (for details of PCR program see section 2.2.5). I began by setting up fifteen reactions according to Table
3. 1, the first seven using a common *GOLGA4* forward primer and the remaining eight using a common *PDGFRB* reverse primer. No product was amplified from any of these PCRs.

<table>
<thead>
<tr>
<th>Reaction No.</th>
<th>Common Primer</th>
<th>Variable primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cGOLGA4 exon 8.1F</td>
<td>PDB.I11.7R</td>
</tr>
<tr>
<td>2</td>
<td>cGOLGA4 exon 8.1F</td>
<td>PDB.I11.6R</td>
</tr>
<tr>
<td>3</td>
<td>cGOLGA4 exon 8.1F</td>
<td>PDB.I11.5R</td>
</tr>
<tr>
<td>4</td>
<td>cGOLGA4 exon 8.1F</td>
<td>PDB.I11.4R</td>
</tr>
<tr>
<td>5</td>
<td>cGOLGA4 exon 8.1F</td>
<td>PDB.I11.3R</td>
</tr>
<tr>
<td>6</td>
<td>cGOLGA4 exon 8.1F</td>
<td>PDB.I11.2R</td>
</tr>
<tr>
<td>7</td>
<td>cGOLGA4 exon 8.1F</td>
<td>PDB.I11.1R</td>
</tr>
<tr>
<td>8</td>
<td>PDB.I11.1R</td>
<td>gGOLGA4.I9.1F</td>
</tr>
<tr>
<td>9</td>
<td>PDB.I11.1R</td>
<td>gGOLGA4.I9.2F</td>
</tr>
<tr>
<td>10</td>
<td>PDB.I11.1R</td>
<td>gGOLGA4.I9.3F</td>
</tr>
<tr>
<td>11</td>
<td>PDB.I11.1R</td>
<td>gGOLGA4.I9.4F</td>
</tr>
<tr>
<td>12</td>
<td>PDB.I11.1R</td>
<td>gGOLGA4.I9.5F</td>
</tr>
<tr>
<td>13</td>
<td>PDB.I11.1R</td>
<td>gGOLGA4.I9.6F</td>
</tr>
<tr>
<td>14</td>
<td>PDB.I11.1R</td>
<td>gGOLGA4.I9.7F</td>
</tr>
<tr>
<td>15</td>
<td>PDB.I11.1R</td>
<td>gGOLGA4.I9.8F</td>
</tr>
</tbody>
</table>

**Table 3.1:** First set of primers used to amplify the genomic breakpoint in the t(3;5) patient. The successful primer pair is highlighted.

As a result, I designed another *GOLGA4* primer in exon 10 (cGOLGA4 exon 10.1F) closer to the mRNA breakpoint, also enabling any breakpoint between the end of the exon and the first intronic primer to be discovered (appendix I, section IX). I carried out the same amplifications, as shown in Table 3.1, but substituted the *GOLGA4* exon 8 primer for the new exon 10 primer. On this occasion a band sized at >2.5 kb was observed as result of amplification with the primer pair gGOLGA4.I9.7F, and PDB.I11.1R (Figure 3.6, Lane 15).
I cloned the amplified band using the TOPO TA cloning kit (Invitrogen) and then sequenced it. In the forward reaction, the traces showed sequence reading from gGOLGA4 I9.7F into GOLGA4 intronic sequence up to nucleotide 9797 (according to my DNA Strider map of intron 10 GOLGA4). The reverse reaction showed only PDGFRB sequence from the mRNA breakpoint at base 1936 reading into PDGFRB for 33 base pairs (bp). Following this result I used a primer internal to gGOLGA4 I9.7F, gGOLGA4.I9.9F (at nucleotide 9684), combined with each of the PDGFRB intronic primers (1aR-7R), performing the PCR from the cloned product (lane 15, Figure 3.6). I used a generic 66°C annealing temperature PCR program with a high fidelity Hotstart Taq DNA polymerase (Qiagen). Figure 3.7 represents the resulting genomic fusion products from this amplification: all but one lane produced a band (L7) indicating that the breakpoint was before PDGFRB intronic primer PDB.I11.7R.

Using the plasmid DNA from the previously cloned product, I then directly sequenced across the genomic breakpoint using PDBI11.6R as this had produced the smallest product size of all the primer pairs. A fusion of GOLGA4 intron 10 to PDGFRB intron 10 was discovered (Figure 3.8A).
1% Agarose gel of the amplification products of the $GOLGA4$-$PDGFRB$ genomic fusion amplified from plasmid DNA diluted 1:1000. Lanes 1-7 represent the common c$GOLGA4$ exon 10.1F primer combined with each of the $PDGFRB$ intronic primers (1aR-7R respectively).

To confirm the genomic breakpoint I designed primers closer to the breakpoint in $GOLGA4$ and $PDGFRB$ and re-amplified the fusion using patient DNA in a single step PCR reaction, using the High Fidelity PCR Master Mix. The forward genomic primer used was in $GOLGA4$ intron 10 (7F) in combination with a new $PDGFRB$ reverse primer in intron 11 (Brkpt.R) (see appendix I, section X for primer sequences). This confirmed the presence of the fusion of intron 10 $GOLGA4$ to $PDGFRB$ intron 10 (Figure 3.8A). Using the same High Fidelity PCR kit and amplification protocol I used PDB.I11.recip.F and PDB.I11.recip.2F combined with a common intronic $GOLGA4$ primer, g$GOLGA4$.I9.recip.R to amplify the reciprocal translocation (for primer sequences see appendix I, section XI). $PDGFRB$-$GOLGA4$ reciprocal products were not detectable by single step or nested PCR (Lanes 3 and 4,
Figure 3.8B). This may have been because of a deletion, commonly seen at the site of the chromosomal breakpoints.

Figure 3.8:  

**GOLGA4-PDGFRB** genomic fusion in patient 1 with t(3;5). (A) Schematic showing the individual sequences and the GOLGA4-PDGFRB fusion, (B) Amplification of GOLGA4-PDGFRB by single step PCR; Lanes 1-2 represent normal controls, lanes 3-4 the reciprocal PDGFRB-GOLGA4 genomic fusion, lane 5 the GOLGA4-PDGFRB genomic fusion and a no DNA control in lane 6 (C) Electropherogram showing the junction sequence in the intron of each gene.
3.2.5  

**Consequence of the t(3;5)**

The experiments performed indicate that the t(3;5) fuses *GOLGA4* exon 10 to *PDGFRB* exon 11. The predicted length of the fusion mRNA is 3662 bp with an open reading frame of 2976 bp that is predicted to generate a chimaeric protein of 992 amino acids, of which 412 amino acids are derived from GOLGA4 and 580 amino acids are derived from PDGFRβ (Figure 3.9). The predicted molecular weight of the protein is 251 kDa.

Normal GOLGA4 is a 2230 amino acid protein. At its c-terminus is a GRIP domain that targets GOLGA4 to the trans-golgi network (TGN) membranes in a G-protein dependent manner (see Figure 3.9 – indicated in yellow). GRIP domain proteins in general contain extensive coiled-coil regions (represented by the blue regions – Figure 3.9), that enable them to self associate and form rod-like homodimers (Luke, et al 2003). The chimaeric protein formed from the *GOLGA4-PDGFRB* fusion is expected to be structurally very similar to other tyrosine kinase fusions reported. It retains the entire catalytic domain of the tyrosine kinase fused to a portion of GOLGA4 containing substantial regions of coiled-coil domains.
**Figure 3.9:** Schematic representation of the normal PDGFRβ, GOLGA4 and GOLGA4-PDGFRβ. The fusion protein retains the entire catalytic tyrosine kinase domain of PDGFRβ and has a substantial proportion of the amino terminal GOLGA4 partner protein containing coiled-coil domains. These oligomerisation motifs in GOLGA4 are likely to be essential for the putative constitutive tyrosine kinase activity of the fusion and the subsequent downstream signalling through proliferation and survival pathways.

### 3.2.6 Identification of a second MPD patient with a GOLGA4-PDGFRB fusion

Following the identification of the GOLGA4-PDGFRB fusion, a second patient with an eosinophilic MPD and a t(3;5)(p21-25;q31-35) was referred to us for analysis (patient reference number 2).
Figure 3.10: Characterisation of the GOLGA4-PDGRFB mRNA fusion in Patient 2. (A) Schematic showing sequence and amino acid translation of the in-frame fusion, (B) The gel picture on the left represents the amplification of the KIAA1449-PDGFRB fusion; lanes 1 & 4 are patient 2, lane 5 a normal control and finally the no DNA controls are shown in lanes 2 and 3. Amplification of the GOLGA4-PDGRFB fusion by single step PCR is represented by the picture on the right, lanes 1 and 5 represent patient 2, lane 2 patient 1 used as a positive control, lane 3 and 4 are no DNA controls and lane 5 a normal control. A band of approximately 500bp is visible for the GOLGA4-PDGRFB fusion only and is the same size as that for the positive control; (C) Electropherogram comparing the mRNA junction sequence of the in-frame fusion in both patients with a GOLGA4-PDGRFB fusion gene. The upper trace corresponds to patient 2 (E923) and patient 1 (P1157) is shown in the lower trace.

In this case, I performed RT-PCR to analyse the patient sample for specific 3p partner genes fused to PDGFRB. I tested for the GOLGA4-PDGRFB fusion, using the primers cGOLGA4 exon 7.1F and PDGFRB exon 12.1R (for primer sequences see
appendix I, section XII). A second variant PDGFRB fusion gene had also been cloned in our group, involving a gene called at 3p22 called WDR48 (also known as KIAA1449). This was also tested for using RT-PCR using the primers KIAA1449.2F combined with PDGFRB.exon15.2R (for primer sequences see appendix I, section XIII). Patient was found to be positive for GOLGA4-PDGFRB and negative for WDR48-PDGFRB. On sequencing of the product, I determined the cDNA breakpoint was the same as that for the first patient (Figure 3.10A, B and C). Again, PDGFRB-GOLGA4 reciprocal products were not detectable by single step amplification.

3.2.7 Two-colour fluorescence in situ hybridisation (FISH)

The BAC clones RP11-100O5 and CTD-2303L4 were again used to confirm the rearrangement involved the PDGFRB gene. 100 interphase cells were counted and of these 52% showed a split in signals, in agreement with the PCR result (Figure 3.11A). The 52% separation in signals observed by the PDGFRB FISH is well above the false positive rate of 3% as determined by Baxter et. al., (2003) confirming the rearrangement of PDGFRB (Baxter, et al 2003). To further confirm the PCR result, two-colour FISH was also performed using a BAC clone that lies centromeric to the PDGFRB breakpoint, RP11-100O5 and a chromosome 3p BAC RP11-158E7 that lies telomeric to the GOLGA4 breakpoint. Interphase FISH verified the presence of the GOLGA4-PDGFRB fusion, as the signals appeared to be fused in 36% of cells (Figure 3.11B), also indicating the fusion initially found by RT-PCR to be real.
Figure 3.11: Fluorescent in situ hybridisation for PDGFRB and GOLGA4. (A) PDGFRB BAC RP11 100O5 (green) and GOLGA4 RP11-158E7 (red) BAC hybridised to an interphase from patient E923, note the fused signal observed in one of the cells; (B) PDGFRB BACs RP11-10005 (green) and CTD-2303L4 (red) hybridised to an interphase from patient E923, the cell shows one pair of signals separated.

3.2.8 Imatinib sensitivity assay

Patients with rearrangements of PDGFRB usually show a good clinical response to imatinib (Apperley, et al 2002). To specifically test if the novel GOLGA4-PDGFRB fusion was sensitive to imatinib, I analysed colony-forming unit, granulocyte-macrophage growth (CFU-GM) production for cryopreserved cells from patient 2, in the presence of 0μmol/L, 1μmol/L or 5μmol/L imatinib. A normal control individual and a BCR-ABL positive CML patient were tested for comparison. Peripheral blood from the normal control showed a reduction in colony numbers of approximately
50% with $1 \mu$mol/L imatinib at 7 and 14 days (Figure 3.12A) and a further 10% reduction was observed at $5 \mu$mol/L imatinib. This degree of non-specific inhibition of CFU-GM numbers has been found previously with a large number of other normal controls, as previously tested by Dr Andy Chase.

![Graphs showing CFU-GM growth](image)

**Figure 3.12:** CFU-GM growth of blood mononuclear cells from (A) normal control, (B) a CML patient and (C) patient 2.

The positive control CML patient, showed a much greater reduction in the number of CFU-GM after exposure to imatinib. Approximately 90% growth inhibition was observed in the presence of $1 \mu$mol/L and $5 \mu$mol/L imatinib at both 7 and 14 days (Figure 3.12B). This replicated similar data obtained in multiple imatinib sensitive CML samples. Marked inhibition was also observed for CFU-GM grown from the
t(3;5) patient, with approximately 95% reduction in colony numbers at day 7 and 14 in the presence of 1μmol/L and 5μmol/L imatinib (Figure 3.12C). The index of response (see methods, section 2.4.1.2) for the t(3;5) patient was 0.05, markedly lower than the 0.2 control which has been used to distinguish imatinib responders from imatinib non-responders. These data suggest that the t(3;5)-associated disease is indeed sensitive to imatinib.

3.2.9 Discussion

I have identified the gene found fused in a recurrent t(3;5) by using a combination of FISH and translocation cloning. I have discovered a novel partner gene, GOLGA4, fused to the tyrosine kinase gene PDGFRB. The breakpoint was analogous to all other reported fusions involving PDGFRB, with the exception of the t(5;14) (Vizmanos, et al 2004). The translocation breakpoint on chromosome 3 lies in intron 10 of GOLGA4. GOLGA4 contains extensive coiled-coils domains that are thought to act as oligomerisation motifs essential for GOLGA4-PDGFRβ constitutive activation.

GOLGA4 (Golgin-245) was first identified by two groups that were probing cDNA libraries with sera from patients with Sjogens syndrome (an autoimmune disease) (Fritzler, et al 1995, Kooy, et al 1992). Golgin-245 is a member of the Golgin family of coiled-coil proteins, associated with the Golgi apparatus. Golgins play a role in tethering events in membrane fusion and as structural supports for Golgi cisternae. Golgin-245 belongs to a subfamily known as the GRIP domain proteins. Van Valkenburgh et. al., (2001) showed that the GRIP domain binds the ARF-like GTPases Arl1 and Arl3 with significantly higher affinity than Rab 6. Recently Rao et. al., (2003), provided evidence for the function of the Arl proteins. Through yeast GFP-GRIP fusion proteins they were able to show a GTPase cycle involving Arl3p that regulates the golgi localisation of Arl1p which in turn binds the GRIP domain and recruits the golgin to the golgi (Rao, et al 2003). Although, it has been suggested that Golgin-245 might function as a vesicle coat, it is perhaps more likely that it is
involved in the organisation of membrane subcompartments of the trans-Golgi
Golgin-245 remains to be fully elucidated.

To date eighteen fusion partners for PDGFRB have been reported: ETV6, TRIP11,
HIP1, CCDC6, RABEP1, PDE4DIP, SPECC1, NIN, KIAA1509, TP53BP1, TPM3,
NDE1, ERC1, PRKG2, GPIAP1, GIT2, SPTBN1 and MYO18A (Abe, et al 1997,
partners known within our laboratory, including GOLGA4 (Hidalgo-Curtis et. al.,
submitted 2009). With the exception of the t(5;14)(q33;q24), all breakpoints within
PDGFRB occur in intron 10 and this is also the case with the novel GOLGA4-
PDGFRB fusion detailed here. The partner proteins in these tyrosine kinase fusions
typically contribute dimerisation motifs leading to the constitutive activation of the
receptor and inappropriate activation of growth signalling and anti-apoptotic
pathways (Cross and Reiter 2002). In an analogous manner to the BCR-ABL
oncoprotein, constitutively active tyrosine kinase fusion proteins deregulate
haematopoiesis.

Interestingly GOLGA4 shows a functional overlap with the previously identified
PDGFRβ fusion partners BIN2, HIP1, RABEP1 and SPTBN1. BAR domain proteins
like BIN2 are involved in membrane curvature and receptor endocytosis (Ge and
Prendergast 2000). Other fusion partners also demonstrate links to intracellular
sorting of receptors, a well known example is BCR. BCR is a component of the
endosomal sorting complex required for transport (ESCRT) (Olabisi OO 2006).
MPDs are well known to be characterised by hyper-responsiveness to growth factors.
Suppression of BCR in HeLa cells by small interfering RNA impairs epidermal
growth factor receptor turnover, and it is conceivable that as components of fusions,
partner proteins such as BIN2, WDR48 and GOLGA4 may contribute to defective receptor internalisation and sorting.

Patients with rearrangements of the PDGFRB gene respond well to treatment with imatinib mesylate (Apperley, et al 2002). The identification of PDGFRB gene rearrangements is therefore critical for disease management. The identification of PDGFRB fusions in these two cases has enabled them to be treated and permitted their response to be monitored by RT-PCR and FISH, techniques which can help to provide early clinical indications of resistance to imatinib therapy.

3.3 IDENTIFICATION OF TWO NOVEL PDGFRA FUSION GENES

The PDGFRA gene is located at 4q12 and to date has been described fused to six different partners: BCR, FIP1L1, KIF5B, CDK5RAP2, STRN and ETV6 (Baxter, et al 2002, Cools, et al 2003, Curtis, et al 2007, Score, et al 2006, Walz, et al 2006). As with PDGFRB fusions, identification of patients that harbour PDGFRA fusions is particularly important as they respond well to targeted therapy with imatinib (Cools, et al 2003, Reiter, et al 2007). PDGFRA fusions are most commonly associated with chronic eosinophilic leukaemia (CEL), with the most frequent abnormality being seen in approximately 12% of cases with a provisional diagnosis of idiopathic hypereosinophilic syndrome (Cools, et al 2003, Griffin, et al 2003), is the FIP1L1-PDGFRα fusion gene. Here I describe two patients with imatinib responsive CEL who presented with acquired chromosomal rearrangements of 4q.

3.3.1 Patient details

Case 1 was diagnosed with CEL and has been described in further detail by Musto et al, 2004. In brief, the patient was a 64 year old male who presented with an eosinophil count of 12x10⁹/L and a bone marrow indicative of an MPD. Cytogenetic analysis revealed 60% of bone marrow metaphases harboured a t(2;4)(p24;q12).
Daily treatment with 100mg imatinib was commenced and following treatment haematologic and cytogenetic remission was achieved. This dose was continued for 9 months and then progressively reduced to 100mg on alternate days followed by 100mg twice a week. Following 24 months of imatinib, the patient refused further therapy despite not having any significant side effects. After 14 months after discontinuation of treatment elevated eosinophil counts were again detected.

Case 2 was another male, 51 years of age, who presented with late onset asthma and Type 2 diabetes. Again the patient exhibited an elevated eosinophil count of 12.9 x 10^9/L and a hypercellular bone marrow suggestive of an MPD. Cytogenetics revealed 100% of bone marrow metaphases carried a t(4;12)(q2?3;p1?2). For over 7 years the patient was managed with hydroxycarbamide but with recurrent problems due to anaemia, thrombocytopenia and very poor control of the same translocation: 46,XY,t(4;12)(q2?3;p1?2)[14]/46,XY[6]. Following the molecular diagnosis as detailed by the work done below, hydroxycarbamide was stopped and imatinib started at 100mg daily. Within four weeks, the full blood counts completely normalised, resulting in significant resolution of symptoms (apart from the asthma) and greatly improved quality of life. Following 9 months of treatment, cytogenetic analysis was again performed and indicated complete cytogenetic remission (30 metaphases).

Upon referral it was known that both cases were negative for BCR-ABL and FIP1L1-PDGFR by RT-PCR or FISH analysis using standard procedures. The presence of a 4q translocation in conjunction with persistent eosinophilia suggested that PDGFRA might be involved.

### 3.3.2 Bubble PCR

To establish if the 4q12 translocations resulted in a novel fusion genes involving PDGFRA, I employed bubble PCR. Bubble PCR (Zhang, et al 1995) is a one sided PCR technique capable of amplifying an unknown sequence that lies adjacent to a known sequence (see section 2.2.6 for more details and appendix I, section I for
primer sequences). Like 5’RACE, this relies on the fact that fusion of tyrosine kinases involves the joining of a partner gene to the kinase in a 5’ to 3’ orientation. This technique also relies on the tight clustering of genomic breakpoints observed in PDGFRA fusions, with almost all involving a truncated exon 12 (Reiter, et al 2007). The technique is also dependent on specific restriction enzyme sites distal to the fusion point but equally within an amplifiable distance. The 4 bp cutters HaeIII and RsaI are used to target this region in the PDGFRA gene and in combination with the primer PDAI12-R4 target genomic regions of 978 bp and 615 bp, respectively (Figure 3.13).

![Diagram of bubble PCR strategy showing PCR primers and restriction sites in the region where all breakpoints have been reported to fall. Arrows indicate primer locations.](image)

Figure 3.13: Schematic of bubble PCR strategy showing PCR primers and restriction sites in the region where all breakpoints have been reported to fall. Arrows indicate primer locations.

On average these enzymes cut approximately every 256 bp, thus products from any rearranged PDGFRA allele are generally expected to be smaller than the wild-type allele. An advantage of using this approach, over that of the 5’RACE kit, is that it relies on genomic DNA not RNA. This was fortunate as the only material available for case 1 was a stained haematological slide. I employed bubble PCR and detected rearranged bands were detected for both cases 1 and 2 (Figure 3.14).
The rearranged PDGFRA alleles were subsequently subcloned into pCR 4-TOPO vector and sequenced using BigDye terminator v1.1 (Applied Biosystems, UK). Sequencing of the rearranged allele for case 1 identified a novel fusion between STRN (NM_003162) and PDGFRA, whereas a novel fusion between ETS-variant gene 6 (ETV6, NM_001987) and PDGFRA was found for case 2.

**Figure 3.14:** Agarose gel picture of nested bubble PCR after digestion with HaeIII and RsaI. The normal controls show the expected germline bands of 1011 bp and 648 bp respectively (the distance between primer PDAI12-R4 and the relevant restriction site plus 33 bp contributed by the bubble oligo). Smaller bands derived from the rearranged allele are visible for cases 1 and 2 after HaeIII digestion and for case 2 after RsaI digestion. The normal allele is often out competed due to size in the amplification reaction.

### 3.3.3 Confirming the STRN-PDGFR A fusion

The rearranged allele resulted in a fusion of STRN intron 6 and a truncated PDGFRA exon 12. Hence, primers were designed to confirm the presence of the fusion (see appendix I, section XVI). Hot-start amplifications were performed for between 30-35
cycles in a MJ Research Inc. Tetrad thermocycler with a 66°C annealing temperature. The genomic \textit{STRN-PDGFR} fusion was amplified using STRN-Fusion-Ex6-1F and PDGFR-Fusion-Ex12-2R. A 285 bp product was amplified from the patient DNA but not from control DNA. A reciprocal \textit{PDGFR-STRN} fusion product was not detected with primers PDGFR-recip-Ex11-1F and STRN-Recip-Int6-1R with either single step or nested PCR (appendix I, section XVI for primer sequences) (Figure 3.15).

\textbf{Figure 3.15:} Characterisation of the \textit{STRN-PDGFR} fusion gene. (A) Electropherogram of the genomic DNA junction amplified by bubble PCR; (B) Sequence of \textit{PDGFR} exon

\begin{align*}
\text{PDGFR exon 12} & \quad \text{ATCACCCAGATGGACATGAATATATGTGGACCCGATG} \\
\text{STRN intron 6} & \quad \text{GCATAAAGAATCCATGTTCTTTAATCTTAACCCGTAGCTTT} \\
\text{STRN-PDGFR} & \quad \text{GCATAAAGAATCCATGTTCTTTAATCTTAACCCGTAGCTTT}
\end{align*}

\begin{align*}
\text{STRN exon 6} & \quad \Delta \text{PDGFR exon 12} \\
(i) & \quad \text{CATTGATACTTCAACACYGCGTTATGACTCAA} \\
(ii) & \quad \text{CATTGATACTTCAACACYTCCCAAGATGGAC}
\end{align*}
12 and STRN intron 6 (underlined) flanking the breakpoints, aligned with STRN-PDGFRα genomic fusion sequence; (C) Amplification of the STRN-PDGFRα but not the reciprocal genomic DNA fusion from case 1 versus normal controls. (D) Predicted in frame STRN-PDGFRα mRNA fusions.

Due to the lack of adequate pre-treatment patient material, identification of the mRNA fusion was not possible. Nonetheless, analogous PDGFRα breakpoints for FIP1L1-PDGFRα usually result in the use of cryptic splice sites within PDGFRα exon 12 downstream of the breakpoint. Two potential AG splice acceptor sites were identified and would result in in-frame mRNA fusions with STRN exon 6. It is likely that either or both of these might represent the real mRNA fusion in case 1 (Figure 3.15D).

3.3.4 Consequence of the STRN-PDGFRα fusion

The t(2;4) is likely to fuse STRN exon 6 to part of PDGFRα exon 12. The predicted length of the fusion mRNA is dependent on the PDGFRα splice site used and I refer to the most likely mRNA fusions as (i) and (ii) (see Figure 3.15D). The predicted length of the mRNA fusion (i) is 5309 base pairs with an open reading frame of 2328 bp. The predicted length of the mRNA fusion (ii) is 5289 bp with an open reading frame of 2304 bp. The t(2;4)(p24;q12) is thus predicted to generate a chimaeric protein of (i) 776 amino acids or (ii) 768 amino acids, of which 265 amino acids are derived from Striatin (Figure 3.16). The predicted molecular weight of the protein is (i) 187.5 kDa or (ii) 189.4 kDa. Normal Striatin is a 780 amino acid protein. At its C-terminus is a region of WD40 repeats and its N-terminus a coiled-coil domain. This coiled-coil region is thought to be crucial for homo- and hetero-oligomerisation of the protein. Striatin is a scaffolding protein involved in calcium dependent signalling (Bartoli, et al 1999) and is also thought to have a putative role in trafficking (Baillat, et al 2001). It is highly expressed in the brain, in particular the neurons, and has been shown to directly bind caveolin (Gaillard, et al 2001). The chimaeric protein formed from the STRN-PDGFRα fusion is structurally very similar to other reported tyrosine kinase fusions. The entire catalytic domain of the tyrosine kinase is retained and
found fused to a portion of Striatin containing coiled-coil domains which probably mediate dimerisation of the fusion protein.

**Figure 3.16:** Structure of Pdgfrα, Striatin and the predicted fusion protein. Relevant domains are shown: transmembrane (tm), WW-like (WW) and tyrosine kinase domains for Pdgfrα; coiled-coil region and WD40 repeats for Striatin. The partially deleted autoinhibitory WW domain of Pdgfrα is shown in Striatin-Pdgfrα.

### 3.3.5 Confirming the ETV6-PDGFRα fusion

As mentioned previously, the sequencing of the rearranged bubble PCR band from case 2 identified a second novel fusion between *ETV6*, intron 6 and *PDGFRA* intron 11 (Figure 3.17A and B). To confirm the fusion a 210 bp product (Figure 3.17C) was specifically amplified from case 2 peripheral blood genomic DNA using breakpoint-specific primers ETV6.Fusion.Int6.4F and PDGFRA.fusion.Ex12.1R on a standard 66°C Hotstart PCR amplification, a band was not observed in control DNA (for all primer sequences see appendix I, section XVII). A reciprocal *PDGFRA-ETV6* fusion product was not detected with PDGFRA-recip-Ex11-1F and ETV6-Recip-Ex7-1R at either single or nested PCR. Expression of *ETV6-PDGFRA* was confirmed by single-step from case 2, using primers ETV6.Fusion.Ex6.1F and PDGFRA.Fusion.Ex12.2R, yielding a 276 bp product that was not observed in normal control cDNA (Figure 3.27C). Sequencing of this product revealed an in-frame whole exon fusion between
ETV6 exon 6 and PDGFRA exon 12 (Figure 3.17D). To further confirm the presence of the ETV6-PDGFRA fusion, two-colour interphase FISH analysis was performed.

**Figure 3.17:** Characterisation of the ETV6-PDGFRA fusion. (A) Electropherogram of the genomic DNA junction amplified by bubble PCR; B) Sequence of PDGFRA intron 12 (italics) and ETV6 intron 6 (underlined) flanking the breakpoints, aligned with the ETV6-PDGFRA genomic fusion sequence; (C) Amplification of the ETV6-PDGFRA fusion from genomic DNA but not the reciprocal genomic DNA fusion from case 2, versus normal controls. (D) Predicted in-frame ETV6-PDGFRA mRNA fusion as determined by sequence analysis of case 2 RT-PCR product.
FISH was only performed on Case 2 due to availability of material. FISH was executed using the bacterial artificial chromosome (BAC) clones RP11-24O10 (3’ of *PDGFRA*) and RP11-434C1 (5’ of *ETV6*) selected from sequences flanking *PDGFRA* and *ETV6* ([www.ensembl.org](http://www.ensembl.org)) or commercially available probe sets. BAC DNA was grown, extracted and labelled by nick translation with digoxygenin or biotin-16-dUTP (Roche, East Sussex) and hybridised using standard procedures. On normal cells the combined BACs gave four separate signals corresponding to the normal chromosomes 4 and 12. However, in 56% of cells for case 2, overlapping probe signals consistent with co-localisation of *ETV6* and *PDGFRA* were observed (Figure 3.18).

**Figure 3.18:** (A) Relative positions of FISH probes with the orientation of each gene indicated by the arrows. (B) A normal cell shows four distinct signals whereas 56% of cells from case 2 showed a fused red/green signal indicative of the t(4;12).
3.3.7 Consequence of the ETV6-PDGFRα fusion

The t(4;12) fuses ETV6 exon 6 to PDGFRα exon 12 resulting in a novel fusion gene. The predicted length of the fusion mRNA is 6015 bp with an open reading frame of 2769 bp. ETV6-PDGFRα fusion is predicted to encode a chimaeric protein of 923 amino acids, of which 384 amino acids are derived from ETV6 and 539 amino acids derived from PDGFRα (Figure 3.19). The predicted molecular weight of the protein is 229.3 kDa. Normal ETV6 is a 452 amino acid protein with two functional domains: a C-terminus pointed domain (PNT) that plays a role in protein-protein interactions including self association and interactions with other proteins. At its C-terminus is the ETS domain that is crucial for DNA-binding. The entire catalytic domain of PDGFRα is retained and found fused to almost the entire ETV6 protein containing the whole pointed domain and a part of the ETS domain.

![Figure 3.19: Structure of PDGFRα, ETV6 and the predicted fusion protein. The pointed and ETS domains are shown for ETV6. The dotted vertical line indicates the breakpoint region in each protein.](image-url)
Characterisation of the fusion in each case enabled a molecular assessment of minimal residual disease to be made. Minimal residual disease (MRD) was analysed for cases 1 and 2 by two step nested PCR to maximize sensitivity of detection. Primers used for the detection of the STRN-PDGFR\(\alpha\) fusion were: STRN-Ex6-1F and PDGFRA-Ex12-2R and for nested PCR, STRN-Ex6-2F and PDGFRA-STRN-Ex12-1R. Primers used for the ETV6-PDGFR\(\alpha\) fusion were: ETV6-Ex6-1F and PDGFRA-Ex12-2R and for nested PCR ETV6.Ex6.2F and the same PDGFRA.STRN.Ex12.1R (see appendix I, section XVIII for all MRD primers). Peripheral blood for cases 1 and 2 was analysed approximately 28 and 9 months after starting imatinib, respectively. MRD was not detected using cDNA or genomic DNA for case 1, indicating the achievement of molecular remission. For case 2, analysis was only performed using genomic DNA. Of four replicate reactions, two were positive indicating low level disease, consistent with a major molecular response.

### 3.3.9 Discussion: PDGFRA fusions

Over the last decade, our understanding of the molecular pathophysiology of IHES and CEL has advanced dramatically. PDGFRA fusions are the most commonly associated with CEL, with the most frequent being FIP1LI-PDGFR\(\alpha\) which is seen in up to 12% of cases provisionally diagnosed as IHES (Cools, et al 2003, Griffin, et al 2003). Using bubble-PCR analysis I have characterised the genomic breakpoints in two CEL cases with chromosome 4q rearrangements and identified STRN at 2p22 and ETV6 at 12p13 as novel PDGFRA fusion partners. This increases the number of known PDGFRA partner genes to six, with PDGFRA being the tenth tyrosine kinase reported to fuse to ETV6. It has not been possible to estimate the frequency of these abnormalities since these fusions were referred specifically for investigation of 4q rearrangements and were not part of a series that underwent systematic cytogenetic investigation. However it is thought they probably only account for < 1% of CEL cases.
*STRN* encodes striatin, a 780 amino acid protein. It is a member of the calmodulin-binding WD repeat family of proteins. It contains four protein-protein interaction domains, a caveolin binding domain (55-63 amino acids), a coiled-coil domain (70-116 amino-acids), a calcium dependent calmodulin-binding domain (149-166 amino acids) and a series of WD repeat motifs (419-780 amino acids) (22). Striatin is thought to be a scaffolding protein that targets proteins such as the oestrogen receptor to the cell membrane allowing the assembly of other proteins required for rapid activation (Lu, *et al* 2004). Interestingly, in case 1 the fusion identified between intron 6 *STRN* to a truncated exon 12 of *PDGFRA*, results in the loss of a large region of WD repeat motifs within striatin. However, the N-terminal coiled-coil domain remains and it is this domain that could act as a necessary oligomerisation motif, constitutively activating the PDGFRα-derived tyrosine kinase moiety. Alternatively, truncation of the WW-like domain may be the critical activating event (Figure 3.26) (Stover, *et al* 2006).

*ETV6* is a member of the *ets* family of transcription factors and is frequently rearranged in both myeloid and lymphoid malignancies (Bohlander 2005). It was first cloned by Golub *et al* (1994) as a result of an *ETV6-PDGFRB* fusion gene in a patient with chronic myelomonocytic leukaemia (CMML) (Golub, *et al* 1994). More then 40 distinct translocations involving ETV6 have been identified cytogenetically with greater then 20 partners described so far (Odero, *et al* 2001, Rowley 1999). *ETV6* fusion partners can be categorised loosely into fusions with 1) protein tyrosine kinases; 2) transcription factors and 3) unproductive fusions, where ectopic expression of nearby genes not normally expressed within the haematopoietic system are thought to be expressed (Bohlander 2005, Cools, *et al* 2002). There is also evidence that *ETV6* functions as a tumour suppressor gene in *RUNX-ETV6* positive childhood ALL, where in 70% of cases the wild-type or normal *ETV6* allele is deleted. In cases where a deletion is not observed a point mutation can be found that abrogates *ETV6* expression (Cave, *et al* 1997, Griesinger, *et al* 2002, Montpetit, *et al* 2002, Patel, *et al* 2003).
BCR was the first described PDGFRα fusion partner. BCR-PDGFRA was amplified in two patients with an atypical chronic myeloid leukaemia (aCML) and a t(4;22) (Baxter, et al 2002). Two more cases with varied haematological malignancies have since been discovered and whilst the fusion breakpoints are diverse in BCR (exons 1, 7, 12 and 17) all show fusion to PDGFRA exon 12, ensuring the entire catalytic domain of PDGFRα is retained (Safley, et al 2004, Trempat, et al 2003). Analysis of FIP1L1-PDGFRA breakpoints show FIP1L1 breakpoints are widely clustered in exons 7–10 and consistently result in truncation of PDGFRA between the two conserved tryptophan residues (the WW domain) of exon 12 (Cools, et al 2003, Griffin, et al 2003, Klion, et al 2003, La Starza, et al 2005, Pardanani, et al 2003, Roche-Lestienne, et al 2005, Vandenberghe, et al 2004). Consistent with all other reported PDGFRA fusions, I found that STRN disrupts the region encoding the negative regulatory WW-like domain of PDGFRα. In contrast, for ETV6-PDGFRA the WW-like domain is intact (Figure 3.29). This is the first case of a PDGFRA fusion where disruption of PDGFRA exon 12 within the WW domain has not occurred. The ETV6-PDGFRA fusion fuses ETV6 exon 6 to the entire length of PDGFRA exon 12. In a recent paper by Stover and colleagues the significance of the breakpoint restrictions in PDGFRA was demonstrated by the production of a truncated exon 12 PDGFRA construct. Despite the absence of any partner gene, truncation of the PDGFRα WW-like domain was sufficient to activate the kinase and transform BaF/3 cells to growth factor independence. This paper may in part explain why the ETV6-PDGFRA fusion is still leukaemogenic, despite the fact that the breakpoint in PDGFRA does not disrupt the WW domain. An artificial ETV6-PDGFRA fusion that retained the region encoding the WW-like domain transformed Ba/F3 cells to growth factor independence, indicating that autoinhibition by the WW-like domain can be overcome by enforced homodimerisation. However the growth of the transformed cells was slower than those transformed by similar fusions in which the WW-like domain was truncated (Stover, et al 2006). It remains unclear why the WW-like domain is truncated with other fusion partners that provide an oligomerisation domain such as BCR, and also why truncation of this domain is not usually seen for fusions involving PDGFRβ.
Both patients are now being treated with imatinib. Case 1 has achieved complete chromosomal and molecular remission whilst on treatment, and Case 2 has been treated with low dose imatinib (100mg/day) which has induced a rapid haematological improvement however, follow up is currently too short to assess cytogenetic or molecular remission. In summary, I have identified two novel PDGFRA fusion genes in male patients presenting with an imatinib-responsive eosinophilic myeloproliferative disorder (MPD). Like PDGFRB, PDGFRA also fuses to diverse partner genes in CEL.

3.4 A NOVEL FGFR1 FUSION GENE

A common theme among the myeloproliferative disorders is the diverse tyrosine kinase fusion genes. In addition to those involving PDGFRA and PDGFRB, recurrent translocations involving chromosome band 8p11 and the tyrosine kinase FGFR1 are also described in myeloid malignancies. The 8p11 myeloproliferative syndrome (EMS), also known as stem cell leukaemia-lymphoma (SCLL) is a chronic MPD that commonly presents with eosinophilia, T-cell proliferation and acquired reciprocal chromosomal translocations with recurrent breakpoints at chromosome band 8p11 that disrupt the receptor tyrosine kinase FGFR1. To date 8 fusion genes have been identified in patients with EMS. The translocations are t(8;13)(p11;q12), t(8;9)(p11;q33), t(6;8)(q27;p11), t(8;22)(p11;q22), t(8;19)(p12;q13.3), t(8;17)(p11;q23), t(7;8)(q34;p11) and t(8;12)(p11;p11) which fuse ZMYM2 (ZNF198/RAMP/FIM), CEP110, FGFR1OP (FOP), BCR, HERV-K, MYO18A, TRIM24 (TIF1) and FGFR1OP2 to FGFR1, respectively (Belloni, et al 2005, Demiroglu, et al 2001, Grand, et al 2004a, Guasch, et al 2003, Guasch, et al 2000, Popovici, et al 1999, Walz, et al 2005, Xiao, et al 1998). Imatinib however is inactive against FGFR1 and cannot benefit patients with EMS. Other compounds that target FGFR1 have been developed including SU5402, PD173074, and TKI258. They function by blocking the ATP binding site in a manner analogous to imatinib, but none are currently in clinical use (Chase, et al 2007, Demiroglu, et al 2001,
Mohammadi, et al 1997). These compounds have been shown to specifically inhibit the growth of ZNF198-FGFR1 and BCR-FGFR1 transformed cells, resulting in increased levels of apoptosis (Chase, et al 2007). In this study, I have investigated a patient that presented with EMS and an acquired chromosomal rearrangement of 8p11, resulting in the fusion of a new partner gene.

3.4.1 Patient details

The patient referred was a 75-year-old female who presented with cervical, left axillary and inguinal lymphadenopathy. A computerised tomography (CT) scan showed right pleural effusion, ascites (abnormal build up of fluid around the abdomen), splenomegaly, and para-aortic nodes. Axillary node biopsy showed effacement of normal architecture by a diffuse population of mononuclear cells. The majority expressed strong surface CD3, but there was a scattered population of cells expressing neutrophil elastase amongst other myeloid markers. The presenting blood count showed haemoglobin of 140 g/L, WBC 16x10^9 L^-1, with neutrophilia (12x10^9 L^-1) and eosinophilia (0.8x10^9 L^-1) plus platelets of 375x10^9 L^-1. Hypercellular bone marrow was observed with an increase in eosinophils and eosinophil precursors but no increase in lymphoid cells or blasts. Cytogenetic analysis revealed 46,XX,t(8;12) (p11;q15)[19]/46, XX. There was rapid clinical deterioration accompanied by pyrexia (fever), lethargy, weight loss, and a widespread rash (biopsy was inconclusive on histology). Intensive chemotherapy was not offered and although there was a transient clinical response to steroids, the patient died 10 weeks later.

3.4.2 Fluorescence in situ hybridisation

FGFR1 FISH had been performed and previously reported as rearranged in this patient (Sohal, et al 2001). The breakpoints for all cases involving the FGFR1 gene thus far reported are tightly clustered within or close to exon 9 of this gene, further investigation was made easier. I explored the possibility that FGFR1 was rearranged by 5’RACE PCR.
3.4.3  5’Rapid Amplification of cDNA Ends (5’RACE) PCR

The partner gene was identified using the Gene-Racer™ Kit (Invitrogen). Briefly, this involves ligating a known sequence onto the 5’ end of the unknown mRNA, enabling subsequent PCR to be performed using an oligonucleotide that is complementary to the ligated sequence. A single RNA sample (E0003) was available for the t(8;12) patient and this was analysed using 5’RACE PCR primers and nested FGFR1 primers as previously described (Grand, et al 2004a). I performed the 5’RACE PCRs with the High Fidelity PCR Master kit (Roche Diagnostics) enabling amplification of up to 3 kb with an annealing temperature of 66°C, according to manufacturer instructions. Products were cloned with the TOPO TA Cloning Kit for Sequencing (Invitrogen) and sequenced. As shown on Figure 3.30, sequencing of the rearranged 5’RACE band revealed a fusion between CPSF6 intron 8 (accession number NM_007007) and FGFR1 at nucleotide 1272 from ATG (accession number M34185). The CPSF6 breakpoint within intron 8 results in the inclusion of a portion of the CPSF6 intron in the fusion mRNA, this fragment of intronic CPSF6 maintains the reading frame with FGFR1 and encodes 11 amino acids.

![Figure 3.20](image)

**Figure 3.20:** Electropherogram (top) and sequence (bottom) of the CPSF6-FGFR1 mRNA fusion determined by cloning and sequencing a 5’RACE PCR product. Region highlighted in red indicates incorporated CPSF6 intronic sequence.
3.4.4 Confirming the CPSF6-FGFR1 fusion

To confirm the CPSF6-FGFR1 fusion, a 224 bp product was specifically amplified from patient cDNA but not from control cDNA. The genomic breakpoint product of 543 bp was also co-amplified in the same reaction and sequenced (Figure 3.31A). A reciprocal FGFR1-CPSF6 fusion product was not detected with either single step or nested PCR. The CPSF6-FGFR1 fusion mRNA was confirmed by sequencing. The genomic breakpoint was confirmed by sequencing the larger amplified 543bp band, this is the same as the mRNA fusion due to inclusion of CPSF6 intronic sequence (detailed below Figure 3.31B).

![Figure 3.31:](image)

**Figure 3.21:** (A) Amplification of the CPSF6-FGFR1 fusion specifically from cDNA in the t(8;12) patient (E0003) but not the normal controls. The reciprocal fusion product was not detected; (B) Sequences surrounding the genomic breakpoints in CPSF6 and FGFR1 (bold).

3.4.5 Consequence of the t(8;12)

The t(8;12) fuses CPSF6 intron 8 to FGFR1 exon 10 and is the ninth reported FGFR1 fusion gene. The CPSF6-FGFR1 mRNA fusion is predicted to encode 5171 bp with an open reading frame (ORF) of 2688 bp including necessary intronic CPSF6 sequence to
maintain the ORF. The chimaeric protein is predicted to encode an 896 amino acid protein of which 501 amino acids are derived from CPSF6 and 395 amino acids from FGFR1. The predicted molecular weight of the protein is 223.5 kDa. The chimaeric protein retains the N-terminal region and the RNA recognition motif (RRM) from CPSF6 fused to the entire tyrosine kinase domain of FGFR1 (Figure 3.32).

**Figure 3.22:** Domain structure of FGFR1, CPSF6, and the predicted fusion protein. Relevant protein domains are detailed: RNA recognition motif (RRM), nuclear targeting domain for CPSF6 and the transmembrane domain, and tyrosine kinase domain for FGFR1.

### 3.4.6 Discussion

*CPSF6* is the ninth partner gene found fused to *FGFR1*. All *FGFR1* fusions to date involve a breakpoint in exon 9 which disrupts the FRS2 binding domain (Hoch and Soriano 2006). No recognisable oligomerisation motifs are identifiable in CPSF6, however it has been demonstrated previously that RNA recognition motifs (RRM), such as the one retained in CPSF6-FGFR1, may mediate homodimerisation (Handa, *et al* 2006, Simpson, *et al* 2004) and thus may activate the fusion protein.

CPSF6 is a cleavage and polyadenylation specificity factor. It plays a major role in pre-mRNA processing and has a domain organisation (N-terminal RNA binding
motif, a proline-rich middle and a C-terminal alternating charge domain that is arginine rich) similar to the superfamily of arginine/serine (R/S)-rich splicing factors which are also involved in pre-mRNA splicing. These R/S splicesomal proteins are thought to bind pre-mRNA and tether other proteins to the RNA via these R/S-domains and commit the pre-mRNA to splicing. It has been suggested that CPSF6 could interact via its alternating charge domain with one or more R/S proteins or R/S related splicing factors (Caceres, et al 1997). In conjunction with this CPSF6 has also been found in purified splicesomes (Dettwiler, et al 2004, Rappsilber, et al 2002, Zhou, et al 2002). During pre-mRNA processing, CPSF6 binds the highly conserved polyadenylation site (AAUAAA) that is present in almost every mRNA precursor, as well as interacting with the poly-A binding protein II (PABII) that binds the growing poly-A tail. Together CPSF6 and PABII activate poly (A) polymerase (PAP), by directing the enzyme to the RNA allowing polyadenylation of the upstream cleavage product (Ruegsegger, et al 1998).

Interestingly, SFPQ and NonO are also found purified in splicosomes along with CPSF6 (Dettwiler, et al 2004, Rappsilber, et al 2002, Zhou, et al 2002). SFPQ has recently been described as a partner protein found fused to the Abl tyrosine kinase in a patient with B-cell progenitor acute lymphoblastic leukaemia (ALL) (Hidalgo-Curtis, et al 2008). It is required early in splicesome formation and is a pre-mRNA splicing factor that binds the poly-pyrimidine tract of mammalian introns. The localisation of both CPSF6 and SFPQ suggests a possible interrelated function in transcription and processing of pre-mRNAs of these proteins. Of further interest is another FGFR1 fusion partner ZMYM2 (ZNF198) (Xiao, et al 1998), that immunoprecipitates with SFPQ (Kasyapa, et al 2005). ZMYM2 is thought to act as a possible scaffold protein bringing RNA-binding proteins together to facilitate pre-mRNA splicing (Kasyapa, et al 2005). In addition, FIP1L1, which fuses to PDGFRα in chronic eosinophilic leukemia (Cools, et al 2003), also plays a role in polyadenylation and pre-mRNA splicing (Kaufmann, et al 2004, Preker, et al 1995). All these interconnected partner proteins found fused to tyrosine kinases in human leukaemia are involved in pre-mRNA processing, the precise reasons for the recurrent
fusion to FGFR1 of this sub-group of interacting partner proteins involved with transcription elongation is still unclear (Hidalgo-Curtis, et al 2008).

3.4 Discussion

In this chapter I describe the discovery of four new partner proteins found fused to three different receptor tyrosine kinases in patients diagnosed with a MPD. All four novel fusions represent rare chromosomal abnormalities with only one of these fusions being recurrent (GOLGA4-PDGFRB).

The identification of rare fusions involving tyrosine kinase genes is clearly important as they are likely to be responsive to targeted therapy with imatinib or other compounds. The functional role of the partner protein is often overlooked, as well as providing dimerisation motifs accumulating evidence suggests that these partner proteins may play additional roles modifying intracellular signalling and disease phenotype (Million, et al 2004).

Groups of partner proteins with similar functions and interactions appear to fuse to tyrosine kinases more frequently than expected. Examples include proteins involved in pre-mRNA processing (Hidalgo-Curtis, et al 2008), endocytosis (this thesis) and a number that are targeted to the centrosome (Delaval, et al 2005b). Partner proteins provide both oligomerisation and targeting motifs for the fusion protein. The FIG-ROS fusion found in glioblastoma and the FGFR1OP-FGFR1 fusion found in patients with 8p11 myeloproliferative syndrome (EMS) are directed by the partner protein to specific sub-cellular compartments, including the centrosome and the Golgi (Charest, et al 2003, Delaval, et al 2005b). In the case of FIG-ROS, this is thought to be essential for the transforming ability of the fusion.

Delaval et. al., (2005) describe the targeting of FGFR1OP-FGFR1, an oncogenic tyrosine kinase, to the centrosome thus activating signalling pathways at this organelle allowing cells to sustain continuous cell cycles. This group have also
identified a second \textit{FGFR1} partner gene localised to the centrosome, and suggest that the fusion partners may not only provide dimerisation domains but also target oncogenic kinases to a specific cellular locations necessary for transformation. They hypothesise that these proteins may exert their oncogenic activity through dysregulation of cell processes associated with the centrosome. Thus further research into the subcellular localisation of oncoproteins indicates that emerging links between partner protein function and localisation may not all be coincidental. Numerous studies have been performed that look specifically at the subcellular localisation of constitutively activated kinases, Wetzler et. al., (1993) investigated the BCR-ABL protein in leukaemic cells and found the fusion resides exclusively in the cytoplasm, unlike the wild type ABL protein that is only found in the nucleus (Wetzler, \textit{et al} 1993). In addition, Ollendorf \textit{et al}., (1999) whom further characterised the FIM-FGFR1 fusion product as a result of the t(8;13) translocation, studied the subcellular localisation of this fusion protein, and in comparison to wild type FIM, they found that it was also entirely cytoplasmic (Ollendorff, \textit{et al} 1999). It is thought that localisation of these fusion proteins to the cytosol is enough to abrogate their entry into the endocytic pathway, and hence prevent lysosomal degradation.

Functionally, the largest related group of partner proteins is those linked to intracellular trafficking and endocytosis. There is increasing evidence of a relationship emerging between endocytosis and leukaemia. Removal of specific receptors from the cell surface by endocytosis and their degradation in lysosomes is an established method for the attenuation of receptor signalling. Genes encoding proteins associated with the endocytic pathway are repeatedly being identified as targets of chromosomal rearrangements in human haematological malignancies. The identification of \textit{GOLGA4-PDGFRB}, adds further weight to this emerging relationship. It is possible that subversion of endocytic control, possibly caused by such oncogenic fusion proteins, could play an important role in hyperproliferative conditions.
Further evidence of the role of endocytic pathways in haematological malignancies is becoming increasingly apparent. At least three reported translocations involving PDGFRB, identified in patients with aCML/CMML (Table 3.2), make in-frame fusions with components of the endocytic machinery. These include Rabaptin-5, Hip1 and PDE4DIP. Rab5 in its active GTP-bound form recruits a number of effectors to the endosome membranes. Rabaptin-5 is proposed to function as molecular linker between Rab5 and Rab4 to coordinate endocytic and recycling traffic (Magnusson, et al 2001, Vitale, et al 1998). Zhu et. al., (2004) has also recently shown through biochemical and functional analysis that the Rab5-Rabaptin-5 complex abrogates endosome fusion if disrupted (Zhu, et al 2004). The Rabaptin-5-PDGFRB fusion results in a loss of the Rab5 binding site and thus might obstruct formation of the Rab5-Rabaptin-5 complex, consequently halting endosome fusion.

A second relevant fusion is Hip1-PDGFRB (Ross, et al 1998). Huntington-interacting protein 1 (Hip1) was the first component of the endocytic machinery to be directly implicated in tumour formation. It interacts with AP-2, phosphoinositides and clathrin and is thought to play a role in clathrin-mediated endocytosis. Over-expression of Hip1 causes cell transformation and increased cell proliferation and the transformed cells also appear to have less clathrin at the plasma membrane. Hyun et. al., (2004) also found that over-expression of Hip1 inhibited the degradation of ligand-stimulated growth factor receptors. Transient transfection of 293T cells with Hip1 stabilised pools of PDGFRB following ligand-induced endocytosis, thus it is proposed that Hip1 may stabilise growth factor receptor levels via altered intracellular trafficking. Hip1 also contains an ENTH (epsin N-terminal homology) domain, that like other ENTH domains binds predominantly to phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P₂), and is important in regulating clathrin-mediated endocytosis. Hyun et. al., (2004) suggest that the ENTH domain may be the factor involved in regulating growth factor endocytosis and thus could be critical in the protein’s ability to promote cell growth or survival (Hyun and Ross 2004). This has been further supported by work done using a mutant construct lacking the ENTH domain, which when transfected into cells induced apoptosis. The lipid binding
<table>
<thead>
<tr>
<th>Protein</th>
<th>Proposed function in receptor downregulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eps15</td>
<td>Ubiquitin-binding protein found in PM clathrin-coated pits and in clathrin coats in sorting endosomes. Possible role in receptor recruitment.</td>
</tr>
<tr>
<td>Endophilin-2</td>
<td>Putative lysophosphatidic acid acyl transferase. Binds dynamin-2 and synaptotagmin. Involved in formation of clathrin-coated vesicles. Member of the membrane-bending/GTPase BAR domain.</td>
</tr>
<tr>
<td>Rabaptin-5</td>
<td>Effector of Rab5. Complexed to the Rab5 GDP/GTP exchange factor Rabex5. Required for endosomal fusion.</td>
</tr>
<tr>
<td>Hip1</td>
<td>Homologue of End4/Sla2, which interacts with actin-binding proteins and is required for endocytosis in Saccharomyces cerevisiae. Role in clathrin-mediated endocytosis. Interacts with AP-2, clathrin and phosphoinositides</td>
</tr>
<tr>
<td>PDE4DIP/Myomegalin</td>
<td>Functions as an anchor to localise components of the cAMP-dependent pathway to the Golgi/centrosome. The cAMP-dependent pathway is thought to be essential for intracellular signalling.</td>
</tr>
<tr>
<td>BIN2</td>
<td>Member of the BAR domain proteins. BAR domain drives membrane curvature, implicating a role in vesicular trafficking. BAR also shares significant homology to GTPase binding domain of Arfaptin 2, an effector of Rho-and Arf-like GTPas. Arfaptin 2 is known to be involved in intracellular transport, especially endocytosis.</td>
</tr>
<tr>
<td>WDR48</td>
<td>WD-repeat endosomal protein.</td>
</tr>
<tr>
<td>Golgin-245</td>
<td>Golgin family of coiled-coil proteins, associated with the Golgi apparatus. Plays a role in tethering events in membrane fusion and as structural supports for Golgi cisternae (subcompartments of the TGN). May also function as a vesicle coat.</td>
</tr>
<tr>
<td>KIF5B</td>
<td>Important for microtubule-based endosome-motility, by regulating early endosome trafficking. If inhibited reduced endosome motility and vesicle fission is observed.</td>
</tr>
<tr>
<td>FIG</td>
<td>Encodes a protein that peripherally associates with the golgi apparatus, and likely plays a role in golgi function.</td>
</tr>
</tbody>
</table>

**Table 3.2:** Components of endocytic machinery found as partners of oncogenic fusion proteins.
characteristics of the ENTH domains were also analysed and unusually found to preferentially bind 3-phosphoinositides not (PtdIns(4,5)P$_2$). Thus, it is suggested that Hip1 acts downstream of receptor uptake to inhibit trafficking of receptor tyrosine kinases to the lysosome for subsequent degradation. It is speculated that this occurs by binding inositol lipid determinants of the early endosome by their ENTH domains, acting to either stabilise the early endosome, favouring continued signalling, or by stimulating growth factor recycling to the surface. This work certainly provides evidence of a strong link between a partner gene of PDGFRB and the targeting of the endocytic pathway.

Other examples of oncoproteins involving members of the endocytic pathway include AF-1p (Bernard, et al 1994) and EEN (So, et al 1997) both of which are partners found fused to the C-terminal region of the MLL gene. In addition, endocytic proteins fused to RTKs in other malignancies have been reported, in particular papillary thyroid carcinomas. Rabes et. al., (2000) observed a fusion between the RET receptor tyrosine kinase and GOLGA5 a golgi integral membrane protein (Bascom, et al 1999, Rabes, et al 2000). Further evidence to support the hypothesis that subversion of endocytic pathways may be important in leukaemogenesis will be described in Chapter 5.

Deregulated, constitutively activated tyrosine kinases are proving to be the main driving force behind the MPDs, as substantiated in this chapter with the identification of four new tyrosine kinase fusion genes: GOLGA4-PDGFRB, STRN-PDGFRA, ETV6-PDGFRB and CPSF6-FGFR1. Moreover, there is now accumulating evidence to suggest an additional role for the partner proteins in modifying intracellular signalling and disease phenotype. Functionally similar groups of partner proteins linked to the centrosome (Delaval, et al 2005a), transcription elongation (Hidalgo-Curtis, et al 2008) and intracellular trafficking (Hidalgo-Curtis et. al., submitted 2009) have been identified that are thought to contribute not only the dimerisation motif but may impair endocytosis and even direct the oncoprotein to specific intracellular compartments, all of which may be critical to their transforming ability.
4 IDENTIFICATION OF Deregulated Expression of Tyrosine Kinase Genes

4.1 Introduction

There is increasing evidence to suggest a role for deregulated tyrosine kinases in the pathogenesis of MPDs. It is particularly important to recognise patients with PDGFRα or PDGFRβ fusions as these cases are likely to be responsive to imatinib (Apperley, et al. 2002, Cools, et al. 2003a, Gotlib 2005, Jones and Cross 2004, Klion, et al. 2004, Pardanani, et al. 2003). Identification of patients who are positive for PDGFR fusions is therefore critical to appropriate clinical management. However these genes have been shown to fuse to diverse partners in MPDs and it seems likely that additional partner genes are yet to be discovered. Furthermore many of these cases are very rare and the breakpoint diversity of many fusions is unknown. Consequently it is difficult to screen for all PDGFR fusions.

Our research group has studied patients for in vitro sensitivity to imatinib, as mentioned above. A number of patients have been found that appear to be responsive to imatinib but are negative for the known common fusions (e.g. FIP1L1-PDGFRα and ETV6-PDGFRβ). This leaves a number of patients where the molecular lesion has not been characterised, but we suspect involves an imatinib responsive gene. Cytogenetic analysis may suggest that either PDGFRα or PDGFRβ might be involved, but some fusions are cytogenetically cryptic (e.g. FIP1L1-PDGFRα) and several translocation targets have been identified on chromosome 5q that are distinct from PDGFRβ, for example FLT4 fusions found in a minority of MDS patients (Armstrong, et al. 1993). Thus cytogenetics alone cannot be used to determine unambiguously whether a patient might be sensitive to imatinib or not. FISH may be used to look for rearrangements of specific genes, but these results may be misleading due to cryptic or complex rearrangements (Kulkarni, et al. 2000). Consequently there
is a need for new methods to detect PDGFRα and PDGFRβ fusion genes. Here I describe the design and application of two generic RT-PCR assays to detect over-expression of PDGFR genes, and an mRNA-based Multiplex Ligation-dependent Probe Amplification (MLPA) test. The main aim was to search for new cryptic gene fusions that might be responsive to imatinib.

4.1.1 PDGFRB rearrangements

Disruption of PDGFRB was first described by Golub et. al., (1994), as a consequence of a t(5;12)(q33;p13), fusing the 5’end of ETV6 (otherwise known as TEL) to the 3’end, encoding the entire tyrosine kinase domain, of PDGFRB (Golub, et al 1994). Subsequently, twenty other translocations involving PDGFRB have been reported (See Table 1.1, Section 1.7.3) including three identified by our laboratory unpublished at present. The breakpoint for twenty out of the twenty one reported oncogenic fusions is within intron 10 of PDGFRB. To date, NIN-PDGFRB is the only reported fusion gene that exhibits a different genomic breakpoint (Vizmanos, et al 2004). The partner genes fuse in frame to the entire PDGFRB tyrosine kinase domain. Each partner protein contains an oligomerisation motif that dimerises the receptor tyrosine kinase, leading to aberrant phosphorylation of downstream proteins promoting growth and resistance to apoptosis.

Under normal conditions PDGFRB mRNA expression is tightly regulated, but when fused to the partner gene expression of the tyrosine kinase portion is under the control of the partner gene promoter and so may be over-expressed compared to the normal, full length PDGFRB mRNA. I developed a PCR-based assay designed to detect the expression of the tyrosine kinase domain of PDGFRB compared to a control band spanning the exons upstream and downstream of the common reported breakpoints.
4.1.2 PDGFRA rearrangements

Disruption of PDGFRA was first described by Baxter et. al., (2002), as a consequence of a t(4;22)(q12;q11), fusing the 5’end of BCR to the 3’end encoding the entire tyrosine kinase domain of PDGFRA. This is a rare variant translocation identified initially in only two patients with a CML-like MPD (Baxter, et al 2002). Five further rearrangements of PDGFRA have been identified and the breakpoints within PDGFRA have been found to be very tightly clustered in all cases, resulting in a truncated PDGFRA exon 12 (Baxter, et al 2002, Cools, et al 2003a, Curtis, et al 2007, Score, et al 2006, Walz, et al 2007). The most common PDGFRA rearrangement is the fusion of FIP1L1 to PDGFRA in HES patients described by Cools et. al., (2003). The fusion is generated by an interstitial deletion on chromosome 4q12. This fusion is cytogenetically cryptic but can be detected by FISH or RT-PCR (Cools, et al 2003a, Pardanani, et al 2003). Detection of this fusion is also complicated by the complex alternative splicing and variability of genomic breakpoints in FIP1LI, in addition to the low expression of the fusion gene in many cases and the significant variability in the proportion of cells involved in the malignant clone (Chung, et al 2006, Cools, et al 2003a). I designed a second assay in the same manner as that for PDGFRB, but for PDGFRA in an attempt to detect novel cryptic rearrangements. It was also hoped that this assay might address the limitation of current procedures in detecting FIP1LI-PDGFRA. I designed a second simple multiplex RT-PCR assay to determine the relative expression of the PDGFRA kinase domain which is retained in all fusions. This was compared to expression of the intact receptor, which is usually not detectable or weakly detectable by standard RT-PCR from peripheral blood or bone marrow leukocytes. Samples with a possible PDGFRA fusion gene were expected to over-express the kinase domain relative to the normal receptor mRNA.
4.2 Methods

4.2.1 Design of the PDGFRB RT-PCR assay

Three primers were designed, two that flank the region where the common breakpoints are reported (PDB Ex9/10.F and PDB Ex15.1R), and a third after the exon in which all the breakpoints are reported (PDBEx13.1F) (Figure 4.1). Primers were designed to anneal at 66°C and the PCR was undertaken in a volume of 10µl, due to limited cDNA available for each patient. The number of patients eligible for screening prompted me to perform all the amplifications in plate format or in 0.2ml strips tubes in a MJ Research Inc. thermocycler (Tetrad). After optimisation, the PCR protocol was finalised to: 0.4µM PDGFRB exon 9/10.F, 0.1µM PDGFRB exon 13.1F, 0.5µM PDGFRB exon 15.1R, 0.2mM dNTP, 1x buffer (Hotstar buffer containing 15mM MgCl₂), 2.5mM MgCl₂ (total), 0.4U Hotstart Taq DNA polymerase (Qiagen, West Sussex, UK). The PCR cycling parameters use a 15 minute 95°C hot-start to activate the enzyme and a 30 second, 66°C annealing temperature for stringency. The PCR was cycled 30 times.
Figure 4.1: A diagram representing the position of each of the primers in the multiplex, in relation to the common reported breakpoints in PDGFRB and to the tyrosine kinase and extracellular domains of PDGFRB.

A patient with normal PDGFRB expression is expected to generate two bands of equal intensity (assuming both PCRs worked with the same efficiency), and a patient with a rearrangement that results in over-expression of the region encoding the tyrosine kinase would exhibit a strong smaller band derived from both the normal receptor mRNA and the fusion mRNA, and a faint larger band derived solely from the normal receptor mRNA (Figure 4.2A).

![Diagram of PDGFRB RT-PCR assay](image)

**Figure 4.2:** PDGFRB RT-PCR assay, a schematic of the expected banding pattern. (A) The banding pattern expected from initial design of the PDGFRB multiplex, (B) Gel picture of observed bands from a number of patients and controls screened using the PDGFRB multiplex.

Figure 4.2B represents an example of a gel of the PDGFRB RT-PCR assay. Lane 1 (L1) represents a normal individual with an additional band visible at 2kb. I cloned and sequenced this aberrant band and found it contained intronic sequence, and was therefore a genomic DNA contaminant. Residual genomic DNA is often found in the cDNA. I altered the PCR conditions by reducing the amount of the 5' primer in the extracellular domain but was unsuccessful at eliminating this band. I also redesigned
the primer, finally resorting to the Ex9/10F primer. This primer was designed to sit across the exon boundaries of exon 9 and 10 to prevent amplification of any genomic DNA. Although this primer helped to reduce the intensity of this contaminating band it was not successful at eradicating it completely. With further work I showed the assay was still informative and that this band did not interfere with interpretation of the results, and can therefore be disregarded. Lane 2 (L2) represents my normal control, for which I cloned each of the fragments and combined them equally to create an artificial mixture. In this lane as expected, the 2kb genomic band is absent. Finally, lane 3 (L3) is my positive control, created again through cloning, from a patient with a novel PDGFRB rearrangement. The positive control shows only the bottom band, representing the tyrosine kinase portion of the gene. Some patients showed this same pattern, with the absence of the larger control band, others showed the differential expression expected (Lanes 1 and 4) (Figure 4.3).

![Agarose gel](image)

**Figure 4.3:** A representative agarose gel picture of patients screened by the PDGFRB multiplex. Lane 1 (L1) represents a positive patient where the tyrosine kinase band is present only; Lanes 2 and 3 (L2 and L3) are negative patients where both bands are clearly visible; Lane 4 (L4) is the positive control created from cloning of a patient with a
known PDGFRB fusion. Lane 5 (L5) shows the no DNA control, and Lanes 6, 7 and 8 (L6, L7, and L8) represent patients where the genomic band has out-competed the presence of the other bands and were classed as fails.

### 4.2.2 Design of the PDGFRA RT-PCR assay

Three primers were again designed, two that flank the region where the common breakpoints have been reported (PDA Ex10.1F and PDA Ex15.1R). The majority of breakpoints in PDGFRA have been reported in exon 12, therefore the third primer, designed to prime after the region of common breakpoints was located in exon 13 (PDA Ex13.1F) (Figure 4.4A). As before, all primers were designed to anneal at 66°C, the PCR volume was 10µl and the amplifications were in plate or strip-tube format, for 30 cycles, in a MJ Research Inc. thermocycler (Tetrad). After optimisation the PCR protocol was finalised to: 0.05µM PDGFRA exon 10.1F, 0.2µM PDGFRA exon 13.2F, 0.4µM PDGFRA exon 15.1R, 0.2µM dNTP, 1x buffer (Hotstar buffer containing 15mM MgCl₂), 2.5mM MgCl₂ (total), 0.4U Hotstart Taq DNA polymerase (Qiagen, West Sussex, UK). The PCR cycling parameter again uses a 15 minute 95°C hot-start to activate the enzyme and a 30 second, 66°C annealing temperature for stringency.
Figure 4.4: PDGFRA RT-PCR assay. (A) A schematic representing the position of each primer in the multiplex in relation to the common reported breakpoints in PDGFRA, (B) The banding pattern expected from initial design of the PDGFRA multiplex, (C) Gel picture of observed bands from a number of patients screened using the PDGFRA multiplex.

A patient with normal PDGFRA expression is expected to yield two bands of equal intensity and a patient with a rearrangement that results in over-expression of the tyrosine kinase would exhibit a strong smaller band and a faint larger band as before (Figure 4.4B). Lanes 1-3 show normal individuals, however only one band was present, and initially presumed to be the larger control band. On analysis of this band, it was clear the product was not the size expected from the primer pair. Upon cloning and sequencing I found that this band was again due to genomic DNA contamination. Alternate primers and differing PCR conditions failed to yield a correct sized control band with no intron, or the expression of both mRNA-derived bands in a normal individual. This is almost certainly because PDGFRA is not normally expressed or
expressed at very low levels in peripheral blood leukocytes. I therefore ordered a cell line, MG-63 that has been shown to express normal \textit{PDGFRA}. MG-63 cDNA produced, as expected, two bands of 650 and 290 bp (Figure 4.5). Although the lower band is stronger than the upper band, both are clearly visible. Eol-1 cells express \textit{FIP1L1-PDGFR\textalpha} and only the smaller 290 bp band derived from the region retained in the fusion is visible (Figure 4.5). Most normal individuals didn’t show either of the \textit{PDGFRA} mRNA-derived bands, but instead the 1049 bp genomic DNA fragment encompassing exons 13–15 was amplified (Figure 4.5). Although an unexpected product, this band served as a useful internal control for successful amplification. On screening positive controls I also discovered that although differential expression was not observed, the smaller band (290bp) was visible only in patients with confirmed \textit{PDGFRA} rearrangements (Figure 4.4C, Lane 4). Thus, the RT-PCR assay allowed identification of over-expression of the tyrosine kinase domain of \textit{PDGFRA}.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.5.png}
\caption{MG-63 cells express normal \textit{PDGFRA} and both 650 bp and 290 bp bands are visible. Eol-1 cells express \textit{FIP1L1-PDGFR\textalpha} but not normal \textit{PDGFRA} and only the smaller 290 bp band is visible. Normal individuals 1–3 express \textit{PDGFRA} very weakly and the predominant amplification product (1049 bp) is exons 13–15 from genomic DNA. Patient E683 shows the 290 bp band but not the 650 bp band, suggesting the presence of a \textit{PDGFRA} fusion gene (discussed in more detail later).}
\end{figure}
4.3 Results

4.3.1 Identification of patients with possible PDGFRB rearrangements

Having optimised the PDGFR RT-PCR assays I went on to screen a cohort of 199 MPD patients for over-expression of PDGFRB. Of these 18 were identified that were regarded as possibly positive for a PDGFRB rearrangement on the basis of increased intensity of the smaller band. Of these 18, 9 were strongly positive with markedly increased relative intensity (Table 4.1) and are considered in detail below.

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Diagnosis/Referral Reason</th>
<th>Previously identified fusion gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>P936</td>
<td>EMS</td>
<td>ZNF198-FGFR1</td>
</tr>
<tr>
<td>P981</td>
<td>MPD? 8p11</td>
<td>FGFR1OP-FGFR1</td>
</tr>
<tr>
<td>P997</td>
<td>MPD t(5;10)</td>
<td>CCDC6-PDGFRB</td>
</tr>
<tr>
<td>P1218</td>
<td>MPD t(5;12)</td>
<td>ETV6-PDGFRB</td>
</tr>
<tr>
<td>E285</td>
<td>MDS transformed t(5;12)</td>
<td>ETV6-PDGFRB</td>
</tr>
<tr>
<td>E287</td>
<td>HES</td>
<td>No</td>
</tr>
<tr>
<td>E354</td>
<td>Eosinophilia</td>
<td>No</td>
</tr>
<tr>
<td>E508</td>
<td>?HES/SM</td>
<td>No</td>
</tr>
<tr>
<td>E521</td>
<td>MPD</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 4.1: Patients observed to have PDGFRB over-expression using the PDGFRB multiplex RT-PCR assay.

The samples were screened blind and the multiplex was able to identify all three patients with previously characterised PDGFRB fusion genes (P997, P1218 and E285), of which two had been found to harbour ETV6-PDGFRB and the third CCDC6-PDGFRB Unexpectedly two patients with over-expression of PDGFRB had previously been found to have rearrangements of FGFR1 (P936 and P981). Finally,
four patients (E287, E354, E508 and E521) had been previously tested negative by RT-PCR for \( ETV6-PDGFRB \) and were investigated in more detail.

My initial analysis focused on patient E521 as this case had the most stored material. I performed FISH and 5’RACE and discovered the patient had a typical \( ETV6-PDGFRB \) rearrangement, despite the previous RT-PCR negative result. Only after I had identified the \( ETV6-PDGFRB \) fusion in this patient did we receive notification of the patient’s karyotype. The fusion is a consequence of a complex rearrangement involving chromosomes 2, 5 and 12 (Figure 4.6). It is not obvious from the karyotype that this fusion might exist but the involvement of 5q33 is suggestive of a \( PDGFRB \) rearrangement. After this discovery the original diagnostic RT-PCR was repeated and the patient sample was found to be positive. Why this fusion was missed initially is unclear however, he had been treated and responded to imatinib. Of the remaining three patients, E508 and E354 were studied extensively performing \( PDGFRB \) split apart FISH and 5’RACE to detect any possible \( PDGFRB \) rearrangements. No evidence for disruption of this gene was found. Very little material was available for study of the fourth case (E287).

---

**Figure 4.6:** Structure of the ETV6-PDGF\( \beta \) fusion with the usual karyotype associated with this fusion and the complex karyotype observed in the bone marrow of patient E521 by conventional cytogenetics

Gain-of-function point mutations have also been discovered in certain receptor tyrosine kinases that act to constitutively activate the kinase, such as the well
characterised V617F change in JAK2 associated with PV and other MPDs. Additionally, point mutations and small deletions in the activation loop and juxtamembrane domains of PDGFRA and KIT have been commonly reported in patients with gastrointestinal stromal tumours (Heinrich, et al 2003, Hirota, et al 1998) and more recently Chiara et. al, (2004) reported the identification of a mutation in the activation loop of the murine PDGFRB gene. The mutation conferred transforming ability to the embryonic fibroblasts (Chiara, et al 2004). Based on this, I screened two patients with available material over-expressing PDGFRB using denaturing High Performance Liquid Chromatography (dHPLC) and sequencing. I began by screening the key regions of PDGFRB, those encoding the juxtamembrane domain and the activation loop (exons 12-18). For both patients each exon was individually amplified and the PCR product mixed in a 1:1 ratio with normal placental DNA PCR product for each individual exon. The samples were then heated to 95°C and slowly allowed to reanneal on a reducing temperature gradient allowing the formation of heteroduplexes. When run through a chromatography column under partially denaturing conditions changes in sequence usually result in extra peaks or shifted peaks compared to normal controls. No activating mutations were identified in either patient in the juxtamembrane or activation loop of PDGFRB. I screened all the remaining exons of PDGFRB to exclude the possibility of a mutation in these regions (exons 1-11 and exons 19-23). Both patients lacked any sequence variants.

A colleague (Kathy Waghorn) had also designed a DNA Multiplex ligation-specific probe amplification (MLPA) assay for selected tyrosine kinase genes. MLPA is a probe based technique that allows multiple targets to be amplified with only a single primer pair (for more detail see Figure 4.8). Amplification of the probe is dependent upon the number of target sites present in the sample, permitting detection of genomic deletions and amplifications. Each probe is fluorescently labelled and has a unique length allowing multiple amplicons to be analysed in one reaction and separated by size by capillary electrophoresis. Comparison of the peak pattern obtained for a given sample against reference samples allows the relative quantity of each amplicon to be determined. The ratio obtained identifies the amount of target
sequence present in the sample DNA. The MLPA assay designed by Kathy included the \textit{PDGFRB} gene, and was used to look for possible duplications and deletions of specific tyrosine kinase genes. Following the \textit{PDGRFB} over-expression RT-PCR result, E508 (only patient with available DNA) was analysed using MLPA. No deletions or duplications were observed (Figure 4.7).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4_7.png}
\caption{MLPA traces of patient E508. MLPA traces showing (A) Normal control displaying full set of candidate tyrosine kinase genes in order: PDGFRA, MET, PDGFRB, RET, FGFR1, SYK, FLT3, ROS, BTK, RRP4, MST4, TRKB, EPHB4, MPL and FGFR3 (Note: \textit{BTK} is an X-linked gene used as control), (B) patient with \textit{FGFR1} duplicated proving its discriminatory power, (C) E508, set of four genes: \textit{EPHB4}, MPL, PDGFRB and MET, used when looking for rearrangements of a specific gene.}
\end{figure}
4.3.2 Identification of patients with possible PDGFRA rearrangements

A total of 270 cases with IHES or persistent unexplained eosinophilia were referred to the Wessex Regional Genetics Laboratory, Salisbury for routine molecular analysis between May 2002 and November 2005. Of these, 229 patients were screened using the PDGFRA multiplex. The patients were simultaneously screened for the more common FIP1L1-PDGFRA fusion using single and nested RT-PCR. All 12 cases in this study group who tested FIP1L1-PDGFRA positive by RT-PCR were also positive by the PDGFRA assay, that is, the 290bp band was clearly visible and the 650bp band was either not visible or was barely visible (Table 4.2), confirming the multiplex worked to identify rearrangements of PDGFRA. There were no cases where a FIP1L1-PDGFRA fusion determined by standard RT-PCR was not identified by my PDGFRA expression assay. Interestingly one patient (E293) turned out to be a sample from a FIP1L1-PDGFRA positive case that had been treated with imatinib and was therefore referred for minimal residual disease analysis. Both single and nested RT-PCR of FIP1L1-PDGFRA proved negative, however my assay observed over-expression of PDGFRA. This is unusual, as it is expected that the nested RT-PCR will be more sensitive than the PDGFRA assay. However, analysis of FIP1L1-PDGFRA breakpoints by a colleague has highlighted breakpoints in regions downstream of the common FIP1L1 breakpoint. Due to the position of the nested RT-PCR primers within PDGFRA, nested RT-PCR using standard primer sets would not be able to identify this case efficiently and thus my PDGFRA assay proved useful in this case for treatment analysis. The diagnosis of the FIP1L1-PDGFRA fusion by RT-PCR specific for the fusion itself is often very difficult and part of the reason may be the fact that some cases have aberrant breakpoints. Some patients are found to be positive only by nested PCR, even at diagnosis.
<table>
<thead>
<tr>
<th>Identifier</th>
<th>FIP1L1-PDGFRA RT-PCR result</th>
</tr>
</thead>
<tbody>
<tr>
<td>243 (^1)</td>
<td>Single step positive</td>
</tr>
<tr>
<td>293 (^1)</td>
<td>Negative</td>
</tr>
<tr>
<td>458</td>
<td>Nested positive</td>
</tr>
<tr>
<td>555</td>
<td>Single step positive</td>
</tr>
<tr>
<td>591 (^2)</td>
<td>Nested positive</td>
</tr>
<tr>
<td>606</td>
<td>Nested positive</td>
</tr>
<tr>
<td>614</td>
<td>Single step positive</td>
</tr>
<tr>
<td>636</td>
<td>Single step positive</td>
</tr>
<tr>
<td>646 (^2)</td>
<td>Nested positive</td>
</tr>
<tr>
<td>759</td>
<td>Single step positive</td>
</tr>
<tr>
<td>794</td>
<td>Single step positive</td>
</tr>
<tr>
<td>905</td>
<td>Single step positive</td>
</tr>
<tr>
<td>939</td>
<td>Single step positive</td>
</tr>
<tr>
<td>1025</td>
<td>Single step positive</td>
</tr>
</tbody>
</table>

\(^1\) Same patient  
\(^2\) Same patient

**Table 4. 2:** Validation of the PDGFRA multiplex. Patients that are FIP1L1-PDGFRA positive were also positive with the PDGFRA RT-PCR multiplex.

A further 14 patients were identified in the screen of the 217 FIP1L1-PDGFRA-negative samples to have over-expression of PDGFRA (~6%) (Table 4.3). These patients were not characterised molecularly and therefore the over-expression of PDGFRA observed by the multiplex offered a possible explanation to the pathogenetic cause of the patients’ leukaemia.
<table>
<thead>
<tr>
<th>Case number</th>
<th>Referral reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td>HES</td>
</tr>
<tr>
<td>200</td>
<td>Idiopathic HES</td>
</tr>
<tr>
<td>271</td>
<td>MPD HES</td>
</tr>
<tr>
<td>289</td>
<td>Eosinophilia</td>
</tr>
<tr>
<td>344</td>
<td>HES</td>
</tr>
<tr>
<td>349</td>
<td>HES</td>
</tr>
<tr>
<td>421</td>
<td>HES</td>
</tr>
<tr>
<td>459</td>
<td>HES</td>
</tr>
<tr>
<td>539</td>
<td>Eosinophilia</td>
</tr>
<tr>
<td>639</td>
<td>CML/MPD</td>
</tr>
<tr>
<td>680</td>
<td>CML</td>
</tr>
<tr>
<td>683/684</td>
<td>Eosinophilic leukaemia</td>
</tr>
<tr>
<td>748</td>
<td>MPD/MDS</td>
</tr>
<tr>
<td>843</td>
<td>MPD</td>
</tr>
</tbody>
</table>

Table 4.3: *FIP1L1-PDGFRA* negative cases that tested positive for the *PDGFRA* RT-PCR assay.

A colleague was carrying out work to characterise the exact genomic breakpoints of patients with *FIP1L1-PDGFRA* using bubble PCR (see section 2.2.6 for more details). Also included in this analysis was patient E683/E684, who had a complex rearrangement; 46XY, del(3)(p21), add(4)(q12),-10,13q?,+der(?)(?→cen→ :
?:4q12→ 4q28.3::10q11.2→ 10qter) and who also tested positive by the *PDGFRA* multiplex in the blood sample (E683) but, interestingly, not the bone marrow (E684). My colleague went on to discover a previously undescribed fusion involving a novel partner gene to *PDGFRA*. The partner gene is a member of the Kinesin family, *KIF5B* and is located at chromosome 10p11.22. The patient was treated with imatinib and six months later had achieved complete cytogenetic and molecular negativity.
(Score, et al 2006). At the time of publication this was the third PDGFRα fusion partner and highlighted the utility of this the over-expression assay to detect rare PDGFRα fusions.

Sufficient material was available for further analysis of six of the fourteen cases found to over-express the PDGFRα kinase domain. Because of the success of bubble-PCR in identifying PDGFRα rearrangements, I used this technique and found a rearrangement in one case with both restriction enzymes. Unexpectedly, sequencing of these rearranged bands revealed FIP1L1 sequence. Reanalysis of patient cDNA confirmed the presence of FIP1L1-PDGFRα mRNA that had been missed on initial analysis. The fusion was detectable by nested but not single-step RT-PCR and this case highlights the problems in reliably detecting FIP1L1-PDGFRα (Score, et al 2009). The reason for PDGFRα over-expression in the remaining cases remains obscure.

4.3.3 Over-expression assays: discussion

Detection of PDGFRα and PDGFRβ rearrangements is of increasing importance but both standard RT-PCR and FISH assays may miss positive cases. Through the blind screening of patients using the newly designed PDGFRα and PDGFRβ multiplexes, I have been able to prove that both assays work well to detect overexpression of these kinases. Both multiplex PCRs identified all patients with previously characterised PDGFR rearrangements.

To date, only one patient (E521) found positive by the PDGFRβ multiplex has actually proved to have a 5q33 rearrangement. The patient was clinically diagnosed as having a CMPD, that was later found to be as a consequence of a previously undetected t(5;12)(q33;p13). None of the remaining patients with over-expression of PDGFRβ have yielded identification of an oncogene or chromosomal rearrangement and the basis for over-expression remains ambiguous. Surprisingly two positive
patients were known to have rearrangements of the \textit{FGFR1} gene. The reason for this is unclear but it is known that active receptor tyrosine kinases activate complex intracellular signalling networks that result in a variety of cell responses including cell maintenance, mitogenesis, migration and differentiation. Thus it is possible that the \textit{FGFR1} fusions might activate \textit{PDGFRB}, however it is unclear why this should lead to specific over-expression of the kinase domain. Potentially this could be an artifact, for example if \textit{PDGFRB} in this context was alternatively spliced such that the normal full length product was not detected. Alternatively, it is known that \textit{PDGFRA} can be expressed from an internal promoter. (Mosselman, \textit{et al} 1994, Mosselman, \textit{et al} 1996) and it is possible that \textit{PDGFRB} might also have an internal promoter, transcripts from which might be scored in my assay as a potential fusion.

Of the 7 patients found to over express \textit{PDGFRA}, one patient who responded to imatinib was investigated in detail and found by bubble-PCR to have a previously undescribed fusion involving a novel partner gene, \textit{KIF5B} at chromosome 10p11.22. Little is known about \textit{KIF5B} however it is a member of the Kinesin superfamily. The kinesins are microtubule-based motor proteins that play important roles in organelle transport and cell division. They have an N-terminal motor domain which converts ATP into a motile force along the microtubules (plus-end direction only), a central alpha-helical rod domain enabling the protein to dimerise and a globular C-terminal domain, that allows interaction with possible organelle receptors. Xia et. al., (1998) have demonstrated in mice that \textit{KIF5B} is ubiquitously expressed, unlike many of the other family members that are expressed only in neuronal tissues (Xia, \textit{et al} 1998).

Confirmation of both the genomic breakpoint and the cDNA breakpoint were later performed, and the fusion mRNA demonstrated by single step PCR amplification. The complex rearrangement involving chromosomes 2, 4 and 10 is predicted to generate a chimaeric fusion protein that is structurally similar to other tyrosine kinase fusions. Therefore, it is likely that the \textit{KIF5B-PDGFRA} fusion deregulates haematopoiesis in a manner analogous to \textit{BCR-ABL} in CML patients. A schematic
representation of the normal and chimaeric fusion proteins is presented in (Figure 4.8).

Figure 4.8: Schematic representation of the normal PDGFRα, KIF5B and the KIF5B-PDGFRα fusion.

The PDGFRA RT-PCR assay has not only pointed the way to a new fusion but also proved to be extremely useful for detecting patients with FIP1L1-PDGFRA. This more common fusion is predominantly tested for diagnostically by RT-PCR using single and nested primers, however studies by a colleague have highlighted the problem of breakpoints further downstream of the common FIP1L1 breakpoint (Score, et al 2009). Clearly, further research is needed to design an optimal protocol for detection of FIP1L1-PDGFRA in patients at diagnosis or after treatment, and my data indicated that simple expression analysis may be very helpful in this regard. In
addition it seems likely that further screening of MPD patients for over-expression of PDGFRA and PDGFRB will identify new uncommon gene fusions.

Also frequently described in myeloid malignancies are recurrent translocations involving chromosome band 8p11. The 8p11 myeloproliferative syndrome (EMS), also known as stem cell leukaemia-lymphoma (SCLL), is associated with disruption of FGFR1. EMS is a CMPD that commonly presents with eosinophilia and associated T-cell proliferation. To date, nine fusion genes have been identified associated with translocations t(8;13)(p11;q12), t(8;9)(p11;q33), t(6;8)(q27;p11), t(8;22)(p11;q22), t(8;17)(p11;q23), t(7;8)(q34;p11), t(8;12)(p11;p11), t(8;19)(p12;q13.3) and t(8;12)(p11;q15) which fuse ZYMYM2 (ZNF198/RAMP/FIM), CEP110, FGFR1OP (FOP), BCR, MYO18A, TRIM24 (TIF1), FGFR1OP2, HERV-K and CPSF6 to FGFR1, respectively (Belloni, et al 2005, Demiroglu, et al 2001, Grand, et al 2004a, Guasch, et al 2000, Guasch, et al 2003b, Hidalgo-Curtis, et al 2008, Popovici, et al 1999, Walz, et al 2005, Xiao, et al 1998).

Imatinib is inactive against FGFR1 and cannot benefit patients with EMS. Nevertheless, other compounds have been developed such as SU5402, PD173074 and TKI258 that function by blocking the ATP binding site analogous to imatinib, but also possess anti-FGFR activity (Chase, et al 2007, Demiroglu, et al 2001, Mohammadi, et al 1997). These compounds have been shown to specifically inhibit the growth of ZNF198-FGFR1 and BCR-FGFR1 transformed cells, and are therefore more likely to be important in the treatment of EMS patients, although currently no compound is currently available for clinical use.

To encompass the FGFR1 gene I first thought about designing another RT-PCR assay analogous to those for the PDGF receptors to assess over-expression in uncharacterised MPD patients. However, this would mean testing every patient three times for each of the assays, so instead I designed an MLPA mRNA detection/quantification assay. MLPA has a further advantage in that it has the potential to be expanded further to include other targets.
4.4 *Multiplex Ligation-dependent Probe Amplification (MLPA)*

MLPA is a probe-based technique used routinely to detect genomic deletions and amplifications (Schouten, *et al* 2002) (Figure 4.9) however it can also be used for mRNA profiling. It is a technique capable of establishing the relative copy number of up to 45 nucleic acid sequences in one single reaction by yielding amplification products of unique size per target. Products are separated by automated fluorescent-based sequence electrophoresis and the copy number of each target is determined from the relative intensities of the products compared to those obtained from controls. As only one pair of PCR primers is used, MLPA results in a very reproducible pattern. Furthermore, MLPA is a fast technique that is very simple to perform and needs only 20ng of DNA. This seemed like an ideal procedure to screen for three tyrosine kinase genes in one assay. In addition more genes can easily be added and it is easily adapted to screen large numbers of patients.

![Figure 4.9: A schematic representation of Multiplex Ligation-dependent Probe Amplification (MLPA). Target A represents a normal MLPA trace, the probe is represented in black, the ‘stuffer’ sequence in grey (alters the size), and the primers are shown in red](image-url)
and green. The 5’ primer (red) is fluorescently labelled. The probes are first hybridised to the target sequence, these are then ligated together and finally amplified. The amplified products are analysed on an ABI 3100 genetic analyser, a representative trace is shown at the bottom. Target B represents a control where some of the peaks are deleted, the star shows an example of a reduction in peak height compared to the normal.

4.4.1 Methods – Probe design

MLPA depends on the use of progressively longer oligonucleotide probes in order to generate locus-specific amplicons of increasing size that can be resolved by electrophoresis. Schouten et. al., 2002 originally designed these long oligonucleotides using a series of proprietary M13-based vectors, however it is now possible to chemically synthesise long oligonucleotides (Schouten, et al 2002). I used a number of rules designated for the design of synthetic MLPA probes published by MRC Holland (www.mrc-holland.com). An important consideration when designing synthetic probes is that the second probe of each pair must be 5’ phosphorylated and significant variation in the quality of long oligonucleotides and the percentage phosphorylation between batches has to be expected.

For each gene I had to design four probes, two that abutted each other within the region encoding the tyrosine kinase domain and two that also abutted each other but were located in the region encoding the extracellular domain. The extracellular peak was then compared to the intracellular peak to determine if there was any over-expression of the kinase domain (retained in gene fusions) compared to the extracellular domain (only present in full length normal transcripts), compared to that seen in controls. As my MLPA assay was based on cDNA I needed to consider specifically how I might design the probes so they would detect only cDNA and not genomic DNA. It was recommended that these probes had an exon boundary close to the ligation site (within 7 nucleotides), thus genomic DNA sequences will not be detected as an intron separates the hybridisation of these two probes. A number of other basic rules have been proposed for optimal probe design: 1) The length of the
two immediately adjacent hybridising sequences must be greater than 21 nucleotides with a predicted melting temperature of more than 70 °C; 2) the first nucleotide following the PCR primer-sequence should preferably not be an adenine as this results in low peak size of the probe amplification product. Specifically the peak size increases in the order A<T<G<C with the difference between A and C greater than two-fold; 3) no more than three G or C residues adjacent to the ligation site; 4) probe target sequences must not overlap as probes will compete for binding and 5) the probes should be 45-60% GC content. An example of one of my MLPA cDNA oligonucleotides is detailed below (Figure 4.10) showing each of the components.

![Figure 4.10: Schematic representing the FGFR1 intracellular domain probe. The target sequence spans exons 12 and 13. The total size of the MLPA oligonucleotide is 95bp, this includes the complement primer sequences (labelled primer sequence in blue, unlabelled primer sequence in purple), the stuffer sequence (green) and the adjacent target probe sequences (red and orange).](image)

The primer sequences for all MLPA cDNA oligonucleotides are exactly the same allowing amplification of all probes in a single reaction. In addition, only one primer (5’primer) is labelled and this is always with 6-FAM which enables the peaks to be distinguished using fluorescent based capillary electrophoresis. The primer sequences are the same as MRC Holland enabling us to be able to use the kit components from this company. The stuffer sequence differs in size enabling each of the probes to be differentiated by size on electrophoresis. The stuffer sequence must fit with the rules
detailed above and not hybridise to the adjacent target sequence. All other probes are
detailed in Appendix I, section XIX.

4.4.2 Methods – Optimisation of the assay

It was anticipated that when samples with known fusion genes were tested the
tyrosine kinase portion would appear amplified compared to the extracellular domain.
My assay contained 8 probes, two each for the tyrosine kinase genes: *FGFR1*,
*PDGFRB*, *PDGFR* and *ABL*. *ABL* was included partly because it is routinely used
as a control for real time quantitative PCR assessments of minimal residual disease
but also because it is a target of leukaemia-associated translocations. These were
added to MRC-Holland MLPA kit components and then used to test sample cDNA
according to the manufacturer’s protocol (see section 2.2.9). The probes were
prepared before use. The cDNA MLPA oligonucleotides arrive lyophilised and were
initially dissolved in sterile dH₂O and diluted to achieve a final concentration of
100pmol/μl. To prevent most of the PCR primers being consumed, low amounts of
the probe oligonucleotides must be used and the suggested working concentration of
each primer per reaction is between 1-4 fmol. To determine the amount of each probe
needed the following calculations were made so a stock solution containing all 16
probes could be prepared for use in the MLPA reaction:

1) \[1.33 \text{ fmol} \times 200\mu l = 2.66 \mu l \text{ (volume of primer required)}\]
   Stock concentration (100pmol/μl)

2) \[2.66\mu l \text{ of each probe (16 X 2.66} \mu l = 42.56\mu l\] and make up to 200μl with
   157.44μl sterile dH₂O

3) A 1:50 dilution of this probe stock is then required (20μl probe stock + 980μl
   sterile dH₂O)
4) Followed by a 1:10 dilution of step 3 (10µl from 1:50 dilution + 90µl sterile dH₂O) to make a final 1:500 dilution. This stock solution is used for the MLPA reaction.

To optimise the assay, available cell line cDNA with known fusions was first used along with a number of normal cDNA controls to enable me to identify any probes that gave false positive or negative results. The cell lines used were EOL-1, K562, SD1, Jurkat, HL60, HEL, and 5637. EOL-1 is an AML cell line which carries FIP1L1-PDGFRα; K562 is a CML blast crisis cell line carrying the BCR-ABL b3a2 fusion gene; SD1 is a pre B-cell ALL which also carries a BCR-ABL fusion (e1a2); Jurkat is a human T-cell leukaemia; HL60 is an AML cell line with amplified MYC; HEL is an erythroleukaemia cell line containing the JAK2 V617F mutation, and finally 5637 is a urinary bladder carcinoma cell line. From these controls it was anticipated that only EOL-1, SD1 and K562 would show any differential expression due to known fusion genes. The remainder should act as normal controls however problems were encountered due to the low expression of some of the targets in normal controls (Figure 4.11).
Figure 4.11: MLPA run 1, electrophoretic traces. (A) A normal control cDNA note the low expression of all probes for most RTKs other than ABL; (B) the HEL cell line also predicted to act as a normal control, note the similarity with A, and (C) K562 CML blast crisis cell line carrying a \textit{BCR-ABL} fusion gene, as expected the C-terminal region of \textit{ABL} is clearly amplified (star) compared to \textit{ABL} 1b/a2 extracellular domain.

Because of its relatively high expression in normal controls, the \textit{ABL} probes were used as a reference control except for cases where a known ABL fusion was present, in which case the \textit{FGFR1} probes were used as a control. The table below shows an example of the Jurkat cell line (normal relative to RTK fusions) (Table 4.4). The expected ratio for all extracellular/intracellular peaks is 1 but this was only seen for \textit{FGFR1}, and a substantial difference was seen for \textit{PDGFRA}.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Intracellular Peak area</th>
<th>Normalised</th>
<th>Extracellular Peak area</th>
<th>Normalised</th>
<th>RATIO Intra/Extra</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGFR1</td>
<td>84234</td>
<td>0.6415678</td>
<td>33229</td>
<td>0.56473487</td>
<td>1.136051</td>
</tr>
<tr>
<td>PDGFRB</td>
<td>317</td>
<td>0.0024144</td>
<td>209</td>
<td>0.00355201</td>
<td>0.679737</td>
</tr>
<tr>
<td>PDGFRA</td>
<td>20012</td>
<td>0.1524213</td>
<td>3059</td>
<td>0.05198844</td>
<td>2.93183</td>
</tr>
<tr>
<td>ABL control</td>
<td>131294</td>
<td></td>
<td>58840</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.4:** Table showing the MLPA results for the cell line Jurkat. Left to right: The RTK probe; the intracellular peak area determined by capillary electrophoresis; the normalised value when divided by the corresponding control ABL peak (shown at bottom); the extracellular peak area the normalised extracellular value when divided by the corresponding ABL control peak, and finally the ratio produced from the normalised values comparing the intra- and extracellular domains. Based on the cut-offs used for DNA analysis, a relative value of 1 would indicate a normal 1:1 intra:extracellular expression, a value of 0.5 would indicate significant under-expression of the intracellular domain and a value of 1.5 or above would indicate over-expression of the kinase (intracellular domain) compared to the extracellular domain.

I decided to repeat the MLPA assay using the same samples to determine if the assay was reproducible or whether the initial results were biased for any technical reasons. Below is a table showing four of cell lines that were tested twice (Table 4.5). Only the urinary bladder cancer cell line (5637) appeared to produce reproducible results. Unexpectedly, EOL-1 with a known FIP1L1-PDGFRA fusion shows over-expression of PDGFRA on the first run as expected, but virtually no expression on the repeat assay. It was unclear why the results were so variable, however DNA quality and quantity on commercially produced MRC-Holland kits is known to be of utmost importance and it seems likely that cDNA quality and quantity is even more difficult to control for.
<table>
<thead>
<tr>
<th>Gene</th>
<th>EOL1</th>
<th>EOL1</th>
<th>K562</th>
<th>K562</th>
<th>Jurkat</th>
<th>Jurkat</th>
<th>5637</th>
<th>5637</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGFR1</td>
<td>0.808339</td>
<td>2.321555</td>
<td>8.151261</td>
<td>1.650277</td>
<td>2.534954</td>
<td>3.1048</td>
<td>2.935601</td>
<td>3.947915</td>
</tr>
<tr>
<td>PDGFRB</td>
<td>174323</td>
<td>2355</td>
<td>0.810345</td>
<td>3.462349</td>
<td>1.516746</td>
<td>1</td>
<td>1.074525</td>
<td>1.190534</td>
</tr>
<tr>
<td>PDGFRA</td>
<td>392</td>
<td>0.001842</td>
<td>4102</td>
<td>0.001351</td>
<td>6.542007</td>
<td>28615</td>
<td>4.304769</td>
<td>4.050285</td>
</tr>
<tr>
<td>ABL control</td>
<td>2.49143</td>
<td>3.098622</td>
<td>7.562993</td>
<td>13.85285</td>
<td>2.231373</td>
<td>1.90531</td>
<td>1.503822</td>
<td>1.519456</td>
</tr>
</tbody>
</table>

Table 4.5: A table showing the comparison between the first MLPA run (black) and the repeat MLPA run (red). Only four cell lines have been shown here for comparison (other cell line data not shown). Only the 5637 bladder cancer cell line shows reproducible results, the remainder are highly variable.

Further experiments were conducted to try and improve the reproducibility using a series of freshly prepared cDNAs from normal controls and Eol1 cells. The results were more promising with the EOL-1 electrophoretic trace showing a clear over-amplification of the intracellular kinase domain of PDGFRA consistent with the presence of the FIP1L1-PDGFR fusion gene (Figure 4.12A). Even the normal control traces looked cleaner and although expression of both portions of every RTK gene was not always observed, the ABL control was always present (Figure 4.12B). Despite cleaner traces, the intracellular:extracellular ratio was rarely calculated as 1 and in fact this still varied considerably, especially where one of the probes was lacking completely. Moreover, in most cases the ABL control looked differentially expressed implying the ABL 1b/a2 probe may not be as sensitive as its counterpart.
Figure 4.12: MLPA run 3, electrophoretic traces. (A) EOL-1 cell line showing a clear amplification of the PDGFRA intracellular domain, the extracellular domain is barely visible; compare this to (B) a normal control where expression of every probe is visible yet very tightly regulated. Note however the differential expression of the ABL probes.

Fresh cDNA samples were then made from patient samples with available material and known fusion genes involving PDGFRA, PDGFRB, ABL and FGFR1 (n=15). All the samples with known fusions showed an amplification of the intracellular domain of the relevant RTK. The results were very encouraging, although CML cases showed a high degree of variation. Figure 4.13 shows an example MLPA trace showing over-amplification for each of the tyrosine kinases in the assay.
Figure 4.13: MLPA run 4, electrophoretic traces. (A) A patient with a known FGFR1 fusion gene, note the over-amplification of FGFR1 intracellular probe compared to the extracellular probe (ratio of ~9); (B) a patient with a ETV6-PDGFRB fusion gene only the intracellular PDGFRB probe is highly over-expressed; (C) a patient with a PDGFRA fusion gene, the extracellular PDGFRA probe is not even visible, and finally (D) a CML patient with a BCR-ABL fusion showing clear over-expression of the ABL C-terminal probe.

It is important to note here that although over-amplification was observed for the intracellular domain of the correct RTK, in all cases the ratios of the other probes were rarely 1.0 and the CML patients often showed ABL and PDGFRA intracellular domain over-expression. As a result of this work I decided to concentrate only on
PDGFR\(A\), PDGFR\(B\) and FGFR\(1\), and use the ABL probe only as a control for normalisation.

It is clear from these results that sample quality and probe efficiency influence the results as well as any genuine difference in expression of the target sequences. In addition, distortions in the data and substantial variations between replicates were seen when peak sizes were zero or very low. Based on the results I therefore created some rules based on +/− 5 standard deviations of the values seen for all normal controls compared to cases known to harbour specific fusions (Table 4.6).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Extracellular =0, or very low</th>
<th>Extra &amp; Intracellular present</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGFR(1)</td>
<td>Fail if extracellular not &gt;50</td>
<td>Positive is &gt;10.87</td>
</tr>
<tr>
<td>PDGFR(B)</td>
<td>Positive if intracellular is &gt;200, ideally &gt;500</td>
<td>Positive is &gt;5.10</td>
</tr>
<tr>
<td>PDGFR(A)</td>
<td>Positive if intracellular is &gt;200</td>
<td>Positive is &gt; 4.39</td>
</tr>
</tbody>
</table>

Table 4.6: Table shows rules for analysing unknown patients to be screened for possible over-expression of the kinase domain (intracellular) compared to a control band (extracellular) based on normal and positive control data. The data has been split according to detection of both probes and only one probe for each RTK.

4.4.3 Validation of cDNA MLPA assay

Having created these rules I had a colleague blind my samples so I could repeat the MLPA assay and determine its predictive value. These were the same 14 patient samples with known fusions plus four normal controls and the EOL-1 cell line. By applying the rules I was able to identify all four patients with known FGFR\(1\) rearrangements, three of four PDGFR\(B\) rearranged patients and four of five PDGFR\(A\) rearranged patients (including the EOL-1 cell line). The two patients missed with a PDGFR\(A\) or PDGFR\(B\) fusion were scored as normal indicating a false negative result of approximately 15%. Nevertheless, all the normal controls were scored as normal therefore no false positives were identified. Analysis of a further blinded series
correctly identified 9 of 9 cases with PDGFRB rearrangements, and a false positive result was seen in one out of sixteen normal controls (6%).

4.4.4 Results

A number of patients (n=16) were then screened that had been previously found by our laboratory to be sensitive to imatinib in vitro or were of particular interest for other reasons such as the presence of a particular chromosome translocation. All of these patients had an unknown molecular pathogenesis however the sensitivity to imatinib implies they harbour a fusion gene or mutation affecting a tyrosine kinase gene that has a response to imatinib such as PDGFRα or PDGFRβ. Other cases included 5 that were chosen due to possible PDGFRβ involvement, including one patient that had been tested previously by Western blot and shown to have massive over representation of the PDGFRβ protein. Another patient had been found positive by my PDGFRA RT-PCR multiplex and the remaining were imatinib responders in vitro.

Of the sixteen patients I found only 4 to have an unusual expression of the intracellular domains of one of the tyrosine kinases. Three of these patients were in vitro imatinib responders, E1009, E281 and E57 and all showed over-expression of PDGFRα. The final patient (E1334) showed over-expression of PDGFRβ. The sample that was screened due to over-expression of PDGFRα by the multiplex assay proved to be normal by MLPA analysis, and also by bubble-PCR. E1334, the patient with over-expression of PDGFRβ was analysed further using 5′RACE PCR but no PDGFRβ fusion was identified. Insufficient material was available for further investigation of the other cases.

More patients were screened using the MLPA assay (n=39), of whom 16 were further candidate imatinib responders, 18 were found previously positive by PDGFRA RT-PCR multiplex and 5 new patients with unusual referrals.
I found 2 of the 16 imatinib responders consistently failed, 8 were shown to be normal, 5 had over-expression of \textit{PDGFRA} and one was borderline \textit{FGFR1} positive. All the \textit{PDGFRA} positive patients were followed up by bubble-PCR and found to be normal. The borderline positive \textit{FGFR1} case was later repeated and found to be normal. The 18 patients with \textit{PDGFRA} over-expression first identified by RT-PCR multiplex were analysed and only 10 of these were confirmed by MLPA assay to also over-express \textit{PDGFRA}. Of the remaining 8 patients, 5 appeared normal and three consistently failed. These patients were also followed up by bubble-PCR to determine if any unusual fusions might be identified but none were discovered. Finally, of the 5 new referrals, 2 were \textit{PDGFRA} positive, 2 were borderline \textit{FGFR1} positive and one was normal. The \textit{PDGFRA} over-expressing patients were analysed by bubble-PCR but no rearrangements were identified. The patients with \textit{FGFR1} over-expression were analysed using 5’RACE PCR as they were both referred with query 8p11 myeloproliferative syndrome (EMS). Again no rearrangements of the \textit{FGFR1} gene were discovered.

In summary, although the MLPA expression assay was validated with a large number of positive controls and normal controls it is far from a perfect screening test and yielded a large number of apparently false positives. Nevertheless, subsequent analysis demonstrated that the test could be useful.

\textbf{4.4.4.1  A cryptic ETV6-PDGFRB fusion gene}

The outcome of the t(5;12)(q31-33;p13) is the \textit{ETV6-PDGFRB} fusion and, although rare, is the most frequent of the known \textit{PDGFRB} fusion genes (Steer and Cross 2002). In patients with a t(5;12)(q31-33;p13), confirmation of the presence of \textit{ETV6-PDGFRB} is particularly important since many cases with this translocation have a distinct molecular aetiology, with upregulation of \textit{IL-3} expression and no involvement of \textit{PDGFRB} (Cools, \textit{et al} 2002). Only cases with a \textit{PDGFRB} fusion are expected to be responsive to imatinib.
The fusion of *ETV6* exon 4 spliced to *PDGFRB* exon 11 is found in the great majority of positive cases, and is typically detected by RT-PCR using primers directed to *ETV6* exons 3 or 4 and *PDGFRB* exon 11 (David, *et al* 2007). In our laboratory, of 10 positive cases detected with these primer combinations, only one had a variant fusion in which *ETV6* exon 4 was fused to *PDGFRB* exon 9. A patient with idiopathic myelofibrosis and an *ETV6* exon 7-*PDGFRB* exon 10 fusion was also recently reported and it was suggested that this novel transcript may have accounted for the somewhat unusual clinical phenotype seen in this case (Tokita, *et al* 2007). Importantly, both these variants are detectable using the standard primer sets above. Here I present an unusual *ETV6-PDGFRB* case that was missed by standard screening.

4.4.4.2  Patient details

The patient was a 19 year old male (E1026) who presented to the outpatient department with a fever for two months. On examination he was found to have generalised lymphadenopathy, hepatosplenomegaly and bony swellings over the left clavicle and right tibia, however there were no bleeding manifestations. A total leukocyte count exceeding 100 x 10^9/L with a left shift was observed on peripheral blood smear. Bone marrow aspiration showed an increase in the basophils and eosinophils, suggestive of an MPD. Fine needle aspiration showed infiltration of the lymph node by blasts that were negative for myeloperoxidase and periodic acid Schiff on cytochemistry. Flow cytometry of the peripheral blood and lymph node aspirate was positive for myeloid markers (CD11c, 13, 15 and 33) along with CD45 and myeloperoxidase. *BCR-ABL* was not detected by RT-PCR and cytogenetics showed a t(5;12)(q31-33;p13), however *ETV6-PDGFRB* was also not detected by RT-PCR. A diagnosis of atypical CML in accelerated phase was made.
Despite being negative for *ETV6-PDGFRB* the patient was started on imatinib mesylate (600mg in divided doses) and exhibited a complete haematological and cytogenetic response to treatment (including complete regression of the lymph nodes, bony swellings and hepatosplenomegaly) at 3 months post treatment. One year after diagnosis the patient continues in remission on 600mg/day imatinib.

### 4.4.4.3 cDNA MLPA analysis of patient E1026

In view of the excellent response to imatinib, we performed further analysis to determine if *PDGFRB* was involved using the cDNA MLPA assay (material was not available for molecular cytogenetic analysis). This assay detected relative overexpression of the *PDGFRB* kinase domain in this patient (E1026) but not in a number of normal controls, suggesting the presence of a *PDGFRB* fusion (Figure 4.14).
Figure 4.14: The cDNA MLPA assay. Probes are in order *FGFR1* tyrosine kinase domain, *FGFR1* 3’extracellular domain, *PDGFRB* tyrosine kinase domain, *PDGFRB* 3’extracellular domain, *PDGFRA* tyrosine kinase domain, *PDGFRA* 3’extracellular domain. Under normal circumstances expression of *FGFR1*, *PDGFRB* and *PDGFRA* is low level in peripheral blood leukocytes. Over-expression of the *PDGFRB* intracellular domain is apparent in both the previously known *ETV6-PDGFRB* case (middle) and also E1026 (top), compared to the normal control (bottom).

4.4.4.4 5’RACE PCR to identify the potential *PDGFRB* fusion gene in patient E1026

I employed 5’rapid amplification of cDNA ends PCR (Invitrogen, Paisley) to characterise the fusion, employing reverse primers derived from sequences located downstream of known *PDGFRB* breakpoint locations (Appendix I, section IV). The amplification was performed under long-range PCR conditions (section 2.2.5) to enable fragments up to 10kb to be amplified. After nested PCR a smear was observed with 4 or 5 possible bands visible, ranging from 250bp – 650bp. The negative control was completely blank indicating one of these bands may have been derived from a *PDGFRB* fusion (Figure 4.15).
Figure 4.15: Agarose gel of nested 5’RACE PCR for patient E1026 versus a control. Lane 1 indicates a no DNA control, only a primer dimer smear is visible below 100bp; Lane 2 shows the nested PCR result after E1026 5’RACE PCR a smear is visible containing several different sized bands ranging from 250-650bp. Red arrows indicate the bands that were cloned and sequenced.

The resulting PCR products were cloned and the digests revealed only two different sized bands one at approximately 500bp and the other at approximately 650bp (indicated by red arrows, Figure 4.15). Sequencing of the smaller product turned out to be PDGFRB and vector sequence only, however sequencing of the larger 650bp product revealed an in-frame mRNA fusion between full length ETV6 exon 7, 34bp derived from ETV6 intron 7 and a truncated PDGFRB exon 12 (Figure 4.16). This fusion involves novel breakpoints, never previously reported.
Figure 4.16: (A) Electropherogram of the cDNA junction amplified by 5’RACE PCR; (B) Sequence of the novel ETV6-PDGFRB fusion showing the insertion of ETV6 intronic sequence ensuring the fusion reads in-frame.

4.4.4.5 Confirmation of the novel ETV6-PDGFRB fusion

The presence of the fusion was confirmed by RT-PCR using the primers ETV6 exon 6.2F and PDGFRB exon 14.1R with the High Fidelity PCR master kit according to manufacturer’s conditions (see appendix I, section XX for primer sequences). The PCR was performed on a generic 66°C annealing temperature PCR protocol. The in-frame fusion was detectable by single step RT-PCR, as seen in Lane 1, Figure 4.17. RT-PCR of the fusion revealed two bands the upper most of which was of the expected size, 415bp. On cloning and sequencing the lower band turned out to be an ETV6-PDGFRB out-of frame fusion between part of ETV6 exon 7 and PDGFRB exon 13. Using the same High Fidelity PCR kit and amplification protocol, I used RT-PCR to try and amplify the reciprocal translocation. PDGFRB-ETV6 reciprocal products were not detectable by single or nested PCR, Lane 2, Figure 4.17.
Figure 4.17: Amplification of the *ETV6-PDGFRB* fusion by single step RT-PCR: Lane 1 represents patient E1026 fusion RT-PCR; Lane 2 represents the reciprocal breakpoint PCR in patient E1026, Lanes 3 & 4 show normal controls, and Lane 5 a no DNA control. Note the reciprocal breakpoint was not observed at single-step or nested PCR (data not shown). The fusion RT-PCR shows two bands of 300 and 415 base pairs, these represent an out-of-frame fusion and the confirmed in-frame fusion.

I also went on to determine the genomic breakpoints of this fusion for further confirmation using the primers ETV6.E1026.Int7.1F and ETV6.E1026.Int7.2F with PDGFRB exon 12R (appendix I, section XX). Both PCRs resulted in a product that was subsequently cloned and sequenced. I was able to show the same fusion breakpoints between exon 12 of *PDGFRB* and intron 7 of *ETV6* in both PCR products (Figure 4.18).
Figure 4.18: (A) Amplification of the genomic ETV6-PDGFRB novel fusion. Lane 1 is amplification of the genomic breakpoint using primer ETV6.Int7.1F, Lane 2 is the genomic breakpoint amplified using a second primer ETV6.Int7.2F, both at single step combined with primer PDGFRB exon12.1R; (B) Schematic showing the individual sequences for PDGFRB (italics), ETV6 and the ETV6-PDGFRB fusion sequence.

4.4.4.6 Further ETV6-PDGFRB analysis

The PDGFRB sequence retained in the novel fusion is downstream of the PCR primer we had initially used for RT-PCR, explaining why ETV6-PDGFRB fusion was not detected. To ascertain if any other similar cases had been missed, 10 cases with a t(5;12) that tested negative for ETV6-PDGFRB using the standard primers were re-analysed using ETV6.Ex2F and PDGFRB.Ex14R. All those tested proved negative, indicating that this variant is not common. Nevertheless, as a result of the finding of this novel ETV6-PDGFRB fusion the above primers are now routinely used to screen t(5;12) cases referred to our laboratory.
In summary, patient E1026 tested negative by RT-PCR for the ETV6-PDGFRB fusion, but following the finding that his disease responded to imatinib, MLPA and 5’RACE analysis showed the patient to have a novel ETV6-PDGFRB variant with breakpoints outside those commonly reported.

The unusual ETV6-PDGFRB fusion is reminiscent of those seen for FIP1L1-PDGFRA, for which breaks almost always fall within PDGFRA exon 12 and FIP1L1 intron-derived sequence is frequently incorporated into the mature mRNA (Cools, et al 2003a). As a consequence the region encoding the PDGFRα auto-inhibitory WW-like domain is disrupted, resulting in partner protein independent activation of the kinase moiety (Stover, et al 2006). The disruption of PDGFRβ WW-like domain in our case is unexpected in view of the fact that ETV6 encodes a dimerisation domain that is required for transformation by ‘normal’ ETV6-PDGFRB (Carroll, et al 1996).

Interestingly, this case presented with advanced phase disease, something not normally associated with this fusion. The fusion itself retained ETV6 exons 1-7, in common with that described by Tokita et. al., (2007) in a case that also presented with advanced features, specifically those of chronic idiopathic myelofibrosis. In contrast to the most common fusion reported that involves ETV6 exons 1-4, these variants are predicted to encode a chimaeric protein that retains the ETV6 internal and ETS DNA-binding domain (Figure 4.19). It is possible that retention of this domain leads to a more aggressive phenotype however it would be premature to draw any clear conclusions from just two cases (Tokita, et al 2007).
Figure 4.19: Schematic of PDGFRβ and the Tel (ETV6) proteins showing relevant domains and arrows indicating reported breakpoints. Below, are the sequences of ETV6-PDGFRB fusions described to date, with the most common variant at the top and the fusion in case E1026 at the bottom.

It is also noteworthy however that our case responded well to imatinib despite being treated in advanced phase. Good responses to imatinib have been described for FIP1L1-PDGFRα associated chronic eosinophilic leukemia in transformation (Metzgeroth, et al, 2007) and this may be a general feature of diseases associated with PDGFR fusions.

It is clear from this case that the breakpoint diversity for ETV6-PDGFRB fusions is greater than originally reported and hence diagnostic screens for this fusion should
employ primers capable of detecting all variants in order to provide an accurate molecular diagnosis allowing for appropriate clinical management.

4.4.5 MLPA Discussion

MLPA is a technique that utilises probes to detect deletions and amplifications of a gene. I designed a cDNA MLPA assay to detect over-expression of the kinase domain of specific tyrosine kinase genes relative to a control probe spanning a number of exons upstream of the common reported breakpoints. The basis of the assay is that under normal conditions the mRNA expression of each tyrosine kinase is tightly regulated, but when fused to a partner gene the expression level of the tyrosine kinase portion of the gene may be increased as a consequence of gene fusion. By screening patients of unknown aetiology I hoped to be able to discover cases with cryptic or novel fusion genes. I designed the assay to incorporate both PDGFR RT-PCR assays into one reaction whilst also accommodating for the FGFR1 and ABL tyrosine kinases. MLPA is ideal for screening multiple tyrosine kinase genes in one assay, and hence is economical on volume of material required. In addition, it is designed to screen high numbers of patients quickly and reproducibly.

The assay was validated using numerous normal control cDNA samples sourced from a non-leukaemic background along with a number of cell lines and patient samples with previously known fusion genes involving one of the tyrosine kinase genes within the assay. Most positive controls were identified, with over-amplification of the intracellular domain of the correct RTK in these cases however, the intracellular:extracellular ratios of the other unaffected probes were rarely 1.0, presumably because of differences in probe efficiencies. The false negative result was 15%, where known positive controls were scored as normal. Analysis of more positive controls was more successful with nine out of nine patients identified correctly. The normal controls were all scored as normal with only one exception that may have been related to poor cDNA quality.
A major drawback of the cDNA MLPA assay was the low expression of the tyrosine kinases under normal circumstances which made reproducibility difficult. The assay was also very dependent on the quality of the cDNA and it appeared that some probes were more sensitive than others. This was particularly apparent when I tested known CML patients for validation of the assay. ABL over-amplification often appeared concomitantly with over-expression of the PDGFRα intracellular domain making these patients difficult to distinguish. Despite redesigning new ABL probes to try and combat this problem differential expression of the ABL probes themselves was always observed. As a result only the PDGFRα, PDGFRβ and FGFR1 probes were used to try and identify possible fusion genes, and the ABL probe was used as a control for normalisation. Despite the drawbacks the results were collated and rules of analysis created based on plus or minus five times the standard deviation for all normal controls compared to positive controls.

Having validated the assay it was used to screen MPD patients of unknown molecular pathogenesis from three basic categories: imatinib responders as previously determined by a colleague using in vitro liquid and colony assays (n=26), patients found to over-express PDGFRα or PDGFRβ by RT-PCR assays (n=24) and finally patients with cytogenetic or other indicators of possible gene fusions (n=6). Apparent over-expression was seen for several cases but only one (2% of all cases) proved to be positive on further investigation.

4.5 Discussion

Persistent aberrant signalling of PDGF and its receptors is thought to be essential for the survival of the many cancer cell types. These findings have influenced the pharmaceutical and research communities to design agents that selectively block PDGFR signalling (Levitzki 2004). Imatinib mesylate has received the most attention and has potent inhibitory activity against both PDGFRα and PDGFRβ. Patients with MPDs and rearrangements of PDGFRα and PDGFRβ, in particular, have been
shown to respond very well to treatment with this compound (Apperley, et al 2002, David, et al 2007). However, a shortcoming of imatinib has been the emergence of acquired resistance in the form of point mutations in the kinase domain in a subset of cases. Furthermore some patients show primary or secondary resistance for unknown reason and others are intolerant of therapy. The development of second generation ATP-competitive tyrosine kinase inhibitors (TKI) has begun to try and overcome these problems. These include nilotinib, a novel aminopyrimidine inhibitor (Weisberg, et al 2005). In pre-clinical studies nilotinib has been shown to be 30-fold more potent than imatinib in treating chronic-phase CML, and Stover et. al., (2005) have also showed it is effective at inhibiting proliferation in vitro of ETV6-PDGFRB, FIP1L1-PDGFRα and imatinib resistant ETV6-PDGFRB/T681I, but not FIP1L1-PDGFRα/T674I mutants (Stover, et al 2005). A second TKI, dasatinib is a potent dual Src/Abl kinase inhibitor that is 67-fold more potent than imatinib (Shah, et al 2004). Its greatest benefits have been observed in patients with CP-CML with 90% achieving a complete haematological response (Steinberg 2007). Dasatinib is also effective against KIT and PDGFRs, in addition to overcoming resistance exhibited by 18/19 imatinib resistant mutations reported (Shah, et al 2006). Sunitinib has also been studied for its effectiveness against PDGFRs although most of this work has centered on PDGFR-associated solid tumours such as colon cancer, non-small cell lung carcinoma and melanoma. However, preclinical trials have shown this TKI has both potent and robust anti-angiogenic and anti-tumour activity (Chow and Eckhardt 2007). Finally, other inhibitors have also been developed that have shown activity against PDGFR, such as sorafenib. This is a small molecule B-Raf inhibitor currently used for the treatment of renal cell carcinoma. It has been shown in vitro to be a potent inhibitor of ETV6-PDGFRB by inducing cell cycle block and apoptosis (Lierman, et al 2007).

To date, 21 PDGFRB and 6 PDGFRA fusion genes have been reported. It seems likely that additional PDGFR rearrangements are yet to be discovered, particularly in view of occasional reports of cases that are responsive to imatinib despite being negative for known fusions and, in most cases, cytogenetic indicators of variant
fusions (i.e. rearrangements involving 4q12 or 5q31-33). To search for new fusions I designed RT-PCR and MLPA assays to detect over-expression of the region encoding the kinase domain of selected tyrosine kinases. This region is retained in gene fusions and may be overexpressed relative to the normal, full length product. Both assays were validated using cases with known rearrangements and normal controls and although there is no doubt that the assays could be considerably improved, e.g. by testing and incorporating other primer combinations, both were applied to screen cases of unknown molecular aetiology.

I screened 229 MPD patients negative for *FIPL1-PDGFRA* by *PDGFRA* RT-PCR and found a total of 14 that over-expressed the kinase domain of *PDGFRA*. Of these cases two yielded identification of unknown fusion genes. The first was a previously unreported novel *KIF5B-PDGFRA* fusion as a consequence of a complex cytogenetic rearrangement involving chromosomes 2, 4 and 10. The second case proved to have a *FIP1L1-PDGFRA* fusion that had not previously been detected by standard RT-PCR. The remaining cases for whom material was available (n=12) were investigated further by bubble-PCR as *PDGFRA* rearrangements are particularly suited to this technique since the breakpoints in all cases reported to date are clustered within exon 12 of this gene. No *PDGFRA* rearrangements were identified after screening these patients in this way, suggesting a false positive rate for the expression assay of ~6%.

I also screened 199 MPD patients looking for over-expression of the *PDGFRB* kinase domain; of these I found 9 to be strongly positive. Surprisingly, two of these were later discovered to have known *FGFR1* rearrangements suggesting interplay between the signalling cascades affecting both these tyrosine kinase genes. Of the remaining 7 patients only one was found to harbour an unknown *PDGFRB* rearrangement. This in fact turned out to be the most common of the *PDGFRB* fusions, *ETV6-PDGFRA*. It is unclear why the patient was initially screened and found negative by standard RT-PCR but clearly the usefulness of screening all MPD patients using this more generalised RT-PCR has proved its worth. Other patients with available material were
subsequently screened by 5’RACE PCR but no further rearrangements were discovered. This gives a false positive rate of 4%.

The MLPA assay proved to be less reliable: of 56 MPD patients screened (some of whom were also screened by the RT-PCR assays above) 25 were found to over-express either \textit{PDGFRA}, \textit{PDGFRB} or \textit{FGFR1}. \textit{PDGFRA} over-expression was the most frequent (n=20), 2 over-expressed \textit{PDGFRB} and three were borderline for over-expression of \textit{FGFR1}. 5’RACE PCR and bubble-PCR of all patients where material was available was performed and with one exception no rearrangements were detected. The exception was a patient referred with a t(5;12) who turned out to have a variant \textit{ETV6-PDGFRB} fusion with breakpoints that could not be detected with standard primers.

The most important conclusion to be drawn from these results is that cryptic fusions of \textit{PDGFRA}, \textit{PDGFRB} and \textit{FGFR1} are very uncommon in MPDs, even in cases that responded to imatinib but tested negative for known fusions. This is because although apparent over-expression of specific kinase domains were seen by the RT-PCR and/or MPLA assays, further analysis failed to reveal gene fusions in most cases. Although RACE PCR might miss some fusions, particularly if the partner gene is very large, we have previously found that bubble PCR is highly effective for isolating genomic \textit{PDGFRA} fusions since the breakpoints within this gene are tightly clustered. Most apparent over-expressors involved \textit{PDGFRA} and thus it seems likely that most were false positives. A second conclusion is that although both assays proved to be useful in identifying new fusions in the research setting, their false positive rates are unacceptable for routine diagnostic use because of the amount of effort required to follow these cases up. It seems likely that the false positive rates could be reduced by the use of further primer sets in the MPLA assay and possibly the more rigorous definitions of cutoffs to define positive and negative results. It is also possible that more quantitative techniques such as real time PCR may be useful to measure the relative expression of the regions encoding the extracellular and
intracellular domains of tyrosine kinases. Other systems such as exon expression arrays might also be used, but would be expensive for routine use.

Overall it is estimated that tyrosine kinase fusions account for only about 2-5% of atypical MPD cases (Reiter, et al 2007). Other abnormalities of tyrosine kinase or downstream signalling components have also been described, e.g V617F JAK2 (10% of cases), NRAS mutations (12.5% of cases) and activating FLT3 or KIT mutations (5% of cases) (Jones, et al 2005). Collectively these abnormalities account for little over a quarter of all atypical MPD cases (Figure 4.20).

Figure 4.20: Pie chart showing the molecular pathogenesis of atypical MPD patients (i.e. atypical CML, CMML, CEL, MPD-U, and MDS/MPD-U). The blue region represents atypical MPD patients that have a normal or aneuploid karyotype and are of unknown aetiology. The section in yellow indicates collectively the number of patients defined molecularly with tyrosine kinase mutations or mutations of downstream signaling components. The section in green shows the small number of patients found specifically to have fusion genes involving tyrosine kinases.

If cryptic tyrosine fusions are uncommon, then other abnormalities must be responsible for the majority of atypical MPDs for whom the molecular pathogenesis
is currently unknown. There are several possibilities: first, they may harbour gene fusions of other tyrosine kinases that I have not investigated. Second, they may harbour activating mutations of tyrosine kinases that would not have been detected by my over-expression assays. Third, they may harbour activating mutations in other signalling components that are either upstream or downstream of tyrosine kinases. Fourth, they may harbour mutations in completely different pathways. In the next Chapter I describe a novel approach aimed at identifying mutations in atypical MPDs.
5 SCREENING FOR NEW ONCOGENES MARKED BY ACQUIRED UNIPARENTAL DISOMY

5.1 Introduction

5.1.1 Loss of heterozygosity (LOH) & acquired uniparental disomy (aUPD)

Loss of heterozygosity (LOH) refers to a widespread phenomenon in cancer whereby normal inherited heterozygous markers in a genomic or chromosomal region are reduced to homozygosity in the malignant clone. LOH frequently marks the presence of a recessive tumour suppressor gene, the classic example being retinoblastoma in which an inherited inactivating mutation in the Rb gene along with flanking heterozygous markers are all reduced to homozygosity in retinal tumours (Knudson 1971). LOH may originate from diverse mutational events such as deletion of the wild-type allele to yield a hemizygous state, mitotic recombination, localised gene conversion or non-dysjunction with either loss or reduplication of the mutant chromosome (Figure 5.1). There are various techniques discussed in detail below that can identify LOH and help to distinguish between the underlying mutational mechanism. In the last few years the discovery of large stretches of somatically acquired LOH associated with no net change in copy number, termed as acquired isodisomy or acquired uniparental disomy (aUPD) has emerged as a common theme, initially in polycythaemia vera but subsequently in acute leukaemia and other malignancies (Irving, et al 2005, Kralovics, et al 2002, Raghavan, et al 2005).

Classically, uniparental disomy (UPD) refers to the finding of an individual inheriting both homologues of a chromosome pair from only one of their parents as a result of an error in meiosis. It is associated with a range of inherited developmental disorders such as Beckwith-Wiedemann syndrome (UPD of chromosome 11p15) and Prader-
Willi syndrome (25% have maternal UPD of chromosome 15). My work is concerned with identifying regions of aUPD that are thought to be caused by mitotic recombination and selection for one of the derivative products. Recent evidence, reviewed below, has suggested that aUPD is usually associated with specific oncogenic mutations. Definition of novel, minimal regions of aUPD may therefore be an important new route to identify novel oncogenes.

Figure 5.1: Schematic illustrating the principal classical mutational events that lead to LOH: A) Interallelic gene conversion, a non-reciprocal exchange occurring between two alleles that reside on homologous chromosomes resulting in loss of the normal allele as shown in the last cell at the top; B) A point mutation may occur through a reading error during semi-conservative replication resulting in an additional mutant chromatid, represented by the end cell at the top, finally C) Mitotic non-dysjunction. This example shows non-dysjunction followed by re-duplication of the mutant chromosome.
5.1.2  \textit{aUPD and MPDs}

The first description of \textit{aUPD} in MPDs was reported by Kralovics and colleagues in 2002 (Kralovics, et al 2002). Previously, studies of rare familial polycythaemia vera (PV) had suggested a possible pathogenic role for loss of heterozygosity (LOH) in this disease. As a consequence they performed whole genome LOH analysis using a series of microsatellite markers on a small number of PV subjects and identified LOH at three new genomic loci, by far the most frequent of which was at 9p. They then showed that there was no change in copy number at 9p, suggesting that mitotic recombination was the mechanism of LOH i.e. \textit{aUPD} rather than allele loss. The prevalence of 9p \textit{aUPD} was observed to be approximately 33\% of all PV cases, and is therefore the most frequent chromosomal lesion seen in these patients. This group mapped the minimal \textit{aUPD} region and identified the possible candidate genes within this region, which included the $\textit{JAK2}$ tyrosine kinase. Although this group went on to screen candidate genes in this region including a partial screen of $\textit{JAK2}$, no mutations were observed and the significance of their observations was unclear.

5.1.3  \textit{aUPD \& the discovery of the JAK2 V617F acquired mutation}

It took a further 3 years before the observations of 9p \textit{aUPD} in PV came to fruition. In 2005 four groups reported the discovery of an acquired G to T mutation at nucleotide 1849 in exon 12 of the $\textit{JAK2}$ gene that results in the substitution of a valine by a phenylalanine (V617F) (Baxter, \textit{et al} 2005, James, \textit{et al} 2005, Kralovics, \textit{et al} 2005a, Levine, \textit{et al} 2005a). The four groups took three different approaches to find this mutation:

(i) James et al., took a functional approach. They observed that the formation of erythropoietin (Epo)-independent colony formation (EEC), one of the biological hallmarks of PV, was impaired by both JAK2 inhibitors and also short interfering RNA (siRNA) against JAK2. They went on to screen the $\textit{JAK2}$ gene and identified
the mutation. They then showed that introduction of V617F into mouse bone marrow cells by retroviral-mediated gene transfer gave rise to a PV-like phenotype which was not seen in controls, thus clearly implicating V617F as causative in the disease process. This seminal discovery again highlighted the importance of aberrant tyrosine kinase signalling in the pathogenesis of MPDs.

(ii) Based largely on the recurrent findings of abnormal tyrosine kinases in atypical MPDs, both Baxter et. al., 2005 and Levine et. al., 2005 performed mutational screens of genes encoding tyrosine kinases.

(iii) Kralovics et. al., 2005 followed up on their earlier work that identified 9p UPD in PV. Analysis of further patients led to a smaller minimally affected region that still included JAK2. Despite having been apparently excluded previously, JAK2 remained an obvious candidate because of its known importance in erythropoiesis. Consequently, this gene was re-screened and the V617F mutation identified in a small region that was apparently not covered in their previous analysis (Figure 5.2).
At the time our group was also screening tyrosine kinase genes for mutations, mainly in atypical MPDs. Once the finding of V617F JAK2 emerged, we looked at a series of cases for the presence of this mutation. In total, we investigated 480 MPDs and found that the percentage of PV, ET and IMF patients positive for the mutation correlated well with the other studies. Interestingly the mutation was also seen in a subset of atypical MPDs: 30/152 (20%) atypical or unclassified MPD, 2/134 (2%) idiopathic hypereosinophilic syndrome were V617F positive but the mutation was not seen in any patients with systemic mastocytosis (SM), AML, secondary erythrocytosis or in normal individuals (Jones, et al 2005).

5.1.4 The association between JAK2 V617F and 9p aUPD

Work performed by Amy Jones using both direct sequencing and pyrosequencing, observed a wide range in the proportion of JAK2 alleles that carried the V617F mutation. In many patients the mutation appeared to be homozygous, i.e. the residual wild type allele was completely or almost completely lost. Homozygosity was only seen in PV and MF and strikingly, the frequency was similar to that of aUPD as described previously (Kralovics, et al 2002). To investigate the relationship between homozygosity for V617F and aUPD 9p in more detail, I used microsatellite PCR analysis to look for chromosome 9p LOH in a series of patients. Microsatellites are repetitive sequence elements that are arranged in tandem and dispersed throughout almost the entire genome. The length of the core repeats are highly variable, and form allelic variants that can be analysed using PCR. I used 8 primer pairs flanking the
dinucleotide (CA) highly polymorphic microsatellite markers on chromosome 9p (Table 5.1).

<table>
<thead>
<tr>
<th>Chromosome band</th>
<th>Marker</th>
<th>Heterozygosity (<a href="http://www.gdb.org/">http://www.gdb.org/</a>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p24.3</td>
<td>D9S1779</td>
<td>0.64</td>
</tr>
<tr>
<td>p24.2</td>
<td>D9S1858</td>
<td>0.58</td>
</tr>
<tr>
<td>p24.2</td>
<td>D9S288</td>
<td>0.86</td>
</tr>
<tr>
<td>p24.2</td>
<td>D9S1813</td>
<td>0.84</td>
</tr>
<tr>
<td><strong>p24.1</strong></td>
<td>JAK2</td>
<td></td>
</tr>
<tr>
<td>p24.1</td>
<td>D9S286</td>
<td>0.88</td>
</tr>
<tr>
<td>p23</td>
<td>D9S254</td>
<td>0.75</td>
</tr>
<tr>
<td>p22.2</td>
<td>D9S157</td>
<td>0.85</td>
</tr>
<tr>
<td>p21</td>
<td>D9S171</td>
<td>0.80</td>
</tr>
</tbody>
</table>

**Table 5.1:** Microsatellite marker name, location on chromosome 9p in relation to the JAK2 gene, and the relative heterozygosity of each microsatellite marker.

I analysed 30 MPD cases that did not have the V617F mutation and 27 V617F homozygous patients. The products were analysed on an ABI3100 using the Genotyper 2.0 program. If only one peak was visible a patient was scored as homozygous. To allow for the presence of background normal cells, homozygosity was also scored if one peak was one third or smaller than the expected size compared to controls. A significant tract of homozygosity was scored if four or more consecutive markers abutted or encompassing JAK2 were homozygous, since the probability of this occurring by chance was well under 5% as determined by known the rates of heterozygosity (http://www.gdb.org/). The results of the analysis are
summarised in Table 5.2. Of the 27 V617F homozygotes, 26 were found to have chromosome 9 UPD (96%). Of the 30 JAK2 negative PV, IMF and ET patients, only four were found to have significant homozygosity (13%) (P<0.0001, Fisher’s exact test). This indicates that LOH is significantly associated with homozygosity for the V617F mutation.

To determine if the observed homozygosity was due to aUPD or LOH, MLPA was performed to look for copy number by another member of our group. One copy of JAK2 would indicate LOH due to either a whole chromosome deletion or specific chromosomal region, whereas two copies would indicate aUPD. All seven cases analysed by MLPA that were homozygous for at least four 9p microsatellites and were homozygous for the V617F mutation were shown to have two copies of JAK2, consistent with chromosome 9 aUPD (Kralovics, et al 2002). Whilst my analysis was ongoing Kralovics et. al., (2005) published similar data showing a strong correlation between V617F homozygosity and 9p aUPD (Kralovics, et al 2005a).
Table 5.2: 9p LOH in V617F homozygous patients compared to V617F negative cases. The V617F genotype of the patient is indicated on the right as either homozygous normal (wt/wt; patients 1-30) or homozygous mutant (M/M; patients 31-57). The microsatellite markers are shown in order from the most telomeric (D9S1779) to the
most centromeric (D9S171) and are scored as homozygous (black), heterozygous (white) or not done (grey).

5.1.5    **Acquired UPD is not limited to MPDs**

More recently it has become apparent that aUPD is widespread in haematological malignancies, for example AML (Raghavan, *et al* 2005), ALL (Irving, *et al* 2005) chronic lymphocytic leukaemia (Pfeifer, *et al* 2007) and multiple myeloma (Walker and Morgan 2006) plus also solid tumours such as basal cell carcinomas (Teh, *et al* 2005), prostate cancer (Torring, *et al* 2007) and colorectal cancer (Gaasenbeek, *et al* 2006).

In contrast to the initial studies in PV which used microsatellite analysis, these more recent studies used genome-wide microarray single nucleotide polymorphism (SNP) analysis. For example, Raghavan *et al*., showed that approximately 20% of normal karyotype AMLs exhibited large regions of homozygosity that was not associated with visible chromosomal abnormalities. They were also able to show that a previously identified homozygous CEBPA mutation coincided with a region of large scale aUPD of chromosome 19 (Raghavan, *et al* 2005). Further studies in AML determined that aUPD for chromosome 13q was associated with activating FLT3 mutations (Griffiths, *et al* 2005), aUPD 11p with inactivating WT1 mutations and aUPD 21q with RUNX1 mutations (Fitzgibbon, *et al* 2005). Mutations in other malignancies have also been linked to homozygosity for specific mutations, e.g. the glucocorticoid receptor gene in pediatric ALL (Irving, *et al* 2005), PTPN11 gene in ALL associated with Noonan syndrome (Karow, *et al* 2007), CDKN2A and TP53 in follicular lymphoma (Fitzgibbon, *et al* 2007) and NF1 in JMML (Flotho, *et al* 2007).

These studies show a strong association between regions of aUPD and specific acquired mutations in oncogenes or tumour suppressor genes. In reports where paired samples have been studied, it appears the mutation precedes the aUPD and hence the aUPD is a ‘second hit’ leading to homozygosity of an oncogenic mutation associated

5.1.6 Aims of my analysis

The primary aim of my analysis was to determine if aUPD was present in atypical MPDs and, if so, to follow up any findings to try and identify new pathogenetic lesions in these diseases. As described above, MPDs are characterised by deregulated tyrosine kinases and therefore our principal hypothesis was that novel activating tyrosine kinase abnormalities were likely to be the root cause of many or all atypical MPDs. Acquired UPD has been associated with two activated tyrosine kinases – JAK2 and FLT3 – and therefore identification of new regions of aUPD may therefore be a particularly good method to provide pointers towards new kinase mutations. In addition, since this was an entirely new and potentially very powerful method of analysis, I also took the opportunity to analyse a range of other cases of haematological disorders.

5.2 SNP CHIPS TO IDENTIFY REGIONS OF UPD

5.2.1 SNP chips

SNPs are believed to be the most common sequence variation in the human genome, occurring every approximately 1,200bp (Sachidanandam, et al 2001) and thus there are likely to be more than 2 million polymorphic sites in the human genome. The identification of SNPs that are roughly evenly spaced has enabled the realisation of whole genome SNP screens. The advent of high density single SNP chips has revolutionised the field of genome-wide allelic screening as they can accurately and reproducibly detect copy number and loss of heterozygosity simultaneously. A few reports have compared LOH differences observed by microsatellite analysis versus
those found by SNP array and demonstrated excellent concordance however, as expected, additional smaller regions of LOH were detected by SNP arrays due to their higher resolution (Dumur, et al 2003, Lindblad-Toh, et al 2000, Wang, et al 2004b).

5.2.2 Plan of analysis and methods

When I started this project there were two platforms available: Affymetrix 10K chips, i.e. chips that harboured 10,000 SNPs across the genome, and Affymetrix 100K chips, which actually consisted of 2 x 50K chips that each contained 50,000 genome-spanning SNPs. The studies that had been published at the time had used 10K arrays and this density was easily sufficient to identify regions of UPD (Irving, et al 2005, Raghavan, et al 2005). However I wanted to explore the possibility of looking at more detail at copy number changes and in particular small deletions or amplifications that may mark the presence of cryptic fusion genes. Because of the relatively high cost of the technique I elected to use just one 50K chip per sample in order to maximise the number of cases that could be analysed.

The generic method of sample preparation for SNP chip use was devised by Kennedy et. al., (2003) (Kennedy, et al 2003). The whole-genome sampling assay (WGSA) is based on a simple restriction enzyme digestion, followed by linker-ligation of a common known adaptor sequence to every fragment generated by digestion. Using a single primer complementary to this adaptor, multiple loci can be simultaneously amplified. Hence, the genomic DNA is converted by PCR into amplified fragments of reduced complexity that can be easily hybridised to the SNP arrays (Figure 5.3). This method was aided by the completion of the human genome project enabling prediction of restriction fragments that should amplify consistently and SNPs that reside within these fragments. Oligonucleotides corresponding to these SNPs are then synthesised onto high-density microarrays. The XbaI 50K mapping array has 58,960 SNPs spaced across the whole-genome with 99.1% of the genome within 500Kb of a SNP, 91.6% within 100Kb and 40% within 10kb (Matsuzaki, et al 2004). For each SNP there are 40 oligonucleotides (20 perfect matches and 20 mismatches) these are
scattered across the chip in order to minimise any problems arising from patchy hybridisation.

**Figure 5. 3:** Schematic representing in brief how the Affymetrix SNP chips work adapted from the website of ImaGenes GmbH, Berlin, Germany. Genomic DNA is first digested using a restriction enzyme that cut regularly throughout the genome, such as XbaI or HindIII. An adapter sequence (represented in blue) is then ligated on to both ends of every fragment. A primer is designed to the adapter sequence that is capable of amplifying all the digested and ligated fragments. These fragments are then end labeled and hybridised to the chip. The results are typed as homozygous (AA or BB) or heterozygous (AB), and a measure of the intensity of the hybridisation signal relates to the copy number of the SNP.

We do not have the local facility to perform Affymetrix analysis ourselves and therefore the laboratory elements were performed as a service by the RZPD in Berlin (now ImaGenes GmbH). The RZPD performed the labelling and hybridisation to the XbaI 50K Affymetrix GeneChip SNP array using our samples. After hybridisation, the array was washed and stained using an Affymetrix Fluidics station. The staining protocol is designed to amplify the signal from the annealed probe. Scanning was
then carried out using an Agilent GeneArray Scanner and the raw data was provided to us as a cabinet (.CAB) file of approximately 50Mb for each sample.

The .CAB file is first read by the Affymetrix GeneChip® Operating System (GCOS) in conjunction with the library files for the 50K mapping array, which contain the details of all SNP and probe locations (http://www.affymetrix.com/index.affx). Further analysis to genotype each call was then carried out using Affymetrix GeneChip® Genotyping Analysis Software (GTYPE 4.1). This is specifically designed to give highly accurate, automated SNP allele calls for the mapping arrays, including the 50K SNP array. The programme includes a new genotyping algorithm, designated BRLMM that produces a quality score for each genotyping call (Rabbee and Speed 2006). Copy-number estimations were then determined using the Affymetrix GeneChip® Chromosome Copy Number Analysis Tool (CNAT 4.0) which is integrated into GTYPE 4.1. The copy number algorithm jointly uses genotype information and perfect match (PM) probe intensity and discrimination ratios between paired PM and mismatched (MM) probe intensity values to identify and estimate genetic copy number changes. Values from an experimental sample are compared with SNP-specific distributions derived from a reference set containing over 100 normal individuals (provided by Affymetrix) to gain statistical power. Once the data was fully analysed using Affymetrix GeneChip® software, it was exported to Excel spreadsheets developed in Newcastle by Professor Andy Hall which displays LOH and copy number changes in ideogram format for individual samples or across multiple samples for the ready detection of common areas of deletion.

5.2.3 Patient samples

DNA from specific patient groups was selected, quantified, tested by agarose gel electrophoresis to check for degradation and 250ng sent to the RZPD. The samples I sent for analysis (in four batches) are shown in Table 5.3. For the cases of CML in blast crisis (CML-BC) cases, samples were selected that had at least 75% blasts. No specific criteria were employed for other patient subgroups, although for
aCML/aMPD I focused preferentially on cases with a high leucocyte count, i.e. > 20x10⁹/L.

<table>
<thead>
<tr>
<th>Leukaemia subgroup</th>
<th>Nos. sent for analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>aCML/aMPD</td>
<td>30</td>
</tr>
<tr>
<td>CML: myeloid blast crisis</td>
<td>10</td>
</tr>
<tr>
<td>CML: lymphoid blast crisis</td>
<td>10</td>
</tr>
<tr>
<td>JAK2 V617F negative IMF</td>
<td>18</td>
</tr>
<tr>
<td>JAK2 V617F positive IMF</td>
<td>10</td>
</tr>
<tr>
<td>JAK2 V617F negative PV</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 5.1: Number of samples sent to the RZPD for 50K SNP chip analysis from each leukaemic subgroup.

Although my main aim was to investigate the pathogenesis of aMPD/aCML patients, I also took the opportunity to investigate other related patient groups for whom the pathogenesis was completely or largely unknown.

5.2.4 SNP array results

5.2.4.1 Overview of results

Data was returned from all samples along with a series of quality metrics, the most important of which was the SNP call rate. Overall a median of 98.2% SNPs gave readable calls with a range from 91.5% - 99.6% for individual samples. Only 6 samples were below 95%. Analysis of the data led to the identification of copy number and LOH anomalies and the results are summarised in detail in Table 5.4.
below. Using the analysis software it was easy to observe the different causes of LOH when the affected regions were relatively large, i.e. deletions, amplifications or copy number neutral loss (aUPD). In Table 5.4 these have been classified as copy number (CN) changes and acquired uniparental disomy (aUPD).

The main aim of my study was to identify large regions of aUPD as an indicator towards novel oncogenes or tumour suppressor genes. However since we did not have constitutional DNA from most of our cases we could not formally prove that regions of homozygosity were acquired and not inherited. Instead we inferred that they are highly likely to be acquired from the patterns of homozygosity seen in normal individuals. A recent study used SNP arrays to genotype 276 DNA samples derived from lymphocytes of elderly neurologically normal subjects. They found extended regions of homozygosity (contiguous tracts >5Mb) in 9.5% patients but structural genomic variations in 66.9%. Of the patients exhibiting extended homozygosity, 42.3% showed multiple tracts ranging from 5-39Mb that were attributed to parental consanguinity. Regions of homozygosity were rarely greater than 15-20Mb in size (Simon-Sanchez, et al 2007). To avoid regions that may be inherited rather than acquired, we therefore focused primarily on larger regions of copy neutral LOH, which we arbitrarily defined as greater than 20Mb running to a telomere (Figure 5.4), although regions >10Mb running to a telomere were also considered. Smaller, interstitial regions of LOH were also identified within most subgroups however, as previously mentioned, the lack of constitutional matched DNA for any of the samples made it impossible to know if these were acquired or inherited. I therefore discounted these regions from the initial analysis.
Figure 5.1: Ideograms showing UPD. A) The graph on the left shows the copy number trace and indicates no large abnormalities. The graph on the right shows zygosity and shows a large block of homozygous SNP calls at 9p. Since there is no copy number change it is likely that the homozygosity arose by aUPD; B) Apparently patchy 11q aUPD observed in some patients that may to be due to poor DNA quality or a high proportion of background normal cells affecting the SNP calls.
<table>
<thead>
<tr>
<th>Leukaemic subgroup</th>
<th>Identifier</th>
<th>aUPD</th>
<th>CN imbalances</th>
</tr>
</thead>
<tbody>
<tr>
<td>aCML/aMPD</td>
<td>E110, E129, E639</td>
<td>7q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P951</td>
<td>1p (~15Mb)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E160</td>
<td>17q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E484, E632, E1191 (granulocyte pellet)</td>
<td>11q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E632</td>
<td>8p (~15Mb)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P1304</td>
<td>20q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E180</td>
<td>whole chr 13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E155</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E1445</td>
<td>21q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P1213</td>
<td>Del 20q, 7q, Del 17q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E674</td>
<td>Del Chr 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P981, E139</td>
<td>Del 8p</td>
<td></td>
</tr>
<tr>
<td>Myeloid blast crisis</td>
<td>mBC3</td>
<td>15p (~10Mb)</td>
<td>Del 17p</td>
</tr>
<tr>
<td></td>
<td>mBC6</td>
<td>Del 2q, Del 11p</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mBC10</td>
<td>Del 7q</td>
<td></td>
</tr>
<tr>
<td>Lymphoid blast crisis</td>
<td>LyBC1</td>
<td>Gain Chr1, Del 16q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LyBC5</td>
<td>Del 9p, Del 11q and Del 20q.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LyBC8</td>
<td>Del 7p, Del 9p</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LyBC10</td>
<td>Del 16q, Gain 1q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LyBC11</td>
<td>Del 7p, Del 9p</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td><strong>JAK2 V617F negative IMF</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E996</td>
<td>1pUPD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E887</td>
<td>9pUPD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>04/1345</td>
<td>11qUPD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>04/4885</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>01/4124</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>01/1755</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>JAK2 V617F positive (homozygous) IMF</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E1936</td>
<td>9p</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E1530</td>
<td>9p</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E1302</td>
<td>9p</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E1190</td>
<td>9p</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E904</td>
<td>9p</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0604544</td>
<td>9p</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0512361</td>
<td>Low level 9p</td>
<td></td>
</tr>
<tr>
<td></td>
<td>00/1180</td>
<td>9p and 5q</td>
<td></td>
</tr>
<tr>
<td></td>
<td>04/3974</td>
<td>9p</td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.2:** Summary of SNP array results. aUPD was scored as regions of LOH greater than 20Mb running to a telomere with no change in copy number. Regions of extended copy neutral homozygosity that were between 10Mb and 20Mb but which extended to a telomere are also shown. Copy number (CN) anomalies (deletions and amplifications) are also shown.
Cytogenetic information was only available for some cases and in general the correspondence between visible cytogenetic abnormalities and SNP array results, particularly deletions, was very good. However since the main focus of my analysis was to identify regions of aUPD and other groups have already shown a good correlation between array results and karyotype, I did not perform a detailed comparison of the two techniques. Candidate regions of aUPD for each patient subgroup are considered in detail in the following sections.

5.2.4.2  \textit{JAK2 V617F negative Polycythaemia Vera}

About 5\% of PV patients are negative for the V617F JAK2 mutation. By screening a small subset of V617F negative patients I hoped to identify another region of recurrent aUPD or imbalance that might indicate a novel chromosomal area harbouring a new candidate gene. However I observed no regions of aUPD or copy number changes in the 10 cases I analysed.

Subsequently it has emerged that many V617F negative PV cases have mutations in exon 12 of the JAK2 gene (Butcher, \textit{et al} 2008, Pietra, \textit{et al} 2008, Scott, \textit{et al} 2007a, Scott, \textit{et al} 2007b, Williams, \textit{et al} 2007) and further analysis by our group indicated that two of the cases analysed were exon 12 mutation positive. However in contrast to V617F, these exon 12 mutations are not generally reduced to homozygosity by aUPD.

5.2.4.3  \textit{JAK2 V617F negative myelofibrosis}

Of the 18 cases analysed, two were apparently positive for regions on 1p and 9p, respectively. The 9p abnormality was a surprise as this is the location of JAK2, and negativity for V617F was confirmed. One possibility was that the homozygous tract was marking the presence of a novel JAK2 mutation, however since the patient was local a mouthbrush sample was obtained in order to see if the small region of homozygosity was really acquired. Rather than performing expensive array analysis
on the mouthbrush sample, microsatellite PCR markers were used to assess the region of 9p LOH in both the mouthbrush DNA sample, the PB DNA sample as well as three normal controls (Table 5.5).

<table>
<thead>
<tr>
<th>Chr location</th>
<th>Marker</th>
<th>E887 (PB)</th>
<th>E2357 (MOUTHBRUSH)</th>
<th>NORMAL1</th>
<th>NORMAL2</th>
<th>NORMAL3</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEL</td>
<td>D9S1779</td>
<td>HOM</td>
<td>HOM</td>
<td>HET</td>
<td>HOM</td>
<td>HOM</td>
</tr>
<tr>
<td>9p24.3</td>
<td>D9S913</td>
<td>HOM</td>
<td>HOM</td>
<td>HOM</td>
<td>HET</td>
<td>HET</td>
</tr>
<tr>
<td>9p24.3</td>
<td>D9S1858</td>
<td>HOM</td>
<td>HOM</td>
<td>HET</td>
<td>HET</td>
<td>HET</td>
</tr>
<tr>
<td>9p24.3</td>
<td>D9S143</td>
<td>HOM</td>
<td>HOM</td>
<td>HET</td>
<td>HOM</td>
<td>HOM</td>
</tr>
<tr>
<td>9p24.2</td>
<td>D9S1871</td>
<td>HET</td>
<td>HET</td>
<td>HET</td>
<td>HET</td>
<td>HET</td>
</tr>
<tr>
<td>9p24.2</td>
<td>D9S228</td>
<td>HOM</td>
<td>HOM</td>
<td>HOM</td>
<td>HOM</td>
<td>HET</td>
</tr>
<tr>
<td>9p24.2</td>
<td>D9S288</td>
<td>HET</td>
<td>HET</td>
<td>HET</td>
<td>HET</td>
<td>HET</td>
</tr>
<tr>
<td>9p24.2</td>
<td>D9S178</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
</tr>
<tr>
<td>9p24.2</td>
<td>D9S1813</td>
<td>HET</td>
<td>HET</td>
<td>HET</td>
<td>HET</td>
<td>HET</td>
</tr>
<tr>
<td>9p24.1</td>
<td>JAK2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3: Results of microsatellite analysis on the MF case with 9p homozygosity.

This analysis revealed first that the region of homozygosity did not include JAK2 (something that was also obvious from the SNP array results) and second, that the region is also present in the mouthbrush DNA and is thus inherited not acquired.

The remaining patient, E996 had 1p UPD (Figure 5.5A). At about this time, a report by Pikman et. al., (2006) detailed the discovery of a mutation in the thrombopoietin receptor (MPL). The somatic activating mutation in the transmembrane domain of MPL (W515L or W515K) was found in 9% of the JAK2 V617F-negative MF patients. Subsequently other variants were found that affected the same residue. The reason this was of interest to me is that this gene is located on chromosome 1p. I investigated the region of 1p UPD in this patient to confirm if this region harboured this gene and indeed this was the case.

We had an assay set up for investigating this mutation by pyrosequencing so I used this to initially test for W515L or W515K mutations (Figure 5.5B). The pyrosequencing trace shows my patient (E996) has a significantly higher percentage of the ‘T’ allele compared to the normal control indicating the G>T mutation.
(W515L). This was further confirmed by sequencing (Figure 5.5C), which clearly indicates that the normal G has been completely replaced by a T. This therefore provides another example of aUPD being associated with a particular oncogenic change.

Figure 5.5: A) the two graphs generated by the customised excel tool supplied by Professor Hall detailing the SNP chip results for chromosome one in this patient (E996). The graph on the left shows the copy number, whilst the graph on the right details regions of loss of heterozygosity; B) shows the pyrograms for E996 (above) versus a normal control (below). E996 shows a greater proportion of the T allele compared to the normal control indicating the G>T mutation; C) the mutation was further confirmed by sequencing. The trace on the left shows the wild-type sequence and the trace on the right E996 showing the homozygous G>T mutation.
5.2.4.4  JAK2 V617F positive myelofibrosis

It is unclear why the V617F mutation is associated with distinct clinical phenotypes, (PV, ET and MF) but it has been suggested that these are multistep diseases with other genetic abnormalities that remain to be identified. In particular, PV and ET may evolve to MF and the appearance of new cytogenetic abnormalities during disease evolution suggests that the transition may be driven by new mutations. I therefore examined 10 V617F positive MF cases with high levels of the mutant clone (>50%) to see if any other abnormalities were apparent.

Of the 10 JAK2 cases, 9 exhibited 9p LOH, all of which showed no change in copy number. The single case that did not show 9p LOH was only 53% V617F positive and hence was probably heterozygous. In addition to 9p, one case also had clear 5q aUPD (Figure 5.6). There are no known haematologically relevant oncogenes at 5q and currently the significance of this finding remains unclear.

**Figure 5.6:** Figure showing the SNP chip results for Patient 00/1180. The patient shows 9p aUPD and 5q aUPD determined by the large region of LOH on the graphs on the right, compared to the left-hand graphs depicting copy number changes of which there are none.
In addition, a second patient had an additional copy number imbalance, specifically deletion of 20q. Deletion of 20q is a known recurrent finding in a range of MPDs.

5.2.4.5  CML in transformation

Although BCR-ABL is probably the sole abnormality that drives chronic phase CML, it is widely believed that a series of additional genetic events precipitate transformation to acute leukaemia, or blast crisis. I analysed 20 cases (10 in lymphoid transformation and 10 in myeloid transformation) to see if these putative changes are associated with aUPD. In fact, none of the lymphoid cases showed copy neutral LOH and just one of the myeloid cases showed one possible region at the tip of 15p that was approximately 10 Mb in size. Unfortunately there was no chronic phase or constitutional DNA available to confirm that this was acquired. Several copy number changes were identified that correlated with the results of chromosome analysis.

5.2.4.6  aCML/aMPDs

Strikingly, this subgroup showed a high frequency of aUPD compared to other patient groups. Of the 30 cases analysed, 12 (40%) were affected by aUPD (including two cases with regions estimated to be 15 Mb in size). In one case (E1191) it was possible to show that the abnormality was indeed acquired by array analysis of DNA extracted from mononuclear cells (MNC; includes B-cells and T-cells that are not usually involved in the disease) and DNA extracted from granulocytes pellet (which would be expected to be predominantly clonal). As shown on Figure 5.7, 11q LOH is clearly visible in granulocyte DNA but is not visible or barely detectable in MNC DNA.
Figure 5.7: Comparison of MNC and granulocyte DNA for patient E1191. No copy number changes are observed in either DNA sample (first and third panels) however clear 11q LOH is observed in the DNA extracted from granulocytes (fourth panel) that is absent in MNC DNA (second panel), indicating an acquired abnormality.

In total, eight different chromosomal regions were affected by aUPD suggesting substantial genetic heterogeneity in these diseases. Recurrent regions were identified at 7q and 11q (3 cases of each, i.e. 10% of cases) and single cases with aUPD were seen at 1p, 8p, whole chromosome 13, 17q, 20q and 21q.

One patient, E632, showed both whole arm 8p aUPD and 11q aUPD. This patient also showed many regions of interstitial LOH, far more than any other patient (Figure 5.8). Limited clinical history suggests this patient may be consanguineous, suggesting that at least some of these multiple homozygous regions may be autozygous, i.e. inherited rather than acquired.
Figure 5.8: Ideograms depicting E632 SNP array results. A) Top panels show tracts of homozygosity at 11q and 8p, both of which extend to a telomere; B) bottom panels show an example of two of the ten interstitial regions of homozygosity.

One patient (E180) showed whole chromosome 13 aUPD (Figure 5.9a). The receptor tyrosine kinase *FLT3* is located on this chromosome and is the most frequently mutated gene in acute myelogenous leukemia (AML), usually associated with an internal tandem duplication (ITD) of the region encoding the juxtamembrane domain. Furthermore it has already been reported that aUPD for chromosome 13 is associated with *FLT3* mutations in AML (Griffiths, *et al* 2005). The ITD is within exon 14 and/or 15 of the *FLT3* gene, so primers were designed initially to amplify exon 14. As shown on (Figure 5.9b), patient E180 is indeed homozygous for a *FLT3* ITD.
Figure 5.9: A) Analysis of aMPD patient E180. The graph on the left shows there has been no copy number change, whilst the graph on the right indicates complete chromosome 13 homozygosity; B) PCR assay for FLT3 ITD. Lanes 1 & 2 represent patient E180, lane 3 is a positive control (an AML patient with a known FLT3 ITD and lane 4 is a normal control. The bottom band is the normal FLT3, the top band is FLT3 with an ITD.

Other regions of aUPD could not be simply linked to known mutations. My analysis focused on the recurrent regions at 7q and 11q and these are discussed in detail below. The abnormalities at 1p, 8p, 17q, 20q and 21q were not investigated. The region at 17q may target *NF1*, a gene that has long been associated with juvenile myelomonocytic leukaemia (JMML) (Side, *et al* 1998). More recently, Flotho et. al., (2007) screened JMML patients and found regions (>55Kb) of copy number neutral 17q LOH leading to loss of the normal *NF1* allele that was associated with an increased incidence of JMML in NF1 patients. However, *NF1* is a large gene with 58 exons and therefore very expensive to screen for mutations (Flotho, *et al* 2007).
In addition to large regions of copy number neutral LOH, copy number imbalances were also observed in aMPD cases. One patient exhibited three deletions of the long arms of chromosomes 20, 17 and 7. As discussed previously deletions of chromosome 20q are a common abnormality associated with myeloid disorders as are deletions of 7q (Luna-Fineman, et al 1995). The 17q deletion may again be associated with an NF1 mutation. Another patient was also found to have a deletion of entire chromosome 7, and two patients had a deletion of 8p. Deletions of 8p have been described previously in association with MDS and the evolution of CML (Kerndrup, et al 1987, Pedersen and Kerndrup 1986, Wada, et al 1994).

5.3 RECURRENT aUPD OF CHROMOSOME 7q

5.3.1 7q abnormalities in myeloid disorders

LOH for chromosome 7, commonly in the form of monosomy 7 or deletion of the long arm of chromosome 7 is a very frequent occurrence in a variety of human malignancies including gastric cancer (Weng, et al 2006), breast cancer (Zenklusen, et al 1994), prostate cancer (Latil, et al 1995), colon cancer and primary cell carcinoma of the head and neck (Zenklusen, et al 1995). It has been hypothesised that a tumour suppressor gene is located at 7q that is ubiquitously expressed thereby accounting for the numerous cancers associated with abnormalities of chromosome 7.

tumor suppressor gene. Furthermore, at least three distinct minimal regions of LOH have been identified suggesting that multiple loci may be important. Some have also argued that the complexity of 7q rearrangements might indicate that this region is prone to breakage and instability as a consequence of different genetic factors that play a general role in oncogenesis rather than the presence of a specific tumour suppressor gene involved in myeloid disorders (Basirico, et al 2003, Gonzalez, et al 2004).

5.3.2  
**Recurrent 7q aUPD identified by SNP chip analysis**

As a result of screening 30 aCML/aMPD patients using Affymetrix 50K SNP arrays I identified three cases with varying sized regions of aUPD of chromosome 7q (Figure 5.10).

---

**Figure 5.10:** Copy number and LOH graphs produced for chromosome 7 for three aCML/aMPD patients: E639, E129 and E110. In all three cases the copy number graph shows no
copy number changes. All three LOH charts show varying sized regions of LOH affecting the q arm of chromosome 7.

No copy number alterations or aUPD was observed elsewhere in the genome of these patients. Cytogenetics was only known for E639 and this was normal. Patients E110 and E129 had regions of aUPD spanning approximately 98Mb, almost the entire length of the long arm of chromosome 7, specifically 7q11.21-tel and 7q11.22-tel, respectively (Figure 5.11). The third patient has a smaller region of partial 7q aUPD spanning 53Mb from cytogenetic bands 7q22.3-tel. This region represents the minimal overlapping region of aUPD observed in all three patients (Figure 5.11).

<table>
<thead>
<tr>
<th>Patient Identifier</th>
<th>Region of LOH/Chromosomal region</th>
<th>Size of region of LOH</th>
</tr>
</thead>
<tbody>
<tr>
<td>E110</td>
<td>63,695,332-158,554,645 (q11.21-qter)</td>
<td>94,859,313 (95Mb)</td>
</tr>
<tr>
<td>E129</td>
<td>71,178,702-158,554,645 (q11.22-qter)</td>
<td>87,375,943 (87Mb)</td>
</tr>
<tr>
<td>E639</td>
<td>105,954,360-158,554,645 (q22.3-qter)</td>
<td>52,600,285 (53Mb)</td>
</tr>
</tbody>
</table>

Figure 5.11: Schematic representation of chromosome 7 and the regions of aUPD (highlighted in red) observed in the three aCML/aMPD patients E110, E129 and E639.
5.3.3 Screening for mutations in 7q aUPD patients

The minimal region 53Mb is still very large and encompasses approximately 400 genes so prioritising candidate genes for screening was of primary concern. Due to the growing evidence linking aberrant tyrosine kinase signalling to the MPDs, I began by screening genes encoding tyrosine kinases and hoped to find mutations that were homozygous as a consequence of the aUPD.

5.3.3.1 Candidate tyrosine kinases located within the region of 7qUPD

There are 3 tyrosine kinase genes in the minimal region, MET (7q31.2), EPHA1 (7q34-35) and EPHB6 (7q34). The MET proto-oncogene, encoding the tyrosine kinase receptor for the hepatocyte growth factor (HGF), controls cell growth, invasiveness, and protection from apoptosis. Deregulated MET signalling has long been associated with tumour progression and metastasis in a variety of tumours (Jeffers, et al 1996), as well as being deregulated in a few cases of human leukaemia and lymphoma (Jucker, et al 1994). Activating MET mutations have also been identified in hereditary and sporadic forms of papillary renal carcinoma (Schmidt, et al 1997). Recently, missense mutations have also been identified in this gene in a variety of human cancers (Lee, et al 2000, Ma, et al 2003). This seemed like an ideal candidate to screen for possible mutations in these three patients with 7q UPD.

The Ephrin receptors and their ligands, the ephrins, regulate numerous developmental and cellular processes, particularly in the nervous system. They are divided into two groups based on their sequence similarities and structures, the ephrin-A (EFNA) class, which are anchored to the membrane by a glycosylphosphatidylinositol linkage, and the ephrin-B (EFNB) class, which are transmembrane proteins. They make up the largest subgroup of receptor tyrosine kinases. Although relatively little is known about these receptors over-expression and loss of expression of some family members have been related to specific cancers (Noren, et al 2006, Song, et al 2007).
5.3.3.2 Expression analysis

It would be expected that any relevant mutated gene would be expressed in haematopoietic progenitor cells and probably in myeloid cells. As a first step towards prioritising these three genes, the expression of each was checked by RT-PCR using cDNA made from peripheral blood (PB) and bone marrow (BM) from 5 cases with other haematological disorders. cDNA-specific primers were designed to amplify the kinase domain (See Appendix I, section XXI) for all three receptor tyrosine kinases. Standard PCR conditions were used (annealing temperature of 66°C and 30 cycles). Clear products for EPHA1 and EPHB6 were seen in all samples but no amplification was seen for MET. I therefore focused initially on analysis of the two Ephrin receptors.

Genomic DNA primers were designed, again to anneal at 66°C, to cover exons encoding the transmembrane and kinase domain as these are known hotspots for oncogenic mutations (for all primer sequences see Appendix I, section XXII). PCR products were directly sequenced and compared to the RefSeq entries for each gene. No mutations were identified so further primers were designed to cover all exons of each gene. As summarised on Table 5.6, no abnormalities were detected.

<table>
<thead>
<tr>
<th>Exon number</th>
<th>E110</th>
<th>E129</th>
<th>E639</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPHA1.5'UTR.F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int1.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int2.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int3.1F</td>
<td>T&gt;C @479 HOM V160A</td>
<td>T&gt;C @479 HOM V160A</td>
<td>T&gt;C @479 HOM V160A</td>
</tr>
<tr>
<td>EPHA1.Int4.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int5.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int6.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int7.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int8.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int9.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int10.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int11.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int12.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int13.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int14.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int15.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>EPHA1.Int16.1F</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>
Table 5.6:  Mutation screening of EPHA1 and EPHB6. The variant found in all three patients in exon 4 of EPHA1 is a known SNP reported in various databases. Exon 1 of EPHA1 proved difficult to amplify, so it was cloned and ten colonies sequenced for each patient.

Although the MET tyrosine kinase was not obviously expressed in the PB or BM of controls, the possibility that its expression might be limited to a small population of myeloid precursor cells plus the strong association of activated tyrosine kinases with MPDs persuaded me that I should screen this gene as well. Genomic DNA primers were designed for the entire coding region of MET (Appendix I, section XXII). An identical intronic sequence change was identified upstream of exon 6 but was ruled out as causative due to the location, 60bp away from the exon start site. Other variants were identified in exons 20 and 21 but corresponded to known reported SNPs. The results are summarised in Table 5.7.

| EPHA1.Int17.1F | N | N | N |
| EPHB6.Int 5'UTR.1F | N | N | N |
| EPHB6.Int 1.1F | N | N | N |
| EPHB6.Int 2.1F | N | N | N |
| EPHB6.Int3.1F | N | N | N |
| EPHB6.Ex4.1F | N | N | N |
| EPHB6.Int4.1F | N | N | N |
| EPHB6.Int5.1F | N | N | N |
| EPHB6.Int6.1F | N | N | N |
| EPHB6.Int7.1F | N | N | N |
| EPHB6.Int8.1F | N | N | N |
| EPHB6.Int9.1F | N | N | N |
| EPHB6.Int10.1F | N | N | N |
| EPHB6.Int11.1F | N | N | N |
| EPHB6.Int12.1F | N | N | N |
| EPHB6.Int13.1F | N | N | N |
| EPHB6.Int14.1F | N | N | N |
| EPHB6.Int15.1F | N | N | N |
| EPHB6.Int16.1F | N | N | N |
Table 5.7: Mutational analysis of the MET gene in the three aMPD patients with 7q aUPD. All exons were directly sequenced apart from exon 14 which was subcloned and then sequenced.

Although it is conceivable that either EPHA1, EPHB6 or MET could be activated by another mechanism which cannot be detected by the strategy I used, e.g. a deep intronic mutation giving rise to a splicing abnormality or an internal deletion/duplication that encompasses at least one of the primer binding sites, it is much more likely that these genes are normal. Since these are the only tyrosine kinases on 7q, I therefore focused on the prioritisation of other candidate genes.
To prioritise further genes I decided to look at those located within the minimal region of aUPD that impact on tyrosine kinase signalling pathways and for which there was a known link with leukaemia or cancer. There were no obvious candidates specifically implicated in leukaemia but there were two genes encoding signalling components more generally associated with cancer: *BRAF* and *PIK3CG*.

The *BRAF* gene is located at 7q34 and seemed an obvious candidate as it encodes a member of the RAS-RAF-MEK-ERK-MAP kinase signalling pathway which is responsible for mediating the transduction of mitogenic signals from the cell membrane to the nucleus. It is a cytoplasmic member of the serine-threonine protein kinase family that is regulated by binding to RAS. Davies et. al., (2002) were the first to report mutations in this gene by high throughput sequence analysis. They screened genomic DNA from a number of different cancer cell lines and identified *BRAF* somatic missense mutations in a wide variety of cancers, notably malignant melanoma (66% of cases). All mutations were within the kinase domain, with a single missense mutation (V599E) accounting for over 80% of the mutations observed. Of 53 leukaemia/lymphoma cell lines analysed no mutations were identified. However, it has been suggested that these mutations are commonly found in cancers that are also known to have a high frequency of *RAS* mutations (Davies, *et al* 2002). *RAS* mutations are seen in a substantial proportion of JMML and CMML patients (Bowen, *et al* 2005, Emanuel 2004, Emanuel 2008). Additionally, Christiansen et. al., (2005) screened 29 AML patients and identified 3 (21%) with *BRAF* mutations so it was entirely possibly that mutations in this gene may be found in patients with MPDs. Since the original report, *BRAF* mutations have been identified in a wide variety of human cancers including papillary thyroid carcinomas (Espinosa, *et al* 2007), non-small cell lung carcinomas (Brose, *et al* 2002), and colorectal carcinomas (Oikonomou and Pintzas 2006) with all mutations being located in either exons 11 or 15.
Based on these reports I decided to screen my aMPD patients with 7q aUPD for these two BRAF exons. Primers were taken from Brose et. al., (2002) and performed as suggested, using standard PCR conditions with a 58°C annealing temperature and 30 cycles, however no sequence variants were found. These negative results fit with a recent report by de Vries et. al., (2007) who screened 65 JMML patients for mutations of the BRAF gene but found no mutations (de Vries, et al 2007).

The second candidate I considered was PIK3CG (PI3-kinase p110 subunit gamma), located at 7q22.3. PIK3CG is highly related to PIK3CA at 3q26, which is commonly mutated in colorectal cancer (Oikonomou and Pintzas 2006) and other cancer types including bladder tumours, breast & ovarian cancer, and brain tumours (Bachman, et al 2004, Broderick, et al 2004, Campbell, et al 2004, Lopez-Knowles, et al 2006). Most of the mutations are clustered in two hotspots within the gene, as seen in Figure 5.12. The mutations are thought to confer resistance to apoptosis and facilitate tumour invasion of solid tumours.

Figure 5.12: Adapted from Samuels et. al., (2004) (Samuels and Velculescu 2004). Clustering of reported mutations in PIK3CA identified in various human cancers. The arrow heads indicate the location of reported missense mutations. The percentage of mutations identified in each tumour group from this paper is indicated.

Although no mutations have been described in PIK3CG, I decided to analyse the regions that were homologous to the hotspots in PIK3CA. Following alignment of the
two genes, I designed primers to amplify exons 2 and 11 of *PIK3CG*, which are homologous to exons 9 and 20 of *PIK3CA*. All primers were designed to anneal at 66°C using standard PCR conditions with 30 cycles (Appendix I, section XXIII) but again no mutations were found.

Having eliminated the likely hotspots in *BRAF* and *PIK3CG*, I considered full mutation screens and also created a shortlist of possible new candidate genes for analysis (Table 5.8). However at this point, the analysis of genes within the second recurrent region of aUPD at 11q was generating interesting results and this is where I focused my attention.

<table>
<thead>
<tr>
<th>Accession No.</th>
<th>Gene Name</th>
<th>Chr. Location</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_005515</td>
<td><em>HLXB9</em></td>
<td>q36.3</td>
<td>Homeobox protein HB9. Fused to ETV6 in infant AML. Activates IL6 via PI3K in Hodgkins Lymphoma. Also fused to MYB. Mutations found in Currarino syndrome.</td>
</tr>
<tr>
<td>NM_002847; NM_130843; NM_130842</td>
<td><em>PTPRN2</em></td>
<td>q36.3</td>
<td>Receptor-type tyrosine-protein phosphatase N2 precursor. It is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signalling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation.</td>
</tr>
<tr>
<td>NM_194071</td>
<td><em>CREB3L2</em></td>
<td>q33-q34</td>
<td>cAMP responsive element binding protein 3-like. The chimeric FUS/CREB3L2 gene is specific for low-grade fibromyxoid sarcoma. It belongs to a family of transcription factors. Fused to FUS in liposarcomas</td>
</tr>
<tr>
<td>NM_178562</td>
<td><em>TSPAN33</em></td>
<td>q32.1</td>
<td>Tetraspanin 33, alternatively known as Penumbra whose mouse homologue has growth suppressive activity</td>
</tr>
<tr>
<td>NM_170606; NM_021230</td>
<td><em>MLL3</em></td>
<td>q36.1</td>
<td>Myeloid/lymphoid or mixed-lineage leukemia 3. It is involved in transcriptional co-activation. May be involved in leukaemogenesis and developmental disorder.</td>
</tr>
<tr>
<td>NM_198267; NM_019071</td>
<td><em>ING3</em></td>
<td>q31.31</td>
<td>Inhibitor of growth protein 3 (p47ING3 protein). The protein is similar to ING1, a tumor suppressor protein that can interact with TP53, inhibit cell growth, and induce apoptosis. This gene can activate p53 trans-activated promoters, including promoters of p21/waf1 and bax. Over-expression of this gene has been shown to inhibit cell growth and induce apoptosis. Allelic loss and reduced</td>
</tr>
</tbody>
</table>
expression of this gene were detected in head and neck cancers.

| NM_004935 | CDK5 | q36.1 | Cell division protein kinase 5 (Tau protein kinase II catalytic subunit). Probably involved in the control of the cell cycle. |

**Table 5.8:** Table showing the gene name, accession number and chromosomal location of some potential candidate genes on 7q. A brief list of known facts relevant to why these genes were short listed is also included.

### 5.4 Recurrent aUPD of chromosome 11q

#### 5.4.1 11q abnormalities in myeloid disorders

Like chromosome 7, chromosome 11 has also long been reported as a region commonly affected by imbalances in cancer. Both gains and losses of chromosome 11 have been described. Gains of 11q are less common than losses but are often seen in neuroblastoma cell lines (Carr, *et al* 2007). LOH of this region is frequent and associated with cervical, ovarian, breast, colon and skin cancers (Bethwaite, *et al* 1995, Carter, *et al* 1994, Foulkes, *et al* 1993, Hampton, *et al* 1994, Keldysh, *et al* 1993, Tomlinson, *et al* 1993). By far the most frequent region of imbalance reported is centered round the cytogenetic band 11q23, in solid tumours as well as haematological malignancies such as T-cell prolymphocytic leukemia (T-PLL), B-cell chronic lymphocytic leukemia (B-CLL) and mantle cell lymphoma (MCL). Amplifications of specific regions of 11q have also been identified in cases of AML and MDS. One group identified 11q genomic gain in more than 40% of AML cases with high level amplification again centered on 11q23, whilst another group reported the rare but recurrent occurrence of 11q amplifications of 11q23 and 11q13 in AML and MDS cases (Rucker, *et al* 2006, Zatkova, *et al* 2006). 11q LOH associated with leukaemia has also been described (Siu, *et al* 2000, Takeuchi, *et al* 2003).
5.4.2 **Recurrent 11q aUPD identified by SNP chip analysis**

From screening 30 aMPD patients with Affymetrix 50K SNP arrays I identified three aCML/aMPD patients with varying sized regions of aUPD at chromosome 11q (Figure 5.13).

**Figure 5.13:** 50K SNP array analysis of chromosome 11 in three aCML/aMPD patients: E484, E632 & E1191. The graphs on the left represent the copy number distribution observed across the chromosome whilst the graphs on the right hand side represent the loss of heterozygosity observed. All three have varying degrees of LOH with no corresponding copy number alterations, indicating aUPD.

Previous cytogenetics was only available for one of the patients, E484, and this was found to be normal. All three patients had partial 11q aUPD of varying size as
illustrated on Figure 5.14. E632 and E1191 had the largest regions spanning approximately 69Mb, almost the entire long arm of chromosome 11. Cytogenetically, these correspond to 11q13.1-qter and 11q12.2-qter in E632 and E1191, respectively (Figure 5.14). E484 had a slightly smaller region of aUPD spanning 57Mb and covering cytogenetic bands 11q13.5-qter. This therefore represents the minimal region and includes about 350 known genes.

### Table 5.9: Patient Identifier, Region of LOH/Chromosomal region, Size of region of LOH

<table>
<thead>
<tr>
<th>Patient Identifier</th>
<th>Region of LOH/Chromosomal region</th>
<th>Size of region of LOH</th>
</tr>
</thead>
<tbody>
<tr>
<td>E484</td>
<td>76,606,755-134,082,843 (q13.5-qter)</td>
<td>57476088 (57Mb)</td>
</tr>
<tr>
<td>E632</td>
<td>64,698,446-134,082,843 (q13.1-qter)</td>
<td>69384397 (69Mb)</td>
</tr>
<tr>
<td>E1191 (Granulocyte)</td>
<td>61,241,557-134,082,843 (q12.2-qter)</td>
<td>67413720 (67Mb)</td>
</tr>
</tbody>
</table>

**Figure 5.14:** Schematic representation of chromosome 11 and the regions of aUPD observed in the three aCML/aMPD patients E484, E632 & E1191.

### 5.4.3 Screening 11q aUPD patients

#### 5.4.3.1 Candidate genes located within the region of 11q aUPD

Interestingly no tyrosine kinase genes are located within the minimal region of 11q aUPD, or indeed anywhere on 11q. Instead, I looked for genes associated with leukaemia or cancer in general within the minimal region. As expected there were several possible candidates (Table 5.9). However, only a few really stood out due to
previous reports of defects associated with specific leukaemias or very specific pathway links to receptor tyrosine kinases (those highlighted in yellow).

<table>
<thead>
<tr>
<th>Accession No.</th>
<th>Gene Name</th>
<th>Chr. Location</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_002576</td>
<td>PAK1</td>
<td>q14.1</td>
<td>p21-activated kinase 1, a serine/threonine kinase.</td>
</tr>
<tr>
<td>NM_138292</td>
<td>ATM</td>
<td>q22.3</td>
<td>Belongs to the PI3/PI4-kinase family. It is an important cell cycle checkpoint kinase. Defects in <em>ATM</em> contribute to T-cell acute lymphoblastic leukemia (TALL) and T-prolymphocytic leukemia (TPLL). Defects in <em>ATM</em> also contribute to B-cell non-Hodgkin lymphomas (BNHL), including mantle cell lymphoma (MCL) and B-cell chronic lymphocytic leukemia (BCLL).</td>
</tr>
<tr>
<td>NM_003904</td>
<td>ZNF259</td>
<td>q23.2</td>
<td>May be a signalling molecule that communicates mitogenic signals from the cytoplasm to the nucleus. Binds to the EGF and PDGF receptors.</td>
</tr>
<tr>
<td>NM_005933</td>
<td>MLL</td>
<td>q23.3</td>
<td>Myeloid/lymphoid or mixed-lineage leukemia. Possibly acts as a transcriptional regulatory factor. Numerous chromosomal aberrations involving MLL are a cause of acute leukaemias. To name a few, translocation t(1;11)(q21;q23) with MLLT11/AF1Q; translocation t(3;11)(p21;q23) with NCKIPS/AF3p21; translocation t(11;19)(q23;p13.3) with MLLT1/ENL; translocation t(11;19)(q23;p23) with GAS7; translocation t(X;11)(q13;q23) with MLLT7/AFX1.</td>
</tr>
<tr>
<td>NM_005188</td>
<td>CBL</td>
<td>q23.3</td>
<td>Participates in signal transduction in haematopoietic cells. Adapter protein that functions as a negative regulator of many signalling pathways that starts from receptors at the cell surface. Acts as an E3 ubiquitin-protein ligase, which accepts ubiquitin from specific E2 ubiquitin-conjugating enzymes, and then transfers it to substrates promoting their degradation by the proteasome. Recognises activated receptor tyrosine kinases, including PDGFA, EGF and CSF1, and terminates signalling. Can be converted to an oncogenic protein by deletions or mutations that disturb its ability to down-regulate RTKs.</td>
</tr>
<tr>
<td>NM_032873.3</td>
<td>STS-1</td>
<td>q24.1</td>
<td>Encodes a protein that contains an ubiquitin associated domain at the N-terminus. The encoded protein was found to inhibit endocytosis of epidermal growth factor receptor (EGFR) and platelet-derived growth factor receptor.</td>
</tr>
<tr>
<td>NM_005238</td>
<td>ETS1</td>
<td>q24.3</td>
<td>v-ets erythroblastosis virus E26 oncogene. ETS is responsible for erythroblast and fibroblast transformation. The juxtaposition of the interferon and c-ETS-1 proto-oncogene may be involved in the pathogenesis of human monocytic leukemia.</td>
</tr>
</tbody>
</table>
Transforming growth factor-beta (TGF-beta) is perhaps the most potent endogenous negative regulator of haematopoiesis.

Inhibits apoptosis by binding to tumor necrosis factor receptor-associated factors TRAF1 and TRAF2, probably by interfering with activation of ICE-like proteases.

This gene encodes a protein that is a member of the cysteine-aspartic acid protease (caspase) family. Sequential activation of caspases plays a central role in the execution-phase of cell apoptosis.

As above. Over-expression of the active form of this enzyme induces apoptosis in fibroblasts.

As above.

A chromosomal aberration involving POU2AF1/OBF1 may be a cause of a form of B-cell leukemia. Translocation t(3;11)(q27;q23) with BCL6.

Friend leukemia virus integration 1. A chromosomal aberration involving FLI1 is a cause of Ewing sarcoma. Translocation t(11;22)(q24;q12) with EWS.

Table 5.9: Initial list of candidate genes on 11q.

Of the three highlighted genes (yellow), I decided to begin with the mixed-lineage leukemia (MLL) gene due to its strong links with acute leukaemia, and its location at 11q23.

5.4.3.2 Screening the Mixed lineage leukaemia (MLL) gene

The MLL gene, otherwise known as ALL-1, HTRX1 or HRX amongst others, encodes a DNA-binding protein that methylates histone H3 lysine 4 as well as positively regulating expression of numerous HOX genes, which are in turn essential for haematopoietic development (Krivtsov and Armstrong 2007). Rearrangements of MLL were first recognised as a consequence of chromosomal translocation in AML (Ziemin-van der Poel, et al 1991). In fact there have now been reports of translocations involving MLL and greater than 50 different fusion partners (Meyer, et al 2006). In addition, partial tandem duplications (PTD) within MLL gene have also been reported in AML with a normal karyotype or trisomy 11 (Caligiuri, et al 1994,
Schichman, et al 1994). Numerous in frame MLL-PTDs have been identified and named according to the fused exons, the most frequent being e9/e3 (under previous nomenclature this was e2/e6), followed by e11/e3 (e2/e8) and e10/e3 (e2/e7). The MLL-PTD is found in approximately 5-20% of adult AML with a normal karyotype and significantly higher proportion (25-54%) in adult AML patients with trisomy 11. Although the precise role of the MLL-PTD in leukaemogenesis remains a matter of debate (Marcucci, et al 1998, Schnittger, et al 1998, Weisser, et al 2005) I decided to see if the 11q aUPD patients were positive for the MLL-PTD.

The single step primers, 3.1c and E3AS, were ordered according to the Schichman et. al., (1994) paper, and nested primers, MLLint and 6.1, were later ordered according to the Schnittger et. al., (1998) paper. RT-PCR was performed as described in Schnittger et. al., (1998) using cDNA from each of the three 11q aUPD patients plus a positive control, the eosinophilic cell line EOL-1 that is known to harbour a MLL-PTD (Schichman, et al 1994, Schnittger, et al 1998). Applied Biosystems AmpliTaq Gold and buffer were used as an alternative to that stated in the paper. The results of the RT-PCR are shown below in Figure 5.15. As can be seen, the EOL-1 cell line clearly shows an MLL-PTD by standard single step PCR, whilst none of the aMPD patients with 11q UPD show any sign of a MLL-PTD by either single step or nested RT-PCR (gel picture not shown).
Figure 5.15: Exon structure of the 5' region of the MLL gene illustrating the common e9/e3 PTD (only the new nomenclature is detailed). The boxes represent individual exons, those in blue are the WT and those in green indicate the PTD. The red arrows indicate the primers as taken from Schnittger et. al.,(1998). The gel picture below shows an Invitrogen 1kb plus ladder (lane 1) and the product observed from all three patients with 11q aUPD (lanes 3-5) versus the cell line EOL-1 (lanes 1-2). The EOL-1 cell line has been shown to harbour a MLL-PTD so was used as a positive control (Saito H 1985).

*MLL* is a very large gene and point mutations have not been reported so I went on to consider other candidates. The *ATM* gene seemed a strong possibility with mutations reported in this gene responsible for genomic instability and an increased risk of cancer. *ATM* gene mutations are seen in subtypes of lymphoid leukaemia and lymphoma (Boultwood 2001) but have not been reported in myeloid disease. The third gene that stood out was *CBL*. This gene encodes a member of the Casitas B-
lineage lymphoma gene family of ubiquitin ligases which function as negative regulators of several receptor protein tyrosine kinase signalling pathways. CBL achieves this by ubiquitinating activated receptor tyrosine kinases, thus promoting their sorting for endosomal degradation or recycling to the plasma membrane. Whilst deliberating which of these candidate genes to screen next, two interesting papers were published reporting the finding of CBL mutations in AML. Sargin et al., found a missense mutations in 1 of 150 cases analysed and demonstrated that the mutation had transforming activity (Sargin, et al 2007). Cagliuri et. al., (2007) found abnormalities in 4 of 12 cases, none of which were functionally analysed (Caligiuri, et al 2007, Langdon, et al 1989). The reason for the marked discrepancy in the proportion of mutation positive cases is unclear, however the technical details for the second paper were very scant and in particular it is unclear whether the sequence changes were verified or not. Nevertheless, given these findings and the known interaction of CBL with tyrosine kinases, I therefore decided to screen this gene next.

5.4.3.3 The CBL gene – introduction

The CBL proto-oncogene (originally called c-CBL) was first discovered as the cellular homologue of v-CBL, a viral transforming gene from the Cas NS-1 murine retrovirus that causes pre-B cell lymphomas and myelogenous leukaemias in mice (Langdon, et al 1989). It is a large, ubiquitously expressed, principally cytoplasmic protein. There are three members of the CBL family: CBL, CBLB and CBLC, all of which play an important role in regulating signalling by many receptor tyrosine kinases. Structurally, the CBL family members are modular proteins that contain a conserved N-terminal tyrosine kinase binding domain (TKB) that has been shown to bind to over 40 proteins involved in cellular regulation. For example it has been shown to bind phosphorylated cytoplasmic kinases such as ZAP70, and several receptor tyrosine kinases such as EGFR and PDGFR. They also have a RING finger domain that has been shown to interact with and recruit the E3 ubiquitin-ligase enzymes (E3) as well as other negative regulators of signalling. Finally, they all contain a proline-rich region that interacts with SH3-domain containing signalling
proteins such as GRB2, SRC and CIN85. As a consequence of this multifunctionality, CBL has been shown to play both a positive and negative role in signalling.

The negative regulatory role of CBL is summarised in the diagram below (Figure 5.16). In brief, activated receptor tyrosine kinases recruit CBL, and CBL mediates the ubiquitination of the receptor on lysine residues. Ubiquitination is the signal that targets the receptors for internalisation via clathrin coated pits, followed by endocytic sorting resulting in either recycling or degradation of the receptor (Shtiegman and Yarden 2003).

In addition, CBL proteins may also play a positive role in mediating positive RTK signalling events to downstream effectors. Schmidt et. al., (2005) have shown that phosphorylation of CBL results in it binding to positive signalling regulators such as SHP-2, Gab2 and PI3-kinase via its C-terminal multi-adaptor domain (Schmidt and Dikic 2005). In other words, CBL can act as a scaffold to the recruitment of downstream effectors of tyrosine kinases.
Receptor tyrosine kinase (RTK) negative regulation and the role of CBL. RTK stimulation with ligand induces dimerisation and hence activation. This leads to autophosphorylation and recruitment of CBL, which itself is then phosphorylated and ubiquitinates the RTK. The ubiquitylated receptors are then sorted into clathrin-coated pits by a multiprotein complex that includes coat adaptors such as Epsin and Eps15. Fission of clathrin-coated vesicles is mediated by a GTPase, dynamin. In addition, CBL-mediated recruitment of CIN85 and endophilin is thought to promote negative membrane curvature and invagination. Shedding of clathrin, a decrease in the internal pH and the accumulation of hydrolytic enzymes, signals progression through the endocytic pathway. The RTK trafficking from early to late endosomes is dependent on its continued association with CBL and its sustained ubiquitination. The endocytic vesicle finally uncoats (facilitated by synaptojanin) and fuses with an early endosome, by a mechanism catalysed by the small GTPase Rab5 and its effectors. From here, the RTK is sorted into intraluminal vesicles of the endosome, in a process mediated by ESCRT-I, ESCRT-II and ESCRT-III. The RTK is probably dephosphorylated and deubiquitinated during this sorting step. The multivesicular body (MVB)/endosomal carrier vesicle is formed in an Annexin-II-dependent fashion, which ultimately fuses with a late endosome or a lysosome. Fusion of the
MVB with the lysosome results in degradation of the contents of the internal vesicles. Recycling of receptors back to the plasma membrane can occur throughout the endocytic pathway, albeit with decreasing efficiency.

5.4.3.4 The CBL gene – screening 11q aUPD patients

Mutations of CBL in AML patients were restricted to the RING domain, encoded by exons 8 and 9. Primers were therefore designed to anneal at 66°C within the introns surrounding these exons to allow the whole exons to be screened using genomic DNA (for primer sequences see Appendix I, section XXIV). PCRs were undertaken in a volume of 25μl using standard conditions and yielded exon 8 and exon 9 PCR products of 424bp and 369bp, respectively (Figure 5.17).

![Figure 5.17](image)

**Figure 5.17:** Agarose gels showing amplification of CBL exon 8 (left) and 9 (right) for all three aMPD patients with 11qUPD. In both cases, the 1Kb plus ladder (Invitrogen) is shown in the first lane to accurately size the products. For both gels: Lane 1, E1191; Lane 2, E632; Lane 3, E484, Lanes 4-6 normal controls, and Lane 7 no DNA control. The correct sized product for CBL exon 8 (~424bp) and exon 9 (~360bp) is seen for all patients and controls.

Direct sequencing identified changes in all three patients. In two patients, E484 and E1191, a T>C change at nucleotide 1277 in exon 8 (Figure 5.18), was seen that is
predicted to change leucine 380 to a proline (L380P). The mutation for E1191 appeared to be predominant with only a small residual wild type T peak visible on the sequence traces. E484 showed a more substantial T peak, which is consistent with the relatively low LOH peaks detected by array analysis as shown in Figure 5.13 above.

![Electropherograms showing part of CBL exon 8 amplified from aCML/aMPD patients E484 (left) and E1191 (right).](image)

**Figure 5.18:** Electropherograms showing part of CBL exon 8 amplified from aCML/aMPD patients E484 (left) and E1191 (right). Each panel show 6 traces: those at the top and bottom (traces 1 and 6) represent the reference CBL sequence (NM_005188). Traces 2 and 5 represent patient sequence in forward and reverse orientations, respectively. The traces in the centre (traces 3 and 4) show a comparison between the reference and the patient sequence highlighting any sequence changes with a peak. Both patients show a T>C mutation at nucleotide 1277 which is predicted to result in the substitution of the amino acid leucine by proline.

The third case (E632) was also found to have a sequence change, this time in exon 9. The mutation is a C>G missense change at nucleotide 1387 that is predicted to result
in the substitution of the proline 417 by alanine (P417A; Figure 5.19). The wild type C residue was barely detectable, consistent with the high level of LOH seen by array analysis. Both sequence changes were different to those described in AML patients.

**Figure 5.19:** Sequence analysis of patient E632 showing a portion of CBL exon 9. The order of the traces is the same as Figure 5.18, with traces 3 and 4 showing the C>G mutation at nucleotide 1387 that results in a P417A amino acid substitution.
5.4.3.5 Screening other MPD patients for CBL mutations

Following the detection of the CBL mutation in all three 11q aUPD patients I wanted to next determine if mutations of this gene might be more widespread in both typical and atypical MPDs. I therefore screened 547 patients as detailed in Table 5.10 by sequence analysis of exons 8 and 9.

I identified mutations in a further 23 patients, making 26 in total. Surprisingly, one patient was found to have a mutation in both exons 8 and exon 9 making a total of 15 cases with exon 8 mutations and 12 cases with exon 9 mutations (Table 5.10). Comparison with MPD subtype indicates that CBL mutations are seen in cases with CMML (12.8%), aCML/aMPD (7.9%) and IMF (5.7%).

<table>
<thead>
<tr>
<th>Disease classification</th>
<th>No. screened</th>
<th>Total no. of patients with mutations</th>
<th>CBL ex8 mutations</th>
<th>CBL ex9 mutations</th>
<th>% with CBL mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV</td>
<td>74</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>IMF</td>
<td>53</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>5.7%</td>
</tr>
<tr>
<td>ET</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>SM</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>CMML/MDS</td>
<td>78</td>
<td>10</td>
<td>8</td>
<td>2</td>
<td>12.8%</td>
</tr>
<tr>
<td>aCML/aMPD</td>
<td>152</td>
<td>12</td>
<td>6</td>
<td>7</td>
<td>7.9%</td>
</tr>
<tr>
<td>HES/Eosinophilia</td>
<td>96</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>CNL</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>AML</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>547</strong></td>
<td><strong>26 patients</strong> (27 mutations)</td>
<td><strong>15</strong></td>
<td><strong>12</strong></td>
<td><strong>4.9%</strong></td>
</tr>
</tbody>
</table>

Table 5.10: Summary of CBL mutations according to disease subtype.

The mutations are summarised in Table 5.11. Most were missense changes, although I identified some candidate splice site mutations (n=4) and a heterozygous deletion (n=1). The majority of these sequence changes were novel however I did detect the R420Q mutation originally found in an AML patient (Sargin, et al 2007) in 4 of my 26 patients. This mutation was present in two IMF, one aCML and one unclassified aMPD cases and thus does not appear to be associated with a specific disease subtype. R420 is highly conserved in all CBL proteins and is the predicted contact
point of interaction between CBL and E2 ubiquitin-conjugating enzymes (Zheng, et al 2000) (Figure 5.20). This mutation was shown by Sargin et. al., (2007) to prevent FLT3 receptor internalisation and ubiquitination and is the first human CBL mutation causatively linked to the development of leukaemia. I also identified one patient with an amino acid alteration of the same codon, 420 that results in an arginine to leucine (R420L) substitution.

<table>
<thead>
<tr>
<th>Patient Identifier</th>
<th>Diagnosis</th>
<th>Exon 8 amplicon</th>
<th>Exon 9 amplicon</th>
</tr>
</thead>
<tbody>
<tr>
<td>E484</td>
<td>aCML</td>
<td>71930T&gt;T/C: 380L&gt;L/P</td>
<td>N</td>
</tr>
<tr>
<td>E632</td>
<td>aCML</td>
<td>N</td>
<td>72252C&gt;G:417P&gt;A</td>
</tr>
<tr>
<td>E1191</td>
<td>aCML</td>
<td>71930T&gt;T/C: 380L&gt;L/P</td>
<td>N</td>
</tr>
<tr>
<td>E001</td>
<td>aCML</td>
<td>71918C&gt;C:376S&gt;S/F</td>
<td>N</td>
</tr>
<tr>
<td>E088</td>
<td>CMML</td>
<td>71933G&gt;A:381C&gt;C/Y</td>
<td>N</td>
</tr>
<tr>
<td>E106</td>
<td>aMPD</td>
<td>71902T&gt;T/C:371Y&gt;Y/H</td>
<td>462R&gt;R/X</td>
</tr>
<tr>
<td>E110</td>
<td>CMML</td>
<td>HET DEL GGTAC (DEL Exon 8)</td>
<td>N</td>
</tr>
<tr>
<td>E435</td>
<td>aMPD</td>
<td>N</td>
<td>72252C&gt;G:417P&gt;A</td>
</tr>
<tr>
<td>NP277</td>
<td>aMPD</td>
<td>N</td>
<td>72262G&gt;G:412R&gt;R/L</td>
</tr>
<tr>
<td>04/1345</td>
<td>IMF</td>
<td>N</td>
<td>72262G&gt;G:412R&gt;R/Q</td>
</tr>
<tr>
<td>E879</td>
<td>IMF</td>
<td>71930T&gt;T/C:380L&gt;L/P</td>
<td>N</td>
</tr>
<tr>
<td>E2167</td>
<td>CMML</td>
<td>71977T&gt;T:396C&gt;C/G</td>
<td>N</td>
</tr>
<tr>
<td>E492</td>
<td>CEL</td>
<td>N</td>
<td>A&gt;AG -2</td>
</tr>
<tr>
<td>04/4357</td>
<td>IMF</td>
<td>N</td>
<td>72262G&gt;G:412R&gt;R/Q</td>
</tr>
<tr>
<td>E140</td>
<td>CMML</td>
<td>72022C&gt;CCT (+4)</td>
<td>N</td>
</tr>
<tr>
<td>E2015</td>
<td>CMML</td>
<td>71886G&gt;G:412C&gt;C/G</td>
<td>N</td>
</tr>
<tr>
<td>P42</td>
<td>aCML</td>
<td>72023G&gt;GA (+5)</td>
<td>N</td>
</tr>
<tr>
<td>E674.2</td>
<td>aCML</td>
<td>N</td>
<td>72262G&gt;G:412R&gt;R/Q</td>
</tr>
<tr>
<td>P607</td>
<td>aMPD</td>
<td>N</td>
<td>72262G&gt;G:412R&gt;R/Q</td>
</tr>
<tr>
<td>0607621</td>
<td>CMML</td>
<td>N</td>
<td>72253C&gt;C/T:417P&gt;P/L</td>
</tr>
<tr>
<td>1383</td>
<td>CMML</td>
<td>N</td>
<td>72363A&gt;A:454N&gt;N/D</td>
</tr>
<tr>
<td>1457</td>
<td>CMML</td>
<td>71941T&gt;T:384C&gt;C/R</td>
<td>N</td>
</tr>
<tr>
<td>1617</td>
<td>CMML</td>
<td>71983C&gt;CCT:398H&gt;H/Y</td>
<td>N</td>
</tr>
<tr>
<td>826</td>
<td>CMML</td>
<td>71942G&gt;G:384C&gt;C/R</td>
<td>N</td>
</tr>
<tr>
<td>03/2231</td>
<td>aCML</td>
<td>72015G&gt;C:408C</td>
<td>N</td>
</tr>
<tr>
<td>06/5197</td>
<td>aCML</td>
<td>N</td>
<td>72257C&gt;CAT:418F/L</td>
</tr>
</tbody>
</table>

Table 5.11: Summary of CBL mutations identified in this study.

Another oncogenic CBL mutation (C381A) in the same region of the protein has been reported in an animal model (Andoniou, et al 1994, Thien, et al 2001). Although I did not detect this mutation in any of my patients, I did discover a mutation that occurred in the same codon, C381Y. This residue is also highly conserved among the CBL
proteins and hence is likely to have a causative affect equivalent to constitutive receptor tyrosine kinase signalling (Figure 5.20).

![Schematic showing the CBL protein domain structure at the top and a 5 species alignment of the RING domain below (Homo sapiens, Pongo Pygmaeus, Bos Taurus, Mus musculus & Drosophila melanogaster). The stars in black at the bottom of each alignment indicate the homology between species. The asterisks at the top of each alignment indicate the site of mutations found in this study.](image)

**Figure 5.20:** Schematic showing the CBL protein domain structure at the top and a 5 species alignment of the RING domain below (Homo sapiens, Pongo Pygmaeus, Bos Taurus, Mus musculus & Drosophila melanogaster). The stars in black at the bottom of each alignment indicate the homology between species. The asterisks at the top of each alignment indicate the site of mutations found in this study.

All other missense mutations affect novel residues in the same region. In addition I identified 5 other sequence changes. The 5bp heterozygous deletion of GGTAC in patient E110 involved the last base pair of exon 8 and was similar to the deletion identified previously in the AML MOLM-13 cell line (Caligiuri, et al 2007) (Figure 5.21).
Figure 5.21: The 5bp deletion in patient E110. Top panel: Comparison of the 14bp deletion in the MOLM-13 cell line with the deletion in patient E110. Bottom panel: the sequencing electropherogram showing the 5bp deletion.

In addition four sequence changes in patients P42, E140, E492 and E2015 were identified in the non-coding regions of exons 8 and 9 (Figure 5.22). It is possible that these mutations represent new splice donor or acceptor sites thus altering the protein structure.
Figure 5.22: Schematic showing the locations and sequencing electropherograms of all four intronic CBL mutations. The trace highlighted in green (1) was discovered in intron 7 a G>C mutation at -1 of exon 8 in patient E2015. The traces highlighted in blue represent the three non-coding mutations identified in intron 8; (2) is the +4C>T substitution in patient E140, (3) shows the +5G>C mutation found in patient P42, and (4) depicts patient E492 and the -2A>G mutation.

The mutation scan that yielded these sequence changes focused exclusively on exons 8 and 9. Upon identification of a mutation the patient DNA was subjected to repeat PCR for the exon in question and resequenced to confirm the mutation was in fact present. Given the strong association of the mutations with an aCML/CMML/MF phenotype, all other CBL exons were amplified and screened for mutations in a total
of 70 of aCML/CMML/MF patients that were negative for exon 8 or 9 mutations. Although a number of known SNPs were identified, no other mutations were found.

5.4.3.6 11q LOH studies

I aimed to determine if the CBL mutations were usually associated with 11q aUPD. Apart from the three initial cases, the other mutation positive cases had not been analysed by SNP arrays. Since sequence analysis is poorly quantitative and arrays are expensive, I used microsatellite PCR analysis to look for chromosome 11q LOH in all 26 CBL mutation positive samples. In total I used 9 fluorescently labelled primer pairs targeting highly polymorphic microsatellite markers spanning chromosome 11q (Table 5.12).

<table>
<thead>
<tr>
<th>Chromosome band</th>
<th>Marker</th>
<th>Heterozygosity (<a href="http://www.gdb.org/">http://www.gdb.org/</a>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11q13.1</td>
<td>D11S1883</td>
<td>0.74</td>
</tr>
<tr>
<td>11q14.1</td>
<td>D11S937</td>
<td>0.88</td>
</tr>
<tr>
<td>11q21</td>
<td>D11S4182</td>
<td>0.73</td>
</tr>
<tr>
<td>11q23.2</td>
<td>D11S1885</td>
<td>0.71</td>
</tr>
<tr>
<td>11q23.3</td>
<td>D11S925</td>
<td>0.85</td>
</tr>
<tr>
<td>11q</td>
<td>D11S4107</td>
<td>0.78</td>
</tr>
<tr>
<td>11q24.1</td>
<td>D11S982</td>
<td>0.77</td>
</tr>
<tr>
<td>11q24.2</td>
<td>D11S934</td>
<td>0.85</td>
</tr>
<tr>
<td>11q25</td>
<td>CA rpt</td>
<td>unknown</td>
</tr>
</tbody>
</table>

Table 5.12: Microsatellite marker name, location on chromosome 11q and heterozygosity where known (www.gdb.org/).
The products were analysed on an ABI3100 using the Genotyper 2.0 program. As previously described for JAK2, if only one peak was visible a patient was scored as homozygous. To allow for the presence of background normal cells, homozygosity was also scored if one peak was one third or smaller than the expected size compared to controls (Jones, et al 2005). Of the 26 CBL mutation positive patients tested, 9 (including all three known 11q LOH patients (E484, E632 & E1191) displayed patterns that were suggestive of LOH (Table 5.13).

<table>
<thead>
<tr>
<th>Patient I.D</th>
<th>D11S1883</th>
<th>D11S937</th>
<th>D11S4182</th>
<th>D11S1885</th>
<th>D11S925</th>
<th>D11S4107</th>
<th>D11S982</th>
<th>D11S934</th>
<th>11q25 CA rpt</th>
<th>Tel</th>
</tr>
</thead>
<tbody>
<tr>
<td>E484</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E632</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1191</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E088</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E106</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E435</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP277</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>04/1345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E879</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2167</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E492</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>04/4357</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E674.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P607</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0607621</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1383</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1457</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1617</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>826</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>03/2231</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>06/5197</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.13:** 11q LOH in CBL mutation positive patients. The microsatellite markers are shown in order from the most centromeric (D11S1883) to the most telomeric (11q25 CA repeat) and are scored as homozygous (black) or heterozygous (white). The marker highlighted in yellow is the first that is distal to CBL.
Of these nine, six showed complete homozygosity at all microsatellite loci tested (the three 11q LOH patients detected by arrays plus E001, E2167 and 1457). The others (E879, 1617 and 03/2231) showed varying degrees of homozygosity however all ran from a centromeric to telomeric marker, consistent with mitotic recombination. A further three patients (04/1345, P42 & E140) also had regions of homozygosity over a region of 4 or 5 microsatellite markers however they did not extend to a telomere. Unfortunately constitutional DNA was not available from any case to determine if these regions were acquired or not, however the mean heterozygosity of the markers employed was 0.79 and thus the chance of four sequential homozygous calls randomly is \((0.21)^4 = 0.0019\), assuming that all are independent of each other.

Patient E106 was heterozygous for all markers. This is the only patient for whom I identified two \(CBL\) mutations, one in exon 8 the other in exon 9. The exon 8 mutation is missense and predicted to result in a unique Y371H amino acid substitution. The exon 9 mutation is also unique to this patient, R462X, and is also the only nonsense mutation identified. I carried out a PCR to determine if these mutations were on the same or different chromosomes as this may tell us more about how they are acting. I designed two intronic primers that would anneal at 66°C, and amplified both exons 8 and 9 of the \(CBL\) gene (see Appendix I, section XXVII) in a single amplicon of 786bp. This was then sub-cloned and several colonies sequenced. I was able to show that the mutations were never in the same clone indicating that the mutations were on different chromosomes. One possibility is that the truncating mutation serves to reduce the Y371H mutation to functional homozygosity.

5.4.3.7  Copy number analysis

I next went on to determine whether the \(CBL\) locus was present in two copies, which would implicate aUPD as the cause of the homozygosity, or one copy which would implicate a deletion. Multiplex ligation-dependent probe amplification (MLPA) is a robust technique used to look at copy number differences and a number of different kits have been developed by MRC Holland (www.mlpa.com) for genes of diagnostic interest. In particular, the Marfan probeset 1 (P065) kit contains 4 control probes at
11q including one at 11q23.3 the exact cytogenetic location of CBL. I therefore used this kit to examine CBL copy number and the results are summarised below in Table 5.14.

<table>
<thead>
<tr>
<th>Patient I.D</th>
<th>MLPA results</th>
</tr>
</thead>
<tbody>
<tr>
<td>E484</td>
<td>N</td>
</tr>
<tr>
<td>E632</td>
<td>N</td>
</tr>
<tr>
<td>E1191</td>
<td>N</td>
</tr>
<tr>
<td>E001</td>
<td>N</td>
</tr>
<tr>
<td>E088</td>
<td>N</td>
</tr>
<tr>
<td>E106</td>
<td>N</td>
</tr>
<tr>
<td>E110</td>
<td>no material</td>
</tr>
<tr>
<td>E435</td>
<td>no material</td>
</tr>
<tr>
<td>NP277</td>
<td>N</td>
</tr>
<tr>
<td>04/1345</td>
<td>N</td>
</tr>
<tr>
<td>E879</td>
<td>N</td>
</tr>
<tr>
<td>E2167</td>
<td>N</td>
</tr>
<tr>
<td>E492</td>
<td>N</td>
</tr>
<tr>
<td>04/4357</td>
<td>no material</td>
</tr>
<tr>
<td>E140</td>
<td>no material</td>
</tr>
<tr>
<td>E2015</td>
<td>N</td>
</tr>
<tr>
<td>P42</td>
<td>N</td>
</tr>
<tr>
<td>E674.2</td>
<td>N</td>
</tr>
<tr>
<td>P607</td>
<td>N</td>
</tr>
<tr>
<td>0607621</td>
<td>no material</td>
</tr>
<tr>
<td>1383</td>
<td>no material</td>
</tr>
<tr>
<td>1457</td>
<td>N</td>
</tr>
<tr>
<td>1617</td>
<td>N</td>
</tr>
<tr>
<td>826</td>
<td>no material</td>
</tr>
<tr>
<td>03/2231</td>
<td>N</td>
</tr>
<tr>
<td>06/5197</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 5.14: Patient identifier versus MLPA copy number result. N= normal (no copy number alteration) indicating the patient has two copies of chromosome 11q.

Of the 26 patients with CBL mutations, only 19 were analysed by MLPA due to the paucity of suitable material for some cases (MLPA is very sensitive to DNA quality). The three original 11q LOH patients identified by SNP array analysis were included as controls and all three were confirmed by MLPA to have two 11q chromosomes. Of the remaining 16 patients, all apparently had two intact chromosomes 11q including the six cases for whom I found LOH by microsatellite analysis. This
implies the LOH is likely to be due to mitotic recombination following an initial \textit{CBL} mutation.

5.4.3.8 \textit{Are the CBL mutations acquired?}

It is important to determine if the \textit{CBL} mutations were acquired or constitutional. An acquired mutation implies the genetic change occurred in a somatic cell and provided it with a selective advantage and thus giving rise to a clone. Since this required obtaining fresh samples, it was only possible to perform the analysis on three cases.

A fresh blood sample was received from E1191 who was positive for the L380P mutation. I set up a phytohaemagglutinin (PHA) culture to enrich for T-cells and compared DNA prepared from a granulocytic fraction to DNA from a T-cell fraction. Previously studies in our group have shown that growth in PHA typically results in >80\% T-cells after 5 days. As shown in Figure 5.23, the mutation is present in granulocytes but absent from T-cells.
Figure 5.23: Electropherogram for patient E1191 showing a) CBL exon 8 reference sequence; b) T-cell enriched DNA; c) comparison of a and b showing no differences in sequence; d) comparison between reference sequence and granulocyte DNA with the peak indicating a sequence difference, and e) granulocytic DNA sequence showing the T>C mutation.

Patient E879 also had the L380P mutation and in this case we were able to obtain a mouthbrush sample. Sequencing showed again that the mutation was only present in the PB DNA sample and not in the epithelial (mouthbrush) fraction (Figure 5.24).
Finally, patient 04/1345 with a R420Q mutation was analysed. The sample I had originally analysed from this case was taken in 2004 when MF was diagnosed, however he had originally presented with ET in 1989. I was able to obtain and extract DNA from bone marrow slides taken in 1989, 1992, 1995 and 1997 and found no evidence of the CBL mutation. Unfortunately no slides were available between 1997 and 2004, however these findings suggest that the mutation may have appeared on transformation from ET to MF.
In conclusion, I have been able to determine that the CBL mutation was acquired in three of three cases analysed. It is likely that the other 23 patients also acquired CBL mutations however this remains a matter of conjecture.

5.4.3.9 Analysis of non-coding CBL mutations

As mentioned above, intronic sequence changes were identified in four cases and a 5bp deletion was found in a further patient. To determine if these variants affected splicing, I performed RT-PCR on the three cases for which material was available: E140 and E2015 with intronic changes and E110 with the heterozygous 5bp deletion.

Primers were designed over two exon boundaries to ensure they would only amplify the coding sequence and not any residual genomic DNA if present in the cDNA sample. The primers were designed to anneal at 66°C and cover exons 6/7 and exons 10/11, thus ensuring the region comprising exons 7-10 would be amplified (for primer sequences see Appendix I, section XXV). A normal control was also run alongside and the resulting amplifications are shown in Figure 5.25.
Figure 5. 25: Agarose gel picture. Lane 1 represents E140; Lane 2, E2015; Lane 3, normal control, and Lane 4, no DNA control run alongside the 1kb Plus ladder (Invitrogen) for sizing. The normal control (Lane 3) is the expected size (~585bp) and when compared to E140 it is clear this patient also only has this normal band suggesting no splicing differences are created with the presence of this mutation. The other patient, E2015 exhibits two bands a normal 585bp band and a smaller band that upon cloning and sequencing revealed a deletion of exon 8.

Patient E140 showed a single band that corresponds to the size of the normal control. Upon sequencing I confirmed exons 7-10 were fully intact thus it appears the variant at 72022C>C/T(+4) does not affect the structure of the protein. Although I did not have material to prove this, it seems likely that the 72023G>G/A(+5) variant seen in patient P42 is also likely to be of no consequence. To test this possibility, I put both these altered sequences into a splice site predictor programme (www.fruitfly.org/seq_tools/splice.html) which predicts if an unknown sequence change might create a better or different splice acceptor or donor site compared to the reference sequence (Reese MG 1997). Neither of the mutations showed improved splice donor or acceptor sites supporting the idea that they are non-functional and are presumably rare polymorphisms or pathogenetically unimportant passenger mutations.
Patients E110 (not shown) and E2015 both exhibited double bands by RT-PCR. Cloning and sequencing showed that the larger band corresponded to the normal product and the smaller band resulted from the complete deletion of exon 8 in both cases. It therefore appears that both patients harbour heterozygous mutations with one normal allele and one allele with a deletion of exon 8. The mutation in MOLM-13 also causes skipping of exon 8 and siRNA targeting the exon 7-9 mutant transcript in this cell line was found to be growth inhibitory strongly suggesting that it is indeed of functional importance (Caligiuri, et al 2007).

5.4.3.10  The relationship between CBL and other mutations in MPDs

It is possible that the CBL mutants provide a growth advantage by interfering with the normal negative regulation of tyrosine kinase signalling, however it is unclear if they might act as sole abnormalities or if they co-operate with other acquired genetic events. All the cases I studied were negative for known tyrosine kinase fusion genes but they had not been tested for other oncogenic mutations seen in MPDs such as the FLT3 ITD, JAK2 V617F and NRAS mutations. These mutations have been reported in approximately 5%, 10-15% and 10% of aMPD cases, respectively (Jones, et al 2005).

The FLT3 ITD analysis was performed as described previously. The JAK2 V617F mutation analysis was carried out by a colleague using a specific ARMS test (Jones, et al 2005). For NRAS I designed primers to exons 2 and 3 which includes the two mutation hotspots at codons 12, 13 and codon 61 (for primer sequences see Appendix I, section XXVI). All CBL mutation positive patients were analysed and the results are shown in Table 5.15.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>JAK2 status</th>
<th>NRAS status</th>
<th>FLT3 ITD status</th>
</tr>
</thead>
<tbody>
<tr>
<td>E484</td>
<td>Normal</td>
<td>N</td>
<td>Normal</td>
</tr>
<tr>
<td>E632</td>
<td>Normal</td>
<td>N</td>
<td>Normal</td>
</tr>
<tr>
<td>E1191</td>
<td>Normal</td>
<td>N</td>
<td>Normal</td>
</tr>
<tr>
<td>E001</td>
<td>Normal</td>
<td>N</td>
<td>Normal</td>
</tr>
<tr>
<td>E088</td>
<td>Normal</td>
<td>N</td>
<td>Normal</td>
</tr>
</tbody>
</table>
Table 5. 15: Results of JAK2 V617F, NRAS & FLT3 ITD mutation analysis in CBL mutated cases. Het = heterozygous mutation.

Only one patient was found to harbour another mutation: E879 was heterozygous positive for JAK2 V617F. Since this finding was unusual, i.e. only seen in a single case, it raised the question whether both mutations arose in the same clone or not. To address this issue a fresh blood sample (taken approximately 2.5 years after the original sample) was obtained and set up in methylcellulose cultures in order to grow haematopoietic colonies for clonal analysis. However in contrast to typical V617F positive cases, the colony growth was very poor and, furthermore, analysis of the blood leucocytes showed the CBL mutation was present as expected but V617F JAK2 was no longer detectable. This raised the question of a technical error at some point in the analysis. To address this, I used the Promega Powerplex® 16 kit to determine if the two samples, plus also other samples we had from this case, were indeed from the same individual. This kit amplifies 16 highly polymorphic loci across the genome and was designed to allow a high discriminatory power for analysis of identity and paternity. I tested four different samples (E40, E879, E2763 and E2772) received at
various time points from this individual. All four samples concurred at all 16 loci, proving they were in fact all from the same individual. These samples and others that were received were analysed for V617F JAK2 and CBL as shown on Table 5.16.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Date</th>
<th>Sample type</th>
<th>JAK2 mutation</th>
<th>CBL mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1064</td>
<td>1998</td>
<td>BM Slide</td>
<td>YES (5% V617F)</td>
<td>NO</td>
</tr>
<tr>
<td>E40</td>
<td>2000</td>
<td>PB</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>E2922</td>
<td>2003</td>
<td>BM slide</td>
<td>not tested</td>
<td>NO</td>
</tr>
<tr>
<td>E879</td>
<td>2005</td>
<td>PB</td>
<td>YES (7% V617F)</td>
<td>YES</td>
</tr>
<tr>
<td>E2763</td>
<td>2008</td>
<td>BM</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>E2772</td>
<td>2008</td>
<td>PB</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>E2915</td>
<td>2008</td>
<td>mouthbrush</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>

Table 5.16: Time course of disease in patient E879 showing the disappearance of JAK2 V617F and the appearance of the CBL mutation over time.

The early samples received were all JAK2 V617F positive (albeit at relatively low levels) and CBL L380P negative. Between 2005 and 2008 this appears to have changed with the CBL L380P clone becoming more dominant and the JAK2 V617F clone being lost. The proportion of CBL mutation positive alleles as called by the mutation surveyor sequencing programme was 86% in 2005 and had increased to 100% by 2008. This transition coincided with an increase in leukocyte numbers that has proved difficult to control by standard therapy. This situation is reminiscent of the evolution of V617F positive PV to acute leukaemia in which it has been reported that roughly half of cases have leukaemia blasts that are V617F negative (Campbell, et al 2006, Theocharides, et al 2007). One possible explanation for this unexpected finding is that an unknown mutation results in a clinically covert clonal expansion prior to the acquisition of V617F. Another explanation is that there may be a hereditary predisposition to acquisition of independent mutations. In our case, the fact that V617F disappeared must presumably mean that the CBL mutation was acquired in a V617F negative clone and that it had a relative growth advantage.

One other case had two mutations: E2167 was originally diagnosed as systemic mastocytosis (SM) before evolving to CMML and had previously tested positive for
the *KIT* D816V mutation before I found a *CBL* mutation in the same sample. Unfortunately it was not possible to obtain any further samples from this case.

## 5.5 DISCUSSION

### 5.5.1 SNP arrays to identify new pathogenetic abnormalities

LOH is a genetic mechanism that creates a homozygous or hemizygous somatic cell by loss of the wild-type allele. This can be as a result of various mutational events such as a deletion, mitotic recombination, localised gene conversion, point mutation, or non-dysjunction. My aim has been to identify new oncogenes or tumour suppressor genes that are marked by these regions of LOH, and specifically regions of aUPD, or copy number neutral LOH formed as a consequence of mitotic recombination.

Regions of aUPD have been described in many human malignancies. Furthermore, these regions are now being commonly associated with oncogenic mutations, the first example of which was the discovery of 9p aUPD and the *JAK2* V617F mutation (Kralovics, et al 2002, Kralovics, et al 2005b). Since these initial reports, other oncogenic mutations have been associated with aUPD involving distinct chromosomal regions, such as *CEBPA* and chromosome 19 (Raghavan, et al 2005), *FLT3* mutations and 13q (Griffiths, et al 2005), *NF1* and 17q (Flotho, et al 2007), *WT1* mutations and 11p and finally *RUNX1* and 21q (Fitzgibbon, et al 2005). Moreover, reports where paired samples have been studied often show that the mutation precedes the aUPD and hence the aUPD represents a ‘second hit’ associated with disease progression (Fitzgibbon, et al 2007, Fitzgibbon, et al 2005). These studies all highlight the potential for discovering new pathogenetic abnormalities by the finding of regions of aUPD. Obviously this approach has limitations, e.g. it will not detect cryptic fusion genes or point mutations that do not have a selective advantage when homozygous. However I was particularly encouraged by the
observations that activating mutations of two tyrosine kinases are subject to aUPD and that these are the gene class that is most commonly mutated in MPDs.

There are various techniques that can aid identification of these regions of LOH however SNP arrays have the advantage of being able to detect copy number in addition to zygosity. In my study, I employed 50K SNP arrays and focused almost exclusively on regions of aUPD however it was clear that large copy number changes were also detected that correlated with cytogenetic data. Although I did not investigate this aspect in any detail, others have found that copy number information generated by 50K arrays is relatively poor but is greatly improved by using arrays of a higher density, in particular the Affymetrix 500K arrays (Mullighan, et al 2007) and more recently SNP 6.0 arrays which include 1.8 million markers. However for the purpose of identifying aUPD the lower density arrays are sufficient.

In total 88 samples were sent to the RZPD German Resource Centre for Affymetrix 50K SNP array analysis, including 30 aMPDs (predominantly aCML), 28 IMF, 10 PV, 10 myeloid & 10 lymphoid CML blast crisis patients. Results were successfully obtained for all samples. The principal weakness of my study was the absence of constitutional DNA, e.g. from mouthbrush swabs or remission DNA and so it was not possible to determine if tracts of homozygosity were indeed acquired or whether they were inherited. Consequently I did not attempt to analyse the smaller regions of homozygosity in any detail, although I did determine that there were no obvious recurrent abnormalities that were restricted to particular patient subgroups and also no recurrent regions that might target tyrosine kinases.

I was fortunate that my principal study group – atypical MPDs – were particularly informative with 40% of cases showing large regions of copy number neutral homozygosity. Because such large regions are not seen in normal individuals it is reasonable to infer that they are most likely acquired, although I was only able to prove this in three cases. I found eight different chromosomal regions affected by aUPD of which two were recurrent (7q and 11q; 3 cases each) and not seen in other
patient subgroups. These regions are discussed in detail below. Acquired UPD was also seen at 1p, 8p, whole chromosome 13, 17q, 20q and 21q. The 1p abnormality was linked to an oncogenic \( MPL \) 515 mutation and the patient with whole chromosome 13 aUPD was homozygous for a \( FLT3 \) ITD. It is likely that these two oncogenic changes are the targets of the aUPD, i.e. homozygosity for these mutations confers a growth advantage over the heterozygotes and so following random mitotic recombination the homozygous clone has a selective advantage. However I cannot exclude the possibility that there are other relevant abnormalities in the affected regions that are being selected for.

In the other patient subgroups, aUPD was uncommon. I was particularly hopeful that SNP analysis might be helpful in understanding disease progression in CML. CML-BC may be myeloid or lymphoid in phenotype and the molecular basis for transformation is poorly defined (Calabretta and Perrotti 2004). It is likely, but not certain, that disease progression involves several genetic steps and that different combinations of a limited number of genes are involved in different patients. The genes involved in myeloid transformation may be distinct from those in lymphoid transformation, but cytogenetic and molecular evidence (Chase, et al 2001, Johansson, et al 2002) suggest that there is an appreciable overlap in the molecular basis of myeloid blast crisis and AML and similarly in the molecular basis of lymphoid blast crisis and ALL. aUPD has been reported in both AML and ALL (Fitzgibbon, et al 2005, Irving, et al 2005, Raghavan, et al 2005) and therefore it was a disappointment that all BC cases were negative, apart from a possible small region at the tip of 15p in one case with myeloid disease. As mentioned above, the copy number resolution of the 50K arrays was insufficient to detect small deletions such as those recently described to target Ikaros in lymphoid BC and Ph+ALL (Mullighan, et al 2007).

Similarly, patients with PV and MF were largely uninformative, although one case with V617F positive MF had extensive aUPD on 5q. The significance of this finding is unclear, although it suggests that there may be a gene on 5q that co-operates with
V617F to induce progression to MF. One possibility is the recently identified 5q-MDS gene RPS14 (Ebert, et al 2008), however this is perhaps unlikely given that RPS14 appears to exert its effects through heterozygous deletion rather than homozygous mutations. Instead of analysing this single abnormality I focused on the recurrent regions in aCML/aMPD patients.

5.5.2 7q aUPD


The region of 7q aUPD in each patient varied, however the minimal overlapping region spanned approximately 53Kb covering cytogenetic bands 7q22.3-qter. This is a colossal amount of sequence and is thought to comprise some 400 different genes. Much work has already been performed on cases in the literature trying to define critical regions of deletion, yet no group has yet identified any definitive tumor suppressor gene(s) that is thought to be located within this region. However I was more concerned with identifying activating mutations, based on the strong linkage between aberrant tyrosine kinase signalling and MPDs. The most obvious candidate genes were the three tyrosine kinases EPHA1, EPHB6 and MET but no mutations were identified. Partial screens of BRAF and PI3KCG were also negative. These genes were selected because they participate in tyrosine kinase signalling pathways and are either known to be mutated in cancer (BRAF) or highly related to a gene that is mutated in cancer (PI3KCG is related to PI3KCA). Since these are such good
candidates it might be worth screening the complete sequence of all exons in case the previously described hotspots are limited to epithelial cell tumours. A similar situation has been described previously: the activating D816V KIT mutation is seen in >80% of cases of systemic mastocytosis but in gastrointestinal stromal tumours, D816V is very uncommon and instead activating mutations are seen elsewhere in the same gene. I looked for other genes on 7q but it is easy to find large numbers of candidates, none of which were obviously outstanding.

Subsequent to my work, a number of papers have been published by Gondek and colleagues (Gondek, et al 2007a, Gondek, et al 2007b, Gondek, et al 2008). This group’s main focus is on MDS and MDS/MPD using 250K SNP arrays. Copy number neutral LOH was seen in 30% of cases, particularly 1p, 1q, 4q, 7q and 11q. Figure 5.26 shows the regions of 7q aUPD in my cases compared to the five identified by Gondek and colleagues, three of which had CMML, one AML and one refractory anaemia (RA, a subtype of MDS).
Figure 5.26: Schematic detailing chromosome 7 and the regions of aUPD identified by Gondek & colleagues (2008), highlighted in green and my aCML/aMPD patients highlighted in red.

The smallest area of overlapping aUPD is now represented by the Gondek patient with RA, but the minimal region is only marginally smaller than that defined by E639. It will be important to analyse more cases to try and narrow the region down and it might be expected that with enough cases the proximal end of the minimal region might converge to a point close to the key gene (Figure 5.27). However this may well require analysis of a very large number of cases.
Figure 5.27:  Schematic detailing a theoretical analysis of chromosome 7 and how identification of more regions of aUPD might narrow down the region and converge at the proximal end (indicated by the red dashed line) near the gene of interest.

Whether the underlying abnormality is the same as the elusive 7q candidate gene believed to be at the root of recurrent deletions in AML is unclear, but since aUPD for 7q is uncommon in AML (Fitzgibbon, et al 2005, Gondek, et al 2008, Raghavan, et al 2005) it may well be that the two abnormalities are different. Over the years much work has been performed in myeloid disorders with 7q deletions and translocations trying to narrow down a minimal region in the hope of highlighting a specific mutated oncogenic or tumour suppressor gene. Some people have suggested that the complexity of 7q rearrangements might simply indicate that this is a region prone to breakage and instability, and thus a synergy of different genetic factors may play a general role in oncogenesis rather than a gene harbouring a specific mutation. However it seems more likely that there is a specific target for the cases affected by
aUPD. Interestingly, one of the 7qUPD cases (E110) was found to have a heterozygous deletion of the \textit{CBL} gene. Assuming that this \textit{CBL} abnormality is functionally important, this might indicate that the 7q gene functionally co-operates with \textit{CBL} to produce an aCML phenotype.

Although the analysis of further cases by arrays will help to define a smaller minimal region, identification of the critical target will probably require the sequence analysis of a large number of genes to identify mutations that are homozygous in cases with 7q aUPD and absent or heterozygous in cases without 7q aUPD. Although this could be achieved by the stepwise of candidate genes such as those listed in Table 5.8, very soon it should be possible to systematically sequence all genes in the minimal region using a combination of pull-down hybridisation arrays covering all 7q exons in the critical region and next generation sequencing, which currently has the capacity to sequence approximately $10^8 - 10^9$ bases per run depending on the platform employed (Chan 2005, Mardis 2008, Morozova and Marra 2008, Schuster 2008, von Bubnoff 2008, Wheeler, \textit{et al} 2008).

\subsection*{5.5.3 11q aUPD}

Chromosome 11 has also been reported in the literature as a region commonly affected by imbalances in human malignancy. Reports of LOH are far more frequent than deletions, and have been associated with cervical, ovarian, breast, colon and skin cancers (Bethwaite, \textit{et al} 1995, Carter, \textit{et al} 1994, Foulkes, \textit{et al} 1993, Hampton, \textit{et al} 1994, Keldysh, \textit{et al} 1993, Tomlinson, \textit{et al} 1993). Nevertheless, deletions and translocations of 11q appear to be more common than LOH in haematological malignancies. LOH has rarely been reported to be associated with leukaemia and so the identification of a recurrent region of 11q aUPD in aMPD patients is particularly interesting and suggests an abnormality specifically associated with an MDS/MPD phenotype.
The smallest overlapping region of aUPD was represented by patient E484 and spans 57Mb covering cytogenetic bands 11q13.5-qter. This covers almost three quarters of the long arm of chromosome 11 and is thought to comprise greater than 350 genes. I had anticipated that the identification of regions of aUPD might provide a pointer towards novel tyrosine kinase mutations, however there are no tyrosine kinase genes on 11q and the three at 7q were eliminated by sequence analysis. A number of candidate genes were then short-listed according to association with leukaemia, and screening for the MLL partial tandem duplication (PTD) was carried out as a priority, but this was found to be negative in all three cases with 11q aUPD. Because of the strong association of MPDs and deregulated tyrosine kinase signalling, the next step was to look at genes that are known to impact on this process.

The CBL gene is located at 11q23.3 and encodes an ubiquitin ligase which functions as a negative regulator of several receptor protein tyrosine kinase signalling pathways. It acts by inducing ubiquitination of the RTK, thus promoting their sorting for endosomal degradation or recycling, however CBL also acts in a positive way by serving as a scaffold for the recruitment of downstream signalling substrates. CBL was an obvious candidate for mutation screening, particularly as two papers were published that reported the discovery of rare CBL mutations in AML patients (Caligiuri, et al 2007, Sargin, et al 2007). Based on the clustering of mutations in these cases I screened the three 11q aUPD cases and found that two had changes in exon 8 predictive of an L380P missense mutation whilst the third had a mutation of exon 9, P417A.

To determine if these mutations were more widespread among the MPDs I screened 547 patients, including cases with PV, IMF, ET, SM, HES, CMML and aMPD. I identified a further 23 patients (total 26 including 11q aUPD patients) with CBL mutations, 15 of that were missense and one nonsense as illustrated on Figure 5.28. One patient had a mutation in both exons 8 and 9 so in total there were 15 exon 8 mutations and 12 exon 9 mutations. The mutations were strongly associated with CMML (12.8%) and aCML/aMPD (7.9%) and in 3 cases I was able to demonstrate that they were acquired. Further analysis is being taken by other members of the
group to try and determine the specific clinical features associated with *CBL* mutations and their impact on prognosis.

I found four cases with R420Q, the same mutation described by Sargin et al., (2007) but the majority of mutations was novel. Codon 420 is highly conserved in all CBL proteins and is the predicted contact point of interaction between CBL and the E2 ubiquitin-conjugating enzyme (Zheng, et al 2000). Functional work by Sargin et al. showed this mutant co-operates with overexpression of FLT3 to induce growth factor independence of 32D cells. R420Q prevents FLT3 receptor internalisation and ubiquitination and is the first description of a *CBL* mutation that is very likely to be a causative event in the development of human leukaemia. There have also been previous reports of animal models with oncogenic mutations in the RING finger.

Figure 5.28: Location of CBL missense mutations. Top panel: schematic showing the protein structure of CBL detailing the tyrosine kinase binding (TKB), RING, proline-rich (P-rich) and ubiquitin-associated/leucine zipper (UBA/LZ) domains. Bottom panel: ClustalW alignment of the entire linker and RING domains plus part of the distal sequence. The ring domain is shaded in grey. Residues affected by missense mutations identified by this study to date are indicated by arrowheads above the clustalW alignment. Those underneath the alignment are previously reported mutations.
domain or linker sequence of CBL, at codon 381 (C381A) (Andoniou, et al 1994, Thien, et al 2001). I also discovered a mutation of the same codon (C381Y).

It is thought that alteration of signals for sorting tyrosine kinases to the lysosomes may contribute to the progression of neoplasia or even be an initiating event (Bache, et al 2004, Peschard, et al 2004). For example, several oncogenic variants of EGFR, MET, CSF1R and KIT have been shown to escape lysosomal degradation by a number of methods such as loss of PTB-mediated CBL binding sites (Peschard and Park 2003), or through the formation of oncoproteins caused by chromosomal translocations. The TPR-MET (Park, et al 1986) fusion gene, found originally in a mutagenised human osteogenic sarcoma cell line, results in deletion of sequence encoding the MET juxtamembrane domain that includes a binding domain site for CBL. As a consequence, the TPR-MET fusion escapes ubiquitination (Peschard, et al 2001) suggesting that its inability to be targeted to the lysosomes for degradation may contribute to its pathogenicity. Analysis performed by Sargin & colleagues proved that R420Q prevented ubiquitination and internalisation of not just FLT3 but also PDGFR and EGFR, suggesting that its oncogenicity might be mediated by failure to properly regulate many different receptors.

The critical analysis that is missing from my study is the functional determination of the consequences of the missense mutations plus the abnormalities leading to skipping exon 8. It was my intention to undertake this analysis but there was insufficient time and this is now being undertaken by another member of the group. The work is being conducted in collaboration with Sargin et al. who have provided the key reagents: pMY based plasmids expressing wild type and R420Q CBL (as a positive control) plus parental 32D cells and 32Ds expressing wild type FLT3. 32D cells are murine myeloid precursor cells that normally require IL-3 for growth but 32D/FLT3 can be propagated in the presence of FLT3 ligand.

Preliminary studies have included transfection of the WT CBL and R420Q expression vectors into 32D and 32D/FLT3 and show they are transforming in our
hands, as well as producing stable clones. To do this, a retroviral transfection protocol was used. The pMY vectors were developed from pMX (originally based on MuLV), which utilise pCMVs long terminal repeat (LTR) required for replication, integration, and expression of viral RNA. They also express green fluorescent protein (GFP) and thus transfected cells can be selected by fluorescent-activated cell sorting (FACS). The selected cells are expanded and proliferation assays (MTS assay and manual cell counting after trypan blue staining) performed after starving the cells of FLT3 ligand overnight. Sargin et al. showed that the R420Q mutant expressed in 32D/FLT3 continued to proliferate under these conditions whereas all other combinations failed to proliferate.

Sargin et al., also demonstrated that R420Q resulted in prolongation of signalling. This will be addressed by starving transfected cells of cytokines and then stimulating them briefly with FLT3 ligand. Samples will be taken over a time course and Western blotting used to look at levels of phospo-ERK compared to total ERK. Other substrates will also be examined; in particular STAT5 which is known to be an important tyrosine kinase substrate in haematopoietic cells (Kato, et al 2005, Spiekermann, et al 2002).

Sargin et al. also demonstrated that the R420Q was deficient in ubiquitination activity and thus the mutant has lost its ability to negatively regulate tyrosine kinases by inducing receptor turnover but has presumably retained its ability to act as a scaffold for the recruitment of other signalling proteins. Loss of ubiquitin ligase activity has been confirmed for some of the mutants by transiently transfecting them into 293T cells along with FLT3 and HA-tagged ubiquitin. After 48 hours, FLT3 ligand was used to stimulate the receptor which was then immunoprecipitated and blotted with anti-HA. The strength of the signal indicated the degree of ubiquitination which was found to be greatly reduced for most CBL mutants (Grand FH 2009).

Another aspect that needs further investigation is functional consequences of homozygosity of CBL mutations. Both functionally important mutations that had
been identified previously (R420Q and C381A) were believed to act in a dominant negative fashion, however the finding of aUPD strongly suggests that there is a further selective advantage by reduction to homozygosity and that therefore the mutations are co-dominant. Following the array analysis, I tested CBL mutation positive cases by a combination of microsatellite PCR and MLPA and found a further 6 of the 26 patients exhibited 11q aUPD. Of the 9 patients in total (includes the 3 aMPD patients diagnosed by SNP array) 5 had aCML/aMPD, 3 CMML and 1 IMF (Table 5.17).

<table>
<thead>
<tr>
<th>Patient Identifier</th>
<th>CBL mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>E484</td>
<td>L380P</td>
</tr>
<tr>
<td>E632</td>
<td>P417A</td>
</tr>
<tr>
<td>E1191</td>
<td>L380P</td>
</tr>
<tr>
<td>E001</td>
<td>S376F</td>
</tr>
<tr>
<td>E879</td>
<td>L380P</td>
</tr>
<tr>
<td>E2167</td>
<td>C396G</td>
</tr>
<tr>
<td>1457</td>
<td>C384R</td>
</tr>
<tr>
<td>1617</td>
<td>H398Y</td>
</tr>
<tr>
<td>03/2231</td>
<td>W408C</td>
</tr>
</tbody>
</table>

Table 5.17: Nine patients with 11q aUPD and a CBL mutation.

The majority of mutations associated with 11q aUPD are found in exon 8 of CBL gene with the exception of E632 (P417A). It also appears that the L380P mutation is significantly associated with homozygosity as all three patients identified with this
mutation also exhibited 11q aUPD. This is not true of the other cases where the same mutation or codon has been altered, such as codon 417 and 384, for which some cases were homozygous and others were heterozygous. Although it is conceivable that another, unknown locus on 11q is the target of aUPD, the finding of one case with a Y371H mutation and a stop codon on opposite alleles strongly suggests that homozygous CBL mutations do indeed confer a direct selective advantage.

There may be an analogy between the CBL mutations and JAK2 V617F, which is also subject to reduction to homozygosity by aUPD in a similar proportion of cases. Homozygosity for V617F is relatively common in PV and MF but rare in ET however many patients with all three disease types are heterozygous. When large numbers of cases are analysed it appears that homozygosity is associated with a slightly more aggressive disease and it would be interesting to determine if cases with homozygous CBL mutants were also more aggressive. Interestingly, most ‘heterozygous’ PV patients have a proportion of haematopoietic colonies that are homozygous for V617F. This is not seen in ET and thus strengthens the association between V617F zygosity and disease phenotype (Scott, et al 2005). It will be interesting to test haematopoietic colonies from CBL mutated cases to see if they are homozygous. This might indicate that the observed heterozygosity was only apparent and in fact caused by a mixture of homozygous mutated cells with normal cells that are not part of the malignant clone.

Finally there is the question of whether CBL mutations are sufficient to give rise to disease by themselves, or if they require other co-operating genetic events. The requirement for over-expression of a tyrosine kinase for transformation of 32D cells may simply reflect the paucity of receptor tyrosine kinase gene expression in this cell line or, alternatively, may indicate that the mutants are only weakly transforming and require the co-operation of other events to give rise to clinically manifest disease. Consistent with the latter hypothesis, I identified two cases in which CBL mutations were acquired following progression of a pre-existing MPD. One of these presented with a relatively low level V617F JAK2 mutation that disappeared on progression to
CBL L380P positive disease, indicating that the two mutations must have arisen independently in different clones. It remains to be established, however, if CBL mutations are usually secondary events or whether the two cases we identified are exceptional.

Further evidence for the requirement for additional co-operating genetic events comes from published mouse studies in which neither complete CBL knockout nor heterozygous knock-in of the RING finger inactivating mutant C379A (mouse equivalent of C381) resulted in evidence of an MPD or other malignancy (Rathinam, et al 2008, Thien, et al 2005). However closer analysis revealed subtle haemopoietic perturbations that clearly relate to the pathogenesis of MPDs. Specifically, haemopoietic stem cells (HSC) from these animals were increased in number, hyper-responsive to thrombopoietin and more potent than wild type HSC in repopulating the haematopoietic system. These effects were associated with increased STAT5 accumulation and phosphorylation. Hyper-responsiveness to growth factors and stem cell involvement are two of the cardinal features of MPDs. Furthermore, the association between STAT5 activation and MPDs is well documented (Cain, et al 2007, Kotecha, et al 2008). Whilst it is possible that overt haematological disease might have been induced by a homozygous knock-in mutant, these findings suggest that CBL mutations may not be solely responsible for a full MPD phenotype. Possibly this could be tested further by transfecting bone marrow cells from CBL knockout mice, which are viable and fertile (Murphy, et al 1998). Nevertheless, since the CBL mutants appear to be acting by enhancing tyrosine kinase signalling, it may prove possible to treat patients in the future with tyrosine kinase inhibitors.
6 DISCUSSION

The unifying theme of my research is the development and use of different approaches to identify deregulated tyrosine kinases and associated signalling components in MPDs. The paradigm of BCR-ABL in CML paved the way for the identification of other deregulated tyrosine kinase genes in the MPDs, although it took several years for the association between MPDs and deregulated signalling to emerge. Initially I focused on the analysis of chromosome translocations, subsequently moving onto screens for over-expression of selected tyrosine kinase genes and finally to genome-wide approaches that took into account the fact that genes other than tyrosine kinase may be responsible for the substantial proportion of MPDs of unknown pathogenesis.

6.1 Novel fusion genes and roles of partner proteins

In this thesis I have described the identification of six novel fusion genes. Four of these were completely novel with partner genes identified by me (GOLGA4, STRN and CPSF6) or by another member of the group (KIF5B) that have not been previously described as being involved in leukaemia. All four are very rare and as yet only GOLGA4-PDGFRB has been found in more than one patient. The fifth resulted from fusion of a known partner protein, ETV6, to PDGFRα. The final fusion was a variant ETV6-PDGFRB fusion with rare breakpoints never reported before.

Five of the six fusions involved PDGFR genes and are therefore diagnostically important because of the known sensitivity of both PDGFRα and PDGFRβ to targeted therapy with the tyrosine kinase inhibitor imatinib. Indeed all five received imatinib, of who four achieved complete cytogenetic remission, and for the fifth case the follow up is still too short to assess remission status. The remaining fusion involved the FGFR1 gene, known to be involved in a specific MPD referred to as
8p11 myeloproliferative syndrome (EMS) that is refractory to imatinib but may be sensitive to inhibitors that are active against FGFR1 (Chase, et al 2007) although none are currently available for clinical use.

Fusions of PDGFRα, PDGFRβ and FGFR1 are all known to be associated with MPDs. Translocations involving PDGFRβ appear the most diverse with reports of partner genes at more than 20 chromosomal locations (Figure 6.1). Only one, ETV6-PDGFRβ, has been found relatively frequently with the remainder being very rare and many only described in one or two patients. FGFR1 in EMS has been reported to fuse to 8 partners in addition to the one I found. The three PDGFRα fusions I found take the current total to six, by far the most common of which is FIP1L1-PDGFRα.

![Spider diagram illustrating translocations associated with MPDs and related disorders. The tyrosine kinase genes are indicated in white and the fusion partner genes in green.](image)

**Figure 6.1:** Spider diagram illustrating translocations associated with MPDs and related disorders. The tyrosine kinase genes are indicated in white and the fusion partner genes in green.
I took a number of different approaches to identify the fusion genes:

- Both *GOLGA4-PDGFRB* and *CPSF6-FGFR1* were identified following initial FISH analysis to confirm the involvement of the respective tyrosine kinase genes, both of which were suspected from the cytogenetic results and clinical phenotype. Once this was confirmed I performed 5’RACE PCR from the kinase to identify the partner gene.

- *STRN-PDGFR* and *ETV6-PDGFR* were cloned by bubble PCR from genomic DNA, exploiting the fact that the genomic breakpoints for all reported *PDGFR* fusions are very tightly clustered. This approach was particularly useful for the *STRN-PDGFR* case since no RNA was available, and in fact the fusion was cloned from DNA extracted from a single bone marrow slide. As above, involvement of *PDGFR* in both cases was suspected by the finding of a translocation involving 4q12 in patients with a MPD and eosinophilia.

- *KIF5B-PDGFR* was identified following a screen for over-expression of *PDGFR* by the generic RT-PCR assay I developed. Of 200 patients screened, just one strongly over-expressed *PDGFR* (although unknown to us at the time, the patient had a complex rearrangement involving 4q12). The fusion was identified by bubble PCR from genomic DNA.

- The variant *ETV6-PDGFR* case was referred with a t(5;12) karyotype and a clinical response to imatinib, strongly suggesting involvement of *PDGFR*, however RT-PCR for *ETV6-PDGFR* was negative, Over-expression of *PDGFR* was confirmed by the MLPA RT-PCR assay and therefore I performed 5’-RACE PCR leading to the unexpected finding of the variant fusion.
All the fusions I found conformed broadly to the consensus structure of tyrosine kinase fusions following the paradigm of BCR-ABL (Figure 6.2). The N-terminal region of the partner gene was fused in-frame to the C-terminal region of the kinase, including the entire catalytic domain.

![Diagram showing the universal structure of tyrosine kinase fusions based on the BCR-ABL paradigm. The N-terminal dimerisation domain of the partner is fused to the entire catalytic C-terminal domain of the tyrosine kinase.](image)

**Figure 6.2:** Schematic showing the universal structure of tyrosine kinase fusions based on the BCR-ABL paradigm. The N-terminal dimerisation domain of the partner is fused to the entire catalytic C-terminal domain of the tyrosine kinase.

The partner genes are widely expressed, usually in all normal tissues but certainly in haematopoietic cells and thus provide a promoter to drive transcription of the fusion. All partner proteins analysed to date (except FIP1L1) contain one or more dimerisation domains that are also required for the transforming activity of the fusion proteins. Homotypic interactions between these domains lead to dimerisation or oligomerisation of the fusion protein, thus mimicking the normal process of ligand-mediated dimerisation that is essential for the activation of normal tyrosine kinases. Although I was unable to test my fusions functionally, all the partner proteins had plausible dimerisation domains that are retained in the fusion genes, e.g.

• STRN: an N-terminus coiled-coil region is thought to be crucial for homo- and hetero-oligomerisation of the protein and hence of the fusion protein (Gaillard, *et al* 2006).


• KIF5B: a central alpha-helical rod domain that enables the protein to dimerise (Huang, *et al* 1994).

• ETV6: the pointed (helix-loop-helix) domain has been established functionally as a self association motif that is essential for the transforming activity of ETV6-PDGFRβ (Carroll, *et al* 1996, Golub, *et al* 1994).

The normal functions of the partner proteins are diverse, however with the characterisation of increasing numbers of tyrosine kinase fusions, some intriguing patterns have emerged. Indeed it is possible that the partner proteins are not just contributing oligomerisation motifs and ubiquitous promoter expression but may also be playing a more direct role in cell dysregulation, as discussed below.

**Deregulated endocytosis.** Several translocations involving PDGFRα or PDGFRβ generate in-frame fusions with components of the endocytic machinery e.g. Rabaptin-5, Hip1 and PDE4DIP. My work adds three further endocytosis-related partners, GOLGA4, KIF5B and STRN and, in addition, a further two PDGFRβ partner genes submitted for publication (*Bin2* and *WDR48*) identified in our laboratory also have endocytic links. Endocytosis is recognised as a fundamental mechanism of signal attenuation by committing receptors to degradation or recycling back to the plasma membrane. Potentially, interference with endocytosis may affect normal receptor down-regulation, resulting in hyper-proliferation (Mosesson, *et al* 2008).

Furthermore, if signalling can occur at the endosomes, as has been reported by Wang Y *et al.*, (2004) for Rabaptin5-PDGFRβ, then this may suggest a mechanism for prolonged signalling (Wang, *et al* 2004a).
**GOLGA4** (tGolgin1; Golgin-245). tGolgin-1 is a member of the Golgin family of coiled-coil proteins associated with specific sub-domains of the **trans**-Golgi network (TGN) and is thought to play a role in Rab6-regulated tethering events in membrane fusion and as a structural support for the Golgi cisternae (Barr 1999, Kjer-Nielsen, *et al* 1999, Munro and Nichols 1999).

The TGN itself is a series of interconnected tubules and vesicles located at the **trans** face of the golgi stack and its purpose is to process and sort glycoproteins and glycolipids at the periphery of the endosomal pathway (Figure 6.3A). The Golgins are further characterised by a highly conserved C-terminal GRIP domain that localises the proteins to the TGN. Over-expression of the isolated GRIP domain has been shown to result in alteration of the pericentriolar distribution of the TGN integral membrane and coat components whilst concomitantly inhibiting vesicular transport from the TGN to the plasma membrane (PM). This suggests that these proteins help maintain the integrity of the TGN by regulating resident protein localisation (Yoshino, *et al* 2003). tGolgin-1 is suggested therefore to be a tubulovesicular carrier that emerges from the TGN and hence is likely to be a regulator of transport from the TGN to the recycling endosome and PM. Recently Lieu et. al., (2008) have further defined the role of tGolgin-1 and shown that not only is it essential for intracellular trafficking but it might additionally act as a vesicle coat, however its complete function still remains to be fully elucidated (Lieu, *et al* 2008).
Figure 6. 3A: Schematic showing anterograde and endocytotic/retrograde transport through cellular organelles. ‘Stop signs’ indicate which pathways might be affected by the GOLGA4-PDGFRβ fusion, leading to possible aberrant endocytosis and increased cellular transformation. tGolgin-1 is a tubulovesicular carrier likely to be a regulator of transport from the TGN to the recycling endosome and PM. This is a retrograde pathway that allows signalling proteins to escape the degradative environment of the late endosomes and lysosomes and to reach internal sites where they can interact with their intracellular targets or be exported back to the PM via the recycling endosomes permitting increased signaling (represented by arrow in green box). EE, early endosome; ERGIC, ER-Golgi intermediate compartment; Glut4, a glucose transporter molecule; LE, late endosome; RE, recycling endosome, and TGN, trans-Golgi network.

It can only be speculated upon how the fusion of tGolgin-1 and PDGFRβ might abrogate endocytosis and enhance the oncogenic potential of the fusion. One such hypothesis is that the partner protein plays a role in localising the fusion to a particular subcellular compartment, in this case the TGN. However for this specific
fusion the c-terminal GRIP domain (essential for localisation to the Golgi) is lost and it is unclear where in the cell the fusion would be. It has also been shown that in general tyrosine kinase fusions are found to be cytosolic due to deletion of the N-terminal signalling peptide from the kinase (Lamorte and Park 2001). Hence they avoid entry to the endocytic pathway anyway, thereby presumably preventing degradation and prolonging signalling.

A second hypothesis concerns the method of uptake of the fusion/constitutively activated receptor. It has been suggested that differential clathrin and caveolae mediated endocytic routes of the same cargo may regulate signalling capacity and endocytic fate (Di Guglielmo, et al 2003, Sigismund, et al 2005). Clathrin mediated endocytosis is a form of membrane invagination and vesicle scission of receptors clustering in clathrin coated pits. The released vesicles discard the coat before fusing with tubulovesicular compartments such as the early endosomes where the cargo is sorted for either degradation in the lysosomes or recycling back to the PM. The caveolae pathway also involves membrane invagination and vesicle budding, however the vesicles are caveolin-positive and are directed first to larger caveolin-enriched organelles termed caveosomes before being further targeted to the early endosomes, thereby allowing convergence with the clathrin mediated pathway. Both pathways are a means of signal attenuation however it is possible that one or the other might promote recycling over degradation further compounded by the role of tGolgin-1 as regulator of transport from the TGN to the recycling endosome and plasma membrane.

Finally and perhaps most interestingly, is the role of tGolgin-1 in retrograde transport from the endosomes to the TGN. This pathway is regulated by specific SNAREs and Rab proteins signifying it is a pre-existing cellular pathway not just a subversive pathway. This retrograde pathway may allow signalling proteins to escape the degradative environment of the late endosomes and lysosomes to reach internal sites where they can interact with their intracellular targets or be exported back to the PM via the recycling endosomes (Figure 6.3A) (Mallard, et al 2002). Therefore, efficient
retrograde transport from the early endosomes to the TGN is important for the recycling of endogenous proteins (Lieu, et al 2007). A role for Golgin-1 in this pathway has been proposed by Yoshino et. al., (2005), who showed that recruitment of microtubule motors (such as the kinesins) to the Golgi requires tGolgin1-dependent retrograde transport from the endosomes. This was observed after experiments over-expressing the GRIP domain were performed that resulted in specific disruption of TGN morphology, protein localisation and function. Concomitantly, vesicular transport from the TGN to the PM was inhibited (Yoshino, et al 2005). This suggests the GRIP domain proteins, such as tGolgin-1 function in TGN maintenance probably by regulating recycling from the endosomes. Thus the TGN requires a functional GRIP domain, shown not only by over-expression of GRIP domains competing for binding, but also by inactivity observed with N-terminal tGolgin-1 fragments that lack the GRIP domain, such as might be observed with the GOLGA4-PDGFRβ fusion. This would imply the fusion may not be recycled back to the PM but instead retained at the TGN or in the early endosomes where conceivably signalling can still occur, aiding oncogenicity.

Yoshino et. al., 2005 have also shown that cells depleted of tGolgin-1 mis-sort internalised glycolipids to the late endosomes/lysosomes and accumulate aberrant endosomal structures, highlighting a possible secondary outcome of disrupting membrane recycling. These results are consistent with the function of golgin-97 (Lu, et al 2004a) and the dominant-negative GRIP-domain over-production in which endosome to TGN recycling is inhibited (Yoshino, et al 2003). The consequence of such mis-sorting to the late endosomes would be eventual depletion from the Golgi of recycling glycosphingolipids, so proteins and receptors cannot be modified and shipped back to the PM. This could mean prolonged signalling and reduced degradation and could also perturb multiple other pathways if required proteins are not modified or transported where required, adding to cellular transformation.

*KIF5B. KIF5B* is an N-terminal (plus-end motor) kinesin that is a member of the superfamily of Kinesin-1 molecular motor proteins that are responsible for
microtubule-dependent transport of cargo in eukaryotic cells (Goldstein and Yang 2000). These motor proteins are powered by the hydrolysis of ATP enabling them to move cargo such as vesicles, proteins and lipids throughout the cell. The plus-end and minus-end microtubule specific motors are responsible for opposing movement along the filaments within the cell (Hirokawa and Takemura 2005). KIF5B in particular, is important for microtubule-based endosome-motility, by regulating early endosome trafficking and has been further implicated specifically in lysosomal and mitochondrial transport (Nakata and Hirokawa 1995, Tanaka, et al 1998).

It is important to highlight the link between the Golgi and the motor protein Kinesin to comprehend how these proteins found fused to a PDGFR tyrosine kinase are related not only to endocytosis but also how these proteins interact at the cellular level. As discussed previously, proteins sequestered at the TGN are selectively sorted and packaged into distinct carrier vesicles destined for either the PM or endosomal compartments. The Golgi is a dynamic structure and to maintain this level of activity it is sustained by a complex cytoskeletal matrix composed of microtubules and actin-spectrin networks as well as by temporal localised proteins such as the Golgins. The interaction between these filaments and the Golgi membrane is further mediated by diverse families of ‘motor proteins’ such as the kinesins. Interestingly, these two endocytic proteins interact at the Golgi and together help maintain the overall integrity of the Golgi structure whilst also supporting protein transport.

Overwhelming evidence now suggests that KIF5B is specifically involved in transporting vesicles from the TGN to the PM or endosomes further localising both KIF5B and tGolgin1 to the same subcellular domain: the TGN. Additionally, if KIF5B expression is reduced the Golgi apparatus has been shown to switch from an extended ribbon to a much more compact structure (Feiguin, et al 1994) with ensuing inefficient protein modification and sorting, and decreased trafficking. An additional layer of complexity within the trafficking of vesicles from the TGN is the requirement of different motor proteins depending on the cargo being transported and the specific cell type, for example epithelial or fibroblasts. Due to the constitutive activity of the kinase caused by the KIF5B-PDGFRA fusion it is likely that KIF5B expression is enhanced. This may cause specific proteins to be modified and
transported far more efficiently than others, resulting in an unbalanced accumulation of specific proteins and a reduction in others creating generalised cell defects perhaps adding to the transforming potential of this fusion.

Another hypothesis for this fusion’s oncogenicity is related to *KIF5B*’s role in plus-end motility of vesicles. It has been shown to work in association with the minus-end motility kinesin *KIFC1*, and if either Kinesin is inhibited reduced vesicle fission occurs suggesting that the opposing forces from the activity of both motors are required for fission to occur (Nath, *et al* 2007). This could imply that vesicles containing activated receptors are unable to be sorted for degradation to the lysosomes due to inefficient vesicle fission. The active receptors may then be able to continue signalling increasing the transformation potential and disrupting receptor attenuation via lysosomal degradation (Figure 6.3B).

*KIF5B* is a tetramer composed of two heavy chains and two light chains that together form a rod-like molecule consisting of two globular head domains, a stalk domain and a fan-like tail domain. The head domain is thought to bind to the microtubules, and the fan-like ends are thought to be associated with vesicles and membranous organelles requiring transport (Hirokawa, *et al* 1989). The N-terminal motor/head domain is present within the fusion with *PDGFRA* so it still retains the ability to move along the microtubules, however with the c-terminal fan-like region absent it may not be able to interact with the vesicles, organelles and proteins requiring intracellular transport. Transport is very important for the proper targeting of lipids, proteins and organelles to distinct subcellular domains without this function it is plausible that more generalised defects within the cell may occur. This also implies the fusion itself would remain cytosolic, possibly bound to the microtubule network, avoiding endocytosis resulting in increased signalling.
Figure 6.3 B: Schematic showing anterograde and endocytotic/retrograde transport through cellular organelles. ‘Stop signs’ indicate which pathways might be affected by the KIF5B-PDGFRα fusion, leading to possible aberrant endocytosis and increased cellular transformation. The plus-end and minus-end microtubule specific motors are responsible for opposing movement along the filaments within the cell, represented by purple arrows. KIF5B is specifically involved in TGN-to-PM or endosome trafficking, enhanced expression of this protein caused by the fusion may cause specific proteins to be modified and transported far more efficiently than others, resulting in an unbalanced accumulation of specific proteins and a reduction in others creating generalised cell defects. Vesicles containing the activated fusion may also be unable to be sorted for degradation to the lysosomes due to inefficient vesicle fission. The active receptors may then be able to continue signalling within the EE or be recycled (represented by arrow in green box) increasing the transformation potential and disrupting receptor attenuation via lysosomal degradation. It is also speculated that overexpression of KIF5B may result in continuous transportation to the required subcellular domains and subsequent enhanced signalling and cell turnover and reduced lysosomal degradation. EE, early endosome; ERGIC, ER-
Golgi intermediate compartment; Glut4, a glucose transporter molecule; LE, late endosome; RE, recycling endosome, and TGN, trans-Golgi network.

Additionally, Cardoso et. al., (2009) have recently reported up-regulation of KIF5B mRNA in several types of cancer including bladder cancer, advanced gastric cancer, squamous cell carcinoma and BRCA1-associated breast cancer. (Cardoso, et al 2009). They were also able to show that when KIF5B was depleted in vitro HeLa cells acquired an elongated cellular phenotype concomitant with significant growth inhibition and cell death. It could be speculated that over-expression of KIF5B may have the opposite effect with effective continuous transportation to the required subcellular domains and subsequent enhanced signalling and cell turnover. Further work from this group implicates that KIF5B may function in transporting lysosomes to the plasma membrane destined for exocytosis and moreover may play a role in positioning lysosomes at distinct sites in the cytoplasm, particularly perinuclear. Consequently, depletion of KIF5B would allow other motor proteins to transport the lysosomes to the cortical areas of the cell resulting in lysosomal aggregation.

**STRN.** Striatin is a multimodular, WD-repeat, and calmodulin-binding protein. It is found in the cytosol in addition to being associated with membranes, and is composed of several different, colinearly aligned protein-protein association modules. These include a caveolin-1 binding motif, a coiled-coil structure, a Ca\(^{2+}\)-calmodulin-binding motif and a large WD-repeat domain (Bartoli, et al 1998, Castets, et al 1996, Gaillard, et al 2001). Striatin is a scaffolding protein involved in calcium dependent signalling (Bartoli, et al 1999) with a putative role in trafficking (Baillat, et al 2001), whilst also being shown to directly bind caveolin-1 (Gaillard, et al 2001).

To further define the role of Striatin, Baillat et. al., 2001 used a yeast two-hybrid strategy to search for partners and identified phocein, an intracellular protein. Its distribution mirrors that of striatin, due to its association and a lack of transmembrane binding motifs. Interestingly, phocein is also highly localised to the Golgi complex indicating a link between this organelle and striatin (Baillat, et al 2001). It is hypothesised that phocein may be a component of a novel vesicular coat allowing
transport to the Golgi complex and further adding to striatins’ role in membrane trafficking.

Additionally, it is speculated that caveolin-1 binding the caveolin binding site of striatin may directly regulate its function by concentrating it within caveolin-enriched domains, concordant with its putative role in membrane trafficking (Baillat, et al 2001). The caveolae themselves are thought to be subcellular domains that regulate among other things, intracellular Ca\(^{2+}\) concentration and Ca\(^{2+}\)-dependent signal transduction (Isshiki and Anderson 1999). This links both the Ca\(^{2+}\)-calmodulin binding domain of striatin and the caveolin-1 binding domain. It is also well known that caveolins interact with signalling proteins from the plasma membrane, such as tyrosine kinases, Ras and G-proteins (Chang, et al 1994, Lisanti, et al 1994, Sargiacomo, et al 1993), and is consistent with a role for caveolae in sequestering signalling molecules to attenuate signalling after receptor stimulation, similar to clathrin coated pits. This further addresses the role of striatin as a possible signalling transduction component due to its links and enrichment at the caveolae. Furthermore, it has been shown that transformation of NIH3T3 cells by various oncogenes leads to a reduction in cellular levels of caveolin, suggesting functional alterations in caveolae may play a critical role in oncogenic transformation, perhaps by disrupting contact inhibition and negatively regulating the activation state of the p42/44 MAP kinase cascade (Engelman, et al 1997, Galbiati, et al 1998, Koleske, et al 1995). Additionally the reduction of caveolae in transformed cells could further prevent the down-regulation of growth-promoting signals, thus contributing to uncontrolled proliferation. It is possible that the STRN-PDGFRα fusion, by interacting with caveolin via the caveolin-1 binding site of striatin, is either maintained at the cell surface constitutively active, or is sequestered from the cell surface and internalised via the caveolae where it may be transported through the endocytic pathway and recycled due to its interaction with caveolin back to the cell surface thereby allowing continuous signalling and promoting further to cell transformation (Figure 6.3C).
In conclusion, it is possible this protein with its calmodulin-binding site a WD-repeat, scaffolding domain and a caveolin-1 binding motif enhances cell transformation, adding to the growing list of proteins now known to have a dual capacity in endocytosis and signalling (Di Fiore and De Camilli 2001).

Figure 6.3 C: Schematic showing anterograde and endocytotic/retrograde transport through cellular organelles. ‘Stop signs’ indicate which pathways might be affected by the STRN-PDGFRα fusion and lead to possible aberrant endocytosis and increased cellular transformation. Reduction of caveolae in transformed cells is thought to prevent the down-regulation of growth-promoting signals, thus contributing to uncontrolled proliferation. It is speculated that the striatin caveolin-1 binding site within STRN-PDGFRα enables it to be either retained at the cell surface constitutively active (represented by arrow in green box), due to reduced levels of caveolae or is sequestered from the cell surface and internalised via the caveolae and recycled allowing continuous signaling whilst promoting cell transformation. EE, early endosome; ERGIC, ER-Golgi intermediate compartment; Glut4, a glucose transporter molecule; LE, late endosome; RE, recycling endosome, and TGN, trans-Golgi network.
Finally, the finding of mutations in *CBL*, a protein that is normally involved in attenuating receptor tyrosine kinase signalling, further strengthens the hypothesis that subversion of endocytic control is pathogenetically important. These mutations have been shown to prevent internalisation and ubiquitination of the RTKs FLT3, PDGFRα/β and EGFR, leading to prolonged signalling (Sargin, *et al* 2007).

*Pre-mRNA processing.* The characterisation of a second novel fusion gene: *CPSF6-FGFR1* in parallel with the discovery in our laboratory of a novel *ABL* fusion: *SFPQ-ABL* led us to the identification of a second functionally related group of partner proteins (Hidalgo-Curtis, *et al* 2008). Both *CPSF6* and *SFPQ* encode proteins that are involved in pre-mRNA processing. SFPQ is an essential factor involved in RNA splicing and was first identified in a complex with polypyrimidine tract-binding protein (PTB) (Patton, *et al* 1993). Interestingly, CPSF6 co-localises in purified spliceosomes with SFPQ and another splicing factor NonO, linking these proteins to both transcription and processing of pre-mRNAs (Dettwiler, *et al* 2004, Rappsilber, *et al* 2002, Zhou, *et al* 2002). ZMYM2 (ZNF198), a partner protein commonly fused to FGFR1, is also thought to be involved in this process, possibly by acting as a scaffolding protein that brings RNA-binding proteins together to facilitate pre-mRNA splicing (Kasyapa, *et al* 2005). Strikingly, ZMYM2 has also been found to immunoprecipitate with SFPQ (Kasyapa, *et al* 2005). Adding further weight to this group of proteins is FIP1L1, which fuses to PDGFRα in CEL (Cools, *et al* 2003a). Not only is this one of the most common fusion genes described in eosinophilic MPD patients, FIP1L1 has also been show to have a role in polyadenylation and pre-mRNA splicing (Kaufmann, *et al* 2004, Preker, *et al* 1995).

How these functionally related proteins might impact on neoplastic growth is not clear. It is possible that the fusions may interfere with normal pre-mRNA processing complexes thereby inhibiting processing of many transcripts. Alternatively they may result in increased processing. Both splicing and polyadenylation are known to be important control points with regard to gene expression and it is conceivable that
alterations could disrupt the fine balance between proliferation, differentiation and apoptosis of these cells by altering mature mRNA levels of key genes. Consistent with this idea, it has been shown that high levels of poly(A)polymerase (PAP), the enzyme that causes elongation of the poly(A) tail, are associated with proliferating cells and that the levels reduce in cells undergoing apoptosis (Thomadaki, et al 2008).

Centrosomal proteins. A third group of functionally related partner proteins has also been observed by Delaval et al., (2005). This group initially observed that FGFR1OP is a centrosomal protein and that the FGFR1OP-FGFR1 fusion is also centrosomal. Furthermore, FGFR1OP1-FGFR1 was shown to be active at the centrosome and activate multiple signalling pathways (Delaval, et al 2005a). Several other centrosomal partner proteins have been described, e.g. CEP110 fuses to FGFR1, PCM1 fuses to JAK2, CDK5RAP2 fuses to PDGFRα, NDE1 and NIN fuse to PDGFRβ.

The centrosome consists of a pair of centrioles surrounded by pericentriolar material and plays an integral role in directing the organisation of the cytoplasmic microtubules and the assembly of the mitotic spindles (Rieder, et al 2001). Aberrations in the number of centrosomes lead to mitotic defects and cause chromosome mis-segregation and centrosome abnormalities and are therefore frequent in many common cancers (Nigg 2002). Delaval et. al., hypothesise that the centrosome, which is linked to the microtubules, is close to the nucleus, and is connected to the Golgi apparatus and the proteasome, could serve to integrate multiple signalling pathways controlling cell division, cell migration and cell fate. Abnormal kinase activity at the centrosome may be an efficient way to pervert cell division in malignancy. In support of this idea, it has been reported that an unidentified centrosomal mechanism controls the number of neurons generated by neural precursor cells (Bond, et al 2005). It is conceivable that similar mechanisms might operate in haematopoietic development.
These three functional clusters of partner proteins are striking and, as discussed above, may directly impact on cellular processes and promote transformation. However since the clusters impact on such functionally diverse processes, it is perhaps unlikely that any of the postulated interactions are critical for transformation, although they might produce cellular changes that modify the phenotype. In support of this, it has been observed that the clinical features in patients with FGFR1 rearrangements are slightly different depending on the identity of the partner protein (Macdonald, et al 2002b). However other interpretations are also possible, for example it might be that endocytic, pre-mRNA processing and centrosomal proteins are simply large families, widely expressed and tend to contain dimerisation domains. It is also possible that altered subcellular localisation may be important, by enabling the kinase to interact with novel substrates. For example, the \textit{FIG-ROS} fusion found in glioblastoma has been shown to be directed by the partner protein to the Golgi (Charest, et al 2003), a localisation that is thought to be essential to its transforming ability. However whether this is due to interactions with novel substrates or whether specific localisation increases the effective concentration of the fusion protein above a critical threshold is unclear.

\textbf{6.2 Routine detection of tyrosine kinase fusions}

Detection of tyrosine kinase fusions is important for several reasons. First, the finding of a fusion such as \textit{FIP1L1-PDGFRA} in a cytogenetically normal patient confirms the diagnosis of a clonal neoplastic disorder. Second, specific fusions are indicators for specific targeted therapies, e.g. patients with \textit{PDGFRA} or \textit{PDGFRB} translocations respond very well to imatinib. Third, the involvement of particular tyrosine kinases may be prognostically important, e.g. patients with \textit{FGFR1} translocations have a strong propensity to progress to AML and should be considered as candidates for allogeneic bone marrow transplantation. It is clearly important therefore that diagnostic procedures are available to detect fusion tyrosine kinases accurately. The problem arises from the sheer number of different fusions that have been described,
which makes it impractical and very expensive to screen for each by RT-PCR. Furthermore it is likely that additional fusions with novel partners remain to be discovered. For many cases it would be possible to use FISH with split-apart probes however rare translocations often mask complex rearrangements at the molecular level and consequently split-apart FISH may fail to detect a rearrangement (Kulkarni, et al 2000, Reiter, et al 1998).

To help address this problem, and also to try and screen for new fusions, I developed two generic RT-PCR assays to detect over-expression of PDGFR genes. The assays were designed to look for over-expression of the tyrosine kinase domain relative to an extracellular control band and which might therefore indicate the presence of a PDGFR fusion gene. Following validation using samples with known PDGFR fusions, I screened approximately 200 patients for both PDGFRA and PDGFRB over-expression. Of the six patients found to over-express PDGFRB, one had a previously undetected variant ETV6-PDGFRB fusion that had arisen as a result of a complex translocation. Unexpectedly, 2 had FGFR1 rearrangements. The reason for this is unclear and requires further investigation, initially to see if this is a consistent finding in cases with FGFR1 fusions. Of the remaining 3 patients, only 2 had material available for further study and no abnormalities were found. Over-expression of PDGFRA was identified in 14 patients. Of the 7 that had sufficient material to investigate further with bubble PCR one had the novel KIF5B-PDGFRA fusion, one had a previously undetected FIP1LI-PDGFRA fusions and five showed no abnormalities.

Overall then, the generic RT-PCR assays led to the identification of fusions that had not been detected previously and is therefore clearly a useful diagnostic tool. However it does produce quite a large number of false positives that require extensive further investigation. It would be desirable to have a technique that reduced these, and was also more readily extendable to screen for fusions involving additional tyrosine kinases. I therefore designed a cDNA MLPA assay to look at expression of the four kinases most commonly involved in MPDs: ABL, FGFR1, PDGFRA and
**PDGFRB.** MLPA is a probe-based technique that was originally designed to detect deletions and amplifications of multiple target genomic DNA sequences in a single reaction. The cDNA MLPA assay was designed on the same principle as that of the generic RT-PCR assays i.e. to look at differential expression of the tyrosine kinase portion of the gene (which would be retained from any fusion) compared to a second probe that hybridised to a portion of the extracellular domain or 5’ region that would not be retained in the fusion.

The assay was validated using normal control cDNA samples, cell lines and patient cDNA with known tyrosine kinase gene fusions. The majority of positive controls were correctly identified, however the intracellular:extracellular ratio of the other unaffected probes was rarely 1.0 as predicted. Furthermore due to the differing levels of expression of each gene it proved difficult to define a single set of conditions that effectively covered all four genes and, in addition, the assay was very dependent on the quality of the cDNA and probe sensitivity. Nevertheless, this technique did lead to the discovery of a novel variant *ETV6-PDGFRB* fusion in one case who over-expressed the *PDGFRB* kinase domain. Overall, the cDNA MLPA seemed to be more useful as a general screen allowing the analysis of 3 tyrosine kinase genes in one test limiting sample volume which is always precious, however it was time consuming and less reproducible. The generic RT-PCRs worked relatively well for the two targets that are clinically most important, *PDGFR* and *PDGFRB*, however over time they also proved temperamental with sensitivity to cDNA quality, primer age and the thermocycler that was used. The yield of previously unknown positive cases from either technique was low and perhaps the most effective approach would be to use the generic RT-PCR assays to screen for *PDGFR* over-expression associated with *FIP1L1-PDGFR* in cases with eosinophilia, and to screen for *PDGFR* and *PDGFRB* abnormalities in cases with a complex karyotype. This strategy is unlikely to miss any true positives but will minimise the number of false positives. Other generic strategies to detect gene fusions are also possible, e.g. exon arrays or real time PCR assays to look at expression across target genes, but these would be much more expensive.
6.3 Loss of heterozygosity analysis as an approach to identify genes harbouring activating mutations

The main focus of my work was to identify new abnormalities in atypical MPDs. LOH is a genetic mechanism that creates a homozygous or hemizygous cell by loss of the wild-type allele as a result of various mutational events. These include deletion, mitotic recombination, localised gene conversion, point mutation, or non-dysjunction. My objective was to identify new oncogenes or tumour suppressor genes that are marked by these regions of LOH, and specifically regions of acquired uniparental disomy (aUPD), or copy number neutral LOH, formed as a consequence of mitotic recombination. Uniparental disomy (UPD) refers to a mechanism in which both copies of a chromosome pair have originated from one parent. This can result in diverse clinical conditions as a consequence of homozygosity for recessive mutations or aberrant patterns of imprinting. Somatic UPD has also been shown to arise as a consequence of recombination occurring during mitotic cell divisions. This acquired UPD (aUPD) has been shown be associated with pathogenetic driver mutations in genes within the region of homozygosity such as JAK2 V617F, FLT3 ITD, CEBPA and WT1 mutations (Fitzgibbon, et al 2005, Griffiths, et al 2005, Kralovics, et al 2005b, Raghavan, et al 2005). I aimed to search for new regions of aUPD as a means to identify regions harbouring genes that are mutated in aMPDs and transformed CML.

Numerous techniques are capable of identifying regions of LOH however SNP arrays have the advantage as they can simultaneously detect copy number and zygosity. For my study, I utilised 50K SNP arrays which are entirely adequate for detecting large regions of aUPD although the resolution of copy number changes turned out to be very limited. A major drawback of my study was that paired sample analysis was not possible due to the lack of constitutional DNA, thus I was unable to determine if tracts of homozygosity were acquired or whether they were inherited. Fortunately though I found that 40% of aMPD cases had large (>20Mb) tracts of copy neutral
homozygosity that are very unlikely to have been inherited. However, aUPD was uncommon in the other patient subgroups (CML-BC, MF and V617F negative PV).

In aMPDs, 8 distinct regions of aUPD were identified indicating considerable genetic heterogeneity in these disorders. However two novel, recurrent regions were seen at 7q and 11q, each in 10% of cases. Initially, because of the strong association between MPDs and tyrosine kinases, I fully re-sequenced the three tyrosine kinase genes on 7q, \(EPHA1\), \(EPHB6\) and \(MET\) but no mutations were found. There are no tyrosine kinase genes on 11q so it was apparent that both abnormalities must be targeting other genes. I then went on to select candidates at both locations for further analysis. Both regions were relatively large and contained several hundred genes, including a large number of candidates that could potentially be involved in malignancy. I focused mainly on those that were known to encode signal transduction components and were either known oncogenes (\(BRAF\), \(CBL\)) or closely related to known oncogenes (\(PI3KCG\)). I also checked for the \(MLL\) ITD as this gene is widely involved in leukaemia. This analysis led to the finding of \(CBL\) mutations in all three cases with 11q aUPD.

The \(CBL\) gene is located at 11q23.3 and encodes a widely expressed ubiquitin ligase that functions as a negative regulator of several receptor protein tyrosine kinases by inducing ubiquitination of the receptors, thereby promoting their sorting for endosomal degradation or recycling. In addition, CBL plays a positive role in signalling by serving as a scaffold for the recruitment of downstream substrates. The mutations I found predicted missense substitutions in the RING/linker domain, L380P in two cases and P417A in the third. This region is responsible for the ubiquitin ligase activity and it is likely that these mutations, like the R420Q substitution previously described in AML, have transforming activity and are therefore causative driver mutations (Sargin, et al 2007).

Having found \(CBL\) mutations in the three 11q aUPD cases, I next screened 547 patients, including those with PV, IMF, ET, SM, HES, CMML, aCML and other
aMPDs. A further 23 patients with *CBL* mutations were identified, the majority of which were missense changes, plus one nonsense mutation and some intronic changes, two of which were shown to result in skipping of exon 8. Although I found the R420Q in 4 cases, the majority of mutations were novel. The mutations were associated predominantly with a diagnosis of aCML, CMML and MF i.e. highly related and relatively aggressive diseases. In three cases for which suitable material was available, the mutations were shown to be acquired. In two cases, the mutations were acquired during progression of a pre-existing MPD indicating that in these cases at least, *CBL* mutations were secondary events.

There are a number of obvious future avenues of investigation that should be undertaken following the results of my studies, e.g.

- Functional analysis to prove that the *CBL* mutations I found are indeed transforming. This has been achieved for some of the mutants by testing if they transform 32D cells over-expressing *FLT3*, as has been described for R420Q (Grand 2009, Sargin, *et al* 2007).

- Investigation of genetic events that might co-operate with *CBL* mutations. One possibility is over-expression of one or more tyrosine kinases, and potentially this could be investigated by gene expression profiling. Another is indicated by the fact that one of my cases has both a *CBL* mutation and 7q aUPD suggesting that they might co-operate. Once the 7q aUPD mutation is identified this can be formally tested, probably most appropriately in mouse models.

- If *CBL* mutants function by increasing tyrosine kinase signalling then it is likely that the malignant clone might be susceptible to inhibition by tyrosine kinase inhibitors. This could be investigated by testing the effects of inhibitors against patient and control cells.

- The clinical phenotype associated with *CBL* mutations needs to be more closely defined. Ultimately it should be possible to see if any clinical features are correlated with specific mutations and also since aUPD is effectively a
second hit, it is possible that homozygous mutations are associated with more aggressive disease.

- Identification of the gene underlying 7q aUPD. Other members of the group have eliminated other candidate genes and are planning to sequence the coding sequence (exons) of all genes in the region using custom designed pull down arrays and Illumina sequencing.

- Investigation of smaller regions of aUPD that may have been missed by my analysis. Efficient algorithms have been described to pick up low level aUPD that is missed by standard approaches (Yamamoto, et al 2001) and this, in combination with the current state-of-the-art SNP 6.0 arrays, may prove to be a powerful approach to identify novel regions harbouring mutated genes.

### 6.4 Current knowledge of the pathogenesis of atypical MPDs

Prior to my findings it was estimated that at least part of the genetic basis of approximately a third of aMPDs was understood, with roughly 2% having tyrosine kinase fusions, 5% having \textit{FLT3} ITD or \textit{KIT} D816V, 12.5% \textit{V617F JAK2} and 12.5% activating \textit{NRAS} mutations. (Cross 2006, Jones, et al 2005). Clearly a lot of detail remains to be filled in, especially the targets of 7q and other regions of aUPD plus how different mutations may co-operate together, but following my work, I estimate that we now know something about two thirds of cases (Figure 6.4) with \textit{CBL} mutations, 7q aUPD and other regions of aUPD each seen in roughly 10% of aMPDs. Further works needs to be performed to determine which combinations of mutations occur in individual patients, how they co-operate and how they might be targeted therapeutically.
Figure 6.4: Pie chart illustrating the proportion of aMPD cases related to known genetic variants, such as V617F and CBL mutations aUPD compared to unknown pathogenicity.
List of References


Armstrong, E., Kastury, K., Aprelikova, O., Bullrich, F., Nezelof, C., Gogusev, J., Wasmuth, J.J., Alitalo, K., Morris, S. & Huebner, K. (1993) FLT4 receptor tyrosine kinase gene mapping to chromosome band 5q35 in relation to the t(2;5), t(5;6), and t(3;5) translocations. Genes Chromosomes Cancer, 7, 144-151.


Bellanne-Chantelot, C., Chaumarel, I., Labopin, M., Bellanger, F., Barbu, V., De Toma, C., Delhommeau, F., Casadevall, N., Vainchenker, W., Thomas, G. &


Bennett, J.H. (1845) Case of hypertrophy of the spleen and liver in which death took place from suppuration of the blood. *Edinburgh Medical and Surgical Journal, 64*, 413-423.


Cragie (1845) Edinburgh Medical and Surgical Journal, 64, 400-412.


cellular oncogene is translocated to the Philadelphia chromosome in chronic

de Vries, A.C., Stam, R.W., Kratz, C.P., Zenker, M., Niemeyer, C.M. & van den
myelomonocytic leukemia. Haematologica, 92, 1574-1575.

centrosome connection. Leukemia, 19, 1739-1744.

Delaval, B., Letard, S., Lelievre, H., Chevrier, V., Daviet, L., Dubreuil, P. &
Birnbaum, D. (2005b) Oncogenic tyrosine kinase of malignant hemopathy targets the

Demiroglu, A., Steer, E.J., Heath, C., Taylor, K., Bentley, M., Allen, S.L., Koduru,
P., Brody, J.P., Hawson, G., Rodwell, R., Doody, M.L., Carnicero, F., Reiter, A.,
leukemia fuses BCR to FGFR1: transforming activity and specific inhibition of
FGFR1 fusion proteins. Blood, 98, 3778-3783.

Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate
RNA binding, protein-protein interactions, and subcellular localization. J Biol Chem,
279, 35788-35797.


Kotecha, N., Flores, N.J., Irish, J.M., Simonds, E.F., Sakai, D.S., Archambeault, S.,
cell profiling identifies aberrant STAT5 activation in myeloid malignancies with
specific clinical and biologic correlates. Cancer Cell, 14, 335-343.

chromosome 9p is a frequent stem cell defect in polycythemia vera. Exp.Hematol.,
30, 229-236.

Kralovics, R., Passamonti, F., Buser, A.S., Teo, S.S., Tiedt, R., Passweg, J.R.,

Kralovics, R., Teo, S.S., Buser, A.S., Brutsche, M., Tiedt, R., Tichelli, A.,
expression in myeloproliferative disorders correlates with activation of signaling by


Kulkarni, S., Heath, C., Parker, S., Chase, A., Iqbal, S., Pocock, C.F., Kaeda, J.,
platelet-derived growth factor receptor beta in BCR-ABL-negative myeloproliferative
disorders with a t(5;10)(q33;q21). Cancer Res., 60, 3592-3598.

Kuno, Y., Abe, A., Emi, N., Iida, M., Yokozawa, T., Towatari, M., Tanimoto, M. &
myelodysplastic syndrome with t(9;12)(q22;p12). Blood, 97, 1050-1055.

La Starza, R., Rosati, R., Roti, G., Gorello, P., Bardi, A., Crescenzi, B., Pierini, V.,
Calabrese, O., Baens, M., Folens, C., Cools, J., Marynen, P., Martelli, M.F., Mecucci,


302


304


and choroidal neovascularization by oral administration of a kinase inhibitor. 
*Am.J.Pathol.*, 154, 1743-1753.


activating mutation in FLT3 respond to a small-molecule FLT3 tyrosine kinase inhibitor, PKC412. *Blood*, **105**, 54-60.


PDGFRB fusion gene in an eosinophilia-associated atypical myeloproliferative neoplasm with a t(5;17)(q33-34;q11.2). *Genes Chromosomes Cancer, 48*, 179-183.


*Haematologica, 90*, 881-889.

*Blood, 83*, 2759-2779.


APPENDIX

I) Bubble-PCR primers

BUB-T:  
5’ AAGGATCCTAGTCTAGCTGTCTGTCGAAGGTAAGGAACGGGACGAGCACTGAG

BUB-B:  
5’ CTCAGTGCTCGTAGTAATCGTTCGCACGAGAATCGCAAGATCTAGGATCCTT

NVAMP1:  
5’ TGCTCGTAGTAATCGTTCGCAC

NVAMP2:  
5’ GTTCGCACGAGAATCGCAAGAT

PDAI12R3:  
5’ AGGTTACCCCATGGAACTACCA

PDAI12R4:  
5’ AAGTTTGTTGCAAGGGAAAGGG

II) PDGFRB Exons 2 to 23

PDBI1F3  5’ AGA GCC AAA GCA AAG GCC AG 3’
PDBI2R3  5’ AGA GCC CAC TGG AAA TCC TG 3’
PDBI2F   5’ CCC CTG TTT CCT GAT GTC TG 3’
PDBI3R   5’ ATG CCT CCA GTT GAC AAG AG 3’
PDBI3F   5’ CTG GCC ACA CAG CAA TTC AG 3’
PDBI4R   5’ CTG AGC ATC AGG CCA GAA AG 3’
PDBI4F   5’ GAC CTA AGC CAA TCT CTC TC 3’
PDBI6R   5’ GAA TTG GGG ATT GGG CTG AG 3’
PDBI6F   5’ TGG CCT CCT TTT GGA TTC AG 3’
PDBI7R2  5’ CCT AGG TTT GTG GCT GAA AG 3’
PDBI8F   5’ TAA CTG TCC TGA CCC TCC CG 3’
PDBI9R  5’ TCC TAG CCA GCT GGG GAC AC 3’
PDBI9F  5’ TCA GTT TCC CTG TCT GCA AG 3’
PDBI10R 5’ GTA GGG ATT GGG ATC GTC AG 3’
PDBI10F 5’ GAT GCC AAA GAT GGG GTG AG 3’
PDBI11R2 5’ ATG TGC CAG TCT TCA CCC AC 3’
PDBI11F 5’ CCT GAG ACT CCC TCT GAT AA
PDBI12R  5’ ACA TGA GGC CTC TCA GGA CT
PDBI12F  5’ CTG GTA GGC CAG GAG CTA AT
PDBI13R  5’ TGG AGG GCT CCA AGG ACT A
PDBI13F  5’ GGG GCA GAA GAG TCA GAA TA
PDBI14R  5’ CCT TGG TGG GCA CT T T
PDBI14F  5’ CCT TGA AGG GAC GCC TGA
PDBI15R  5’ AAA GAT TCA GTC CCT GGC CT
PDBI15F  5’ CTG CAG GCT CTC TCT GGA GT T
PDBI16R  5’ AGC CTG TTT GTA TGT GGG GT 3’
PDBI16F  5’ TAG CAG GTG ACC CTC TGT TGT T
PDBI18R  5’ AGG GCC CAG GGA AGG TA T
PDBI18F  5’ CAC TCA GCA CCT GTC CTG A 3’
PDBI19R  5’ GGC TGG AGG AGG AAG CAA 3’
PDBI19F  5’ CCC CTG AGC CTG TGT ATC T 3’
PDBI20R  5’ TCC CAA ATG CAT GAG ACT CCA 3’
PDBI20F  5’ AAG GGA ATA TCC AGA GAC AGG A 3’
PDBI21R  5’ TAA ATG CCA GCC CAT CAC GCA 3’
PDBI21F  5’ CGG TTA GAA GAT CCC TGA AG 3’
PDBI22R2 5’ TGT GCA CAA TTT CCT TGG CC 3’
PDBI22F  5’ ATG AGC CGG AGT GTG TGA AG 3’
PDBE23R  5’ GGG ACA GCT GAT AAG GGC AG 3’
### III) PCR and dHPLC conditions

<table>
<thead>
<tr>
<th>GENE</th>
<th>EXON</th>
<th>FORWARD PRIMER</th>
<th>REVERSE PRIMER</th>
<th>ANNEALING TEMPERATURE</th>
<th>DHPLC TEMPERATURE</th>
<th>TIME-SHIFT (SEC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDGFRB</td>
<td>2</td>
<td>PDBI1F3</td>
<td>PDBI2R3</td>
<td>62</td>
<td>63, 64</td>
<td>0, -0.5</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>PDBI2F</td>
<td>PDBI3R</td>
<td>60</td>
<td>61, 62, 63</td>
<td>-0.5, 0, 0.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>PDBI3F</td>
<td>PDBI4R</td>
<td>60</td>
<td>59, 60, 61</td>
<td>-0.5, 0, 0.5</td>
</tr>
<tr>
<td></td>
<td>5+6</td>
<td>PDBI4F</td>
<td>PDBI6R</td>
<td>60</td>
<td>60, 61, 63</td>
<td>-0.5, 0, 1</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>PDBI6F</td>
<td>PDBI7R2</td>
<td>62</td>
<td>64, 65</td>
<td>0.5, 1</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>PDBI7F</td>
<td>PDBI8R</td>
<td>60</td>
<td>62, 63, 64</td>
<td>-0.5, 0, 0.5</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>PDBI8F</td>
<td>PDBI9R</td>
<td>64</td>
<td>63, 64, 65</td>
<td>0, 0, 0.5</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>PDBI9F</td>
<td>PDBI10R</td>
<td>60</td>
<td>64, 65, 66</td>
<td>0, 0, 0.5, 1</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>PDBI10F</td>
<td>PDBI11R</td>
<td>62</td>
<td>61, 62, 63</td>
<td>0, -0.5, 0, 0.5</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>PDBI11F</td>
<td>PDBI12R</td>
<td>62</td>
<td>62, 63, 64</td>
<td>0, 0, 0.5, 1</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>PDBI12F</td>
<td>PDBI13R</td>
<td>62</td>
<td>60, 62, 63</td>
<td>-0.5, 0, 0.5</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>PDBI13F</td>
<td>PDBI14R</td>
<td>64</td>
<td>63, 64</td>
<td>-0.5, 0</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>PDBI14F</td>
<td>PDBI15R</td>
<td>60</td>
<td>62, 63, 64, 66</td>
<td>-0.5, 0, 0.5, 1</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>PDBI15F</td>
<td>PDBI16R</td>
<td>60</td>
<td>60, 61, 62, 63</td>
<td>-0.5, 0, 0.5, 1</td>
</tr>
<tr>
<td></td>
<td>17+18</td>
<td>PDBI16F</td>
<td>PDBI18R</td>
<td>62</td>
<td>60, 61, 62</td>
<td>-1, -0.5, 0</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>PDBI18F</td>
<td>PDBI19R</td>
<td>60</td>
<td>60, 62, 63</td>
<td>-1, 0, 0.5</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>PDBI19F</td>
<td>PDBI20R</td>
<td>62</td>
<td>60, 63, 64</td>
<td>-1.5, 0, 0.5</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>PDBI20F</td>
<td>PDBI21R</td>
<td>62</td>
<td>60, 61, 62</td>
<td>-1, -0.5, 0</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>PDBI21F</td>
<td>PDBI22R</td>
<td>58</td>
<td>61, 62, 63</td>
<td>-0.5, 0, 0.5</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>PDBI22F</td>
<td>PDBI23R</td>
<td>62</td>
<td>62, 63, 64</td>
<td>0, 0.5, 0</td>
</tr>
</tbody>
</table>

### IV) PDGFRB

**First step**
- PDGFRB exon 15.1R: 5’ agg tag tcc acc agg tct ccg ta 3’

**Nested**
- PDGFRB exon 14.1R: 5’ aag ggc ttg ctt ctc act gcg 3’

### V) Sequencing

**M13F**
- 5’ gta aaa cga cgg cca g 3’

**M13R**
- 5’ cag gaa aca gct atg ac 3’

### VI) Fusion
### GOLGA4 exon 8.1F
5’ act tat cac tca gtt gcg tga tg 3’

### PDGFRB exon 12.1R
5’ acc ttc cat cgg atc tgg taa cgt 3’

#### Reciprocal

### GOLGA4 exon 10.1R
5’ ttc ctt ttt aag act gat ggc tt 3’

### PDGFRB exon 9.1F
5’ aca tca tct ggt ctc cct gca g 3’

#### VII) PDGFRB

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB.I11.1R</td>
<td>5’ gag agc agg cca tga gca aac 3’</td>
</tr>
<tr>
<td>PDB.I11.1aR</td>
<td>5’ gag acc ccc agc ctg atg aat 3’</td>
</tr>
<tr>
<td>PDB.I11.2R</td>
<td>5’ cat gag gtc cta tgg aac tca tt 3’</td>
</tr>
<tr>
<td>PDB.I11.3R</td>
<td>5’ tcc cgc tca ggc ttt agt cac 3’</td>
</tr>
<tr>
<td>PDB.I11.4R</td>
<td>5’ ggc agc tct aga cag gca cg 3’</td>
</tr>
<tr>
<td>PDB.I11.5R</td>
<td>5’ aag gtg tgc tcc aat tat atg ctt 3’</td>
</tr>
<tr>
<td>PDB.I11.6R</td>
<td>5’ cag gca tag tgc aca ggt gta 3’</td>
</tr>
<tr>
<td>PDB.I11.7R</td>
<td>5’ tgc cag aat gtc cta tgc at 3’</td>
</tr>
</tbody>
</table>

#### VIII) GOLGA4

(Note: all primers are designed in intron 10, just labelled wrong)

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>gGOLGA4.I9.1F</td>
<td>5’ tct taa cta tag agg ctc cta ata a 3’</td>
</tr>
<tr>
<td>gGOLGA4.I9.2F</td>
<td>5’ gaa gac tga ctt gaa gca gtt tg 3’</td>
</tr>
<tr>
<td>gGOLGA4.I9.3F</td>
<td>5’ cag aac tgt tca agt agg aag tg 3’</td>
</tr>
<tr>
<td>gGOLGA4.I9.4F</td>
<td>5’ ctt tgg ctc act tat tct gga tgt 3’</td>
</tr>
<tr>
<td>gGOLGA4.I9.5F</td>
<td>5’ tct tat cat ggg tac tta gaa taa g 3’</td>
</tr>
<tr>
<td>gGOLGA4.I9.6F</td>
<td>5’ cca tgg atg tga tgt aac ctc a 3’</td>
</tr>
<tr>
<td>gGOLGA4.I9.7F</td>
<td>5’ caa tgg agc tgc atc cat ttc ag 3’</td>
</tr>
<tr>
<td>gGOLGA4.I9.8F</td>
<td>5’ tca atg gta gag gga agg ctg g 3’</td>
</tr>
<tr>
<td>gGOLGA4.I9.9F</td>
<td>5’ gca gaa aat tat atc taa aca gga ac 3’</td>
</tr>
<tr>
<td>gGOLGA4.I9.10F</td>
<td>5’ gtc cca agt tac aca cct gtt g 3’</td>
</tr>
<tr>
<td>gGOLGA4.I9.11F</td>
<td>5’ tta tac agt aca gaa ata ggg aca t 3’</td>
</tr>
</tbody>
</table>

#### IX) GOLGA4 RT-PCR

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>cGOLGA4.exon 10.1F</td>
<td>5’ cag atg act acc cag gga gag 3’</td>
</tr>
</tbody>
</table>
X) **GOLGA4 genomic fusion**

- gGOLGA4.I9.7F: 5’ caa tgg agc tcg atc cat ttc ag 3’
- PDB.I11.Brkpt.R: 5’ aac aga tgt ctg gga att agg ca 3’

XI) **Reciprocal PDGFRB**

- PDB.I11.recip.F: 5’ gtg agt ccc cag cct cag tg 3’
- PDB.I11.recip.2F: 5’ atg cat agg aac aca ttc tgg ca 3’

**Reciprocal GOLGA4**

- gGOLGA4.I9.recip.R: 5’ gat tcc tat tct aac ttt act g 3’

XII) **RT-PCR Fusion**

- cGOLGA4 exon 7.1F: 5’ tcc agc aaa gag tga agc gtc a 3’
- PDGFRB exon 12.1R: 5’ acc ttc cat cgg atc teg taa cgt 3’

XIII) **KIAA1449**

- KIAA1449.2F: 5’ cat acc gag tcc atg atg aag 3’
- PDGFRB.exon15.2R: 5’ tgc tgc agg aag gtg tgt ttg ttg 3’

XIV) **Bin2-PDGFRB construct primers**

**Full-length fusion amplification**

- Bin2-PDB.Fus.1F: 5’ agt tgg cag gat ggc aga ggg 3’
- Bin2-PDB.Fus.2F: 5’ gtt ggc agg atg gca gag g 3’
- Bin2-PDB.Fus.2R: 5’ acg gcc cct gca gtt ttc tt 3’

**Overlapping primers for sequencing**

- M13F & M13R: (as previous – section V)
- Bin.e5.1F: 5’ ttg gga aga cta cga gga 3’
- Bin.e9.1F: 5’ cta cga ggt gat gag caa g 3’
- PDB.e15.2F: 5’ caa caa aca cac ctt cct gca gca 3’
- PDB.e18.1F: 5’ att aca tct cca aag gca gca cct 3’
- PDB.e15.2R: 5’ tgc tgc agg aag gtg tgt ttg ttg 3’

**pcDNA3.1(+) vector primers**
XV) **GOLGA4-PDGFRB construct primers**

**Fragment 1 construct amplification (GOLGA4 only)**  
*NheI.1F* 5’ tac gct agc cca tgt tca aga aac tg 3’  
cGOLGA4.e10.3R 5’ cct ctc cct ggg tag tca tct g 3’

**Overlapping primers for sequencing fragment 1**  
M13F & M13R (as previous – section V)  
BC/GOL.3F 5’ tgtg aac gaa gct taa gta gct gc t3’  
BC/GOL.2F 5’ gca cag aag ctc cag ctc cg 3’  
cGOLGA4.e6.2F 5’ tgc atc ttt aga gga gaa aga tca 3’

**Fragment 2 construct amplification (GOLGA4 only)**  
cGOLGA4.e7.1F 5’ tcc agc aaa gag tga agc gtc a 3’  
PDGFRB.e15.1R (as previous – section IV)  
cGOLGA4.e8.1F 5’ tta tca ctc atg tgc gtg atg ca 3’

XVI) **STRN-PDGFRA fusion primers**

**Fusion**  
STRN.Fusion.Ex6.1F 5’ tct gag tta aca gat tct gcc tc 3’  
PDGFR.A.Fusion.Ex12.2R 5’ caa gca cta gtc cat ctc ttg g 3’

**Reciprocal**  
PDGFR.A.recip.Ex11.1F 5’ acg gtg gct gct gca gtc ct 3’  
STRN.Recip.Int6.1R 5’ act atc ttc atc tgc ccc ata gaa 3’

XVII) **ETV6-PDGFRB fusion primers**

**Genomic DNA Fusion**  
ETV6.Fusion.Int6.4F 5’ ctt cta aga agg cag tgg gta at 3’  
PDGFRB.A.fusion.Ex12.1R 5’ caa gca ggg atc ata cct cgg 3’

**mRNA Fusion**  
ETV6.Fusion.Ex6.1F: 5’ act gta gac tge ttt ggg att ac 3’  
PDGFRB.A.Fusion.Ex12.2R 5’ caa gca cta gtc cat ctc ttg g 3’

**Reciprocal**
XVIII) **STRN-PDGFRA & ETV6-PDGFRA MRD primers**

**STRN Fusion – SINGLE STEP**
- STRN.Fusion.Ex6.1F 5’ tct gag tta aca gat tct gcc tc 3’
- PDGFRA.Fusion.Ex12.2R 5’ caa gca cta gtc cat ctc ttg g 3’

**STRN Fusion – NEST STEP**
- STRN.Ex.6.2F 5’ tct gag tta aca gat tct gcc tc 3’
- PDGFRA.STRN.Ex12.1R 5’ cac tag tcc atc ttc tgg aaa ct 3’

**ETV6 Fusion – SINGLE STEP**
- ETV6.Fusion.Ex6.1F: 5’ act gta gac tgc ttt ggg att ac 3’
- PDGFRA.Fusion.Ex12.2R 5’ caa gca cta gtc cat ctc ttg g 3’

**ETV6 Fusion – NEST STEP**
- ETV6.Ex6.2F 5’att ccg gat agt gga tcc caa c 3’
- PDGFRA.STRN.Ex12.1R 5’ cac tag tcc atc ttc tgg aaa ct 3’

XIX) **MLPA cDNA oligonucleotides**

Labelled primer
Target sequence X
Target sequence Y
Stuffer Sequence
Other Primer

1. **FGFR1 intracellular (exons 12/13) (95bp)**

A:  
GGG TTC CCT AAG GGT TGG AGC ATG GAG TAT CTG GCC TCC  
AAG AAG

B:  
TGC ATA CAC CGA GAC CTG GCA GCC CTG TCT AGA TTG GAT  
CTT GCT GGC AC
2. FGFR1 extracellular (exons 3/4) (100bp)

A: GGG TTC CCT AAG GGT TGG AGA GCC ATA ACA CCA AAC CAA ACC GTA TGC

B: CCG TAG CTC CAT ATT GGA CAT CCC GAA TCT CTA GAT TGG ATC TTG CTG GCA C

3. PDGFRB intracellular (exons 20/21) (105bp)

A: GGG TTC CCT AAG GGT TGG AGT ACT AC G CCT GCC CAT GCC TCC GAC GAG AT

B: CTA TGA GAT CAT GCA GAA GTG CTG GCA CTA TGT CTA GAT TGG ATC TTG CTG GCA C

4. PDGFRB extracellular (exons 4/5) (110bp)

A: GGG TTC CCT AAG GGT TGG ACT GAC GTA CG C CTA CTA TGT CTA CAG ACT CCA GG

B: TGT CAT CCA TCA ACG TCT CTG TGA CCT AGC TGA GTC TAG ATT GGA TCT TGC TGG CAC

5. PDGFRA intracellular (exons 20/21) (115bp)

A: GGG TTC CCT AAG GGT TGG AGT ATG ACT GTA CG C CCT GCC CAT GCC CAC GCT ACC AGT GAA GT

B: CTA CGA GAT CAT GGT GAA ATG CTG CTG ACA TAG CAG TCT AGA TTG GAT CTT GCT GGC AC

6. PDGFRA extracellular (exons 2/3) (120bp)

A: GGG TTC CCT AAG GGT TGG ACA CTC TAG GTA CGT G TG GTC TTA GGC TGT CTT CTC ACA G

B: GGC TGA GCC TAA TCC TCT GCC AGC GTC AGA GAC CGA TAC TCT AGA TTG GAT CTT GCT GGC AC
7. ABL C-terminal domain (exons 8/9)  (125bp)

A:  GGG TTC CCT AAG GGT TGG AGC TGA ATC AGA CCT ACC AGG 
    TCT ATG AAC TCA TGC GAG CAT

B:  GTT GGC AGT GGA ATC CCT CTG ACC CAG TCA ACA GGC TCG 
    ATC TCT AGA TTG GAT CTT GCT GGC AC

8. ABL 1b start (exons 1b/a2)  (130bp)

A:  GGG TTC CCT AAG GGT TGG AGC TAC CGC TAC AAG TTC TAG 
    GGT CCA CAC TGC AAT GTT TTT GTG

B:  GAA CAT GAA GCC CTT CAG CGG CCA CTA GAT CCA TTA TCG 
    TAT ACT CTA GAT TGG ATC TTG CTG GCA C

XX)  

Cryptic ETV6-PDGFRB fusion primers

Fusion
ETV6.Ex6.2F  5’ATT CCG GAT AGT GGA TCC CAA C 3’
PDGFRB.Ex14R  5’- CCC CAA CA G GTT GAC CAC GTT CAG -3’

Genomic Fusion
ETV6.E1026.Int7.1F  5’ AGG ATA GGA TAG GAT AGA GTT 
    GAA -3’
ETV6.E1026.Int7.2F  5’ AAT GCC AGA CTG CAT GAC ATG -3’
PDGFRB.Exon 12R  5’- CCA CGT AGA TGT ACT CAT GGC -3’

XXI)  

Primer sequences designed to amplify the kinase domain of the RTKs EPHA1, EPHB6 & Met

EPHA1.Ex12F  5’ tgg aag gcg tgc tca caa age 3’
EPHA1.Ex16R  5’ cca tcc tca atg ctc ttc ata ac 3’
EPHB6.Ex11F  5’ gat cct gct tat atc aag att gag  3’
XXII) **EPHA1 genomic DNA primers to screen the whole gene**

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPHA1.5'UTRF</td>
<td>gacgcctcccggtcttgaatttgag</td>
</tr>
<tr>
<td>EPHA1.i1R</td>
<td>aacatggecctaggaatg</td>
</tr>
<tr>
<td>EPHA1.i1F</td>
<td>tttttgtctttaaccccaagcg</td>
</tr>
<tr>
<td>EPHA1.i2R</td>
<td>cttgtcttccacccacctct</td>
</tr>
<tr>
<td>EPHA1.i2F</td>
<td>gggatgatcgtgctcctgaactct</td>
</tr>
<tr>
<td>EPHA1.i3R</td>
<td>cttctctggaccccaatagga</td>
</tr>
<tr>
<td>EPHA1.i3F</td>
<td>agaagcggaccttgccatggt</td>
</tr>
<tr>
<td>EPHA1.i4R</td>
<td>gtactctgcaatctccccgagt</td>
</tr>
<tr>
<td>EPHA1.i4F</td>
<td>tgggtggtgtctgctgctgtaaa</td>
</tr>
<tr>
<td>EPHA1.i5R</td>
<td>ccacccctgacaaatgctccagca</td>
</tr>
<tr>
<td>EPHA1.i5F</td>
<td>gaggacatgagaaagttggtgtggt</td>
</tr>
<tr>
<td>EPHA1.i6R</td>
<td>tatgctctctgccatcagt</td>
</tr>
<tr>
<td>EPHA1.i6F</td>
<td>gatgggcaaggaacccataagaaa</td>
</tr>
<tr>
<td>EPHA1.i7R</td>
<td>ggcctggaactgaagacccatct</td>
</tr>
<tr>
<td>EPHA1.i7F</td>
<td>ggctcagcactgccacattaat</td>
</tr>
<tr>
<td>EPHA1.i8R</td>
<td>aggtgtatgtgtgtgtggcagag</td>
</tr>
<tr>
<td>EPHA1.i8F</td>
<td>gatgggcaggaagccatagaaaa</td>
</tr>
<tr>
<td>EPHA1.i9R</td>
<td>ggcattgacactgaagacccatct</td>
</tr>
<tr>
<td>EPHA1.Int8.1F</td>
<td>ccctagggactgtgggttcaccaca</td>
</tr>
<tr>
<td>EPHA1.Int9.1R</td>
<td>tactaggggattgagggtgctgttt</td>
</tr>
<tr>
<td>EPHA1.Int9.1F</td>
<td>tctggggaaagatcgcagccagag</td>
</tr>
<tr>
<td>EPHA1.Int10.1R</td>
<td>gatccctccctcaggggtgaatcctc</td>
</tr>
<tr>
<td>EPHA1.Int10.1F</td>
<td>cctgtagtgggtgctcctccacc</td>
</tr>
<tr>
<td>EPHA1.Int11.1R</td>
<td>ctgcttgggtggccttctctg</td>
</tr>
<tr>
<td>EPHA1.Int11.1R</td>
<td>gggccacgcgctggtttttcttc</td>
</tr>
<tr>
<td>EPHA1.Int12.1R</td>
<td>agctgagggagaccactcagctg</td>
</tr>
<tr>
<td>EPHA1.Int12.1F</td>
<td>gaaagggggtggagacccactgga</td>
</tr>
<tr>
<td>EPHA1.Int13.1R</td>
<td>cagctagtccccttgcccctcctca</td>
</tr>
<tr>
<td>EPHA1.Int13.1F</td>
<td>cccctcctgactccaccact</td>
</tr>
<tr>
<td>EPHA1.Int14.1R</td>
<td>cagtttacgggcaatggtaatg</td>
</tr>
<tr>
<td>EPHA1.Int14.1F</td>
<td>cccctgaccatgcttcttag</td>
</tr>
<tr>
<td>Primer name</td>
<td>Primer sequence</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>EPHA1.Int15.1R</td>
<td>gggaggacctgggaggaataa</td>
</tr>
<tr>
<td>EPHA1.Int15.1F</td>
<td>ttctcatgggcctccagttgtag</td>
</tr>
<tr>
<td>EPHA1.Int16.1R</td>
<td>ctgctctgtgtccctgtagtct</td>
</tr>
<tr>
<td>EPHA1.Int16.1F</td>
<td>gcacaccagatcgacccatga</td>
</tr>
<tr>
<td>EPHA1.Int17.1R</td>
<td>cccagacagtcagctctttac</td>
</tr>
<tr>
<td>EPHA1.Ex17.1F</td>
<td>caccatggagtgtgtgtgaga</td>
</tr>
<tr>
<td>EPHA1.Ex 18.1R</td>
<td>agggtctctgaaagtggcc</td>
</tr>
<tr>
<td>EPHB6.Int5'UTR.F</td>
<td>aaggtcgacagaccttttaacca</td>
</tr>
<tr>
<td>EPHB6.Int.1R</td>
<td>gcctactcctgttgtgacctgt</td>
</tr>
<tr>
<td>EPHB6.Int.1F</td>
<td>cctgcctcttcttgttgtggtgt</td>
</tr>
<tr>
<td>EPHB6.Int.2R</td>
<td>actcctcctctaactgggtctc</td>
</tr>
<tr>
<td>EPHB6.Int.2F</td>
<td>gcacacttgtgcccccttattt</td>
</tr>
<tr>
<td>EPHB6.Int3R</td>
<td>aacctcaccagggtctcttc</td>
</tr>
<tr>
<td>EPHB6.Int.3F</td>
<td>ctaaagctctcctggcccttcc</td>
</tr>
<tr>
<td>EPHB6.Int4R</td>
<td>caattggtcacccttgtgtgtcag</td>
</tr>
<tr>
<td>EPHB6.Int.4F</td>
<td>ccgggagaccttcacctttact</td>
</tr>
<tr>
<td>EPHB6.Int.5R</td>
<td>aagtgatacctgctgtgagagtc</td>
</tr>
<tr>
<td>EPHB6.Int5F</td>
<td>cgaggatacctgctgtgtgagagtc</td>
</tr>
<tr>
<td>EPHB6.Int6R</td>
<td>acttctgcccctctcaagctct</td>
</tr>
<tr>
<td>EPHB6.Int6F</td>
<td>ggtctctctctctcaatatec</td>
</tr>
<tr>
<td>EPHB6.Int.7R</td>
<td>ccaggtctggagatcaggagagttgag</td>
</tr>
<tr>
<td>EPHB6.Int.7F</td>
<td>atctctgctcagggactctat</td>
</tr>
<tr>
<td>EPHB6.Int.8R</td>
<td>tgaggctctgagtctctcactcag</td>
</tr>
<tr>
<td>EPHB6.Int.7.1F</td>
<td>cagggttgtgaggtgttttaccc</td>
</tr>
<tr>
<td>EPHB6.Int 8.1R</td>
<td>actacccacctggggctcactc</td>
</tr>
<tr>
<td>EPHB6.Int 8.1F</td>
<td>gcacatggacaagcaacctgtgag</td>
</tr>
<tr>
<td>EPHB6.Int 9.1R</td>
<td>actgggcagaggctgtttgag</td>
</tr>
<tr>
<td>EPHB6.Int 9.1F</td>
<td>gtcatccccagtcaccgttttgt</td>
</tr>
<tr>
<td>EPHB6.Int 10.1R</td>
<td>gcccacacttccacagaaagt</td>
</tr>
<tr>
<td>EPHB6.Int10.1F</td>
<td>ctgctcttacaacacagcctcataa</td>
</tr>
<tr>
<td>EPHB6.Int11.1R</td>
<td>cctgcgtggtttgattgagggact</td>
</tr>
<tr>
<td>EPHB6.Int11.1F</td>
<td>agggecaaaagccagggatacag</td>
</tr>
<tr>
<td>EPHB6.Int12.1R</td>
<td>cctgcgtcctgaaggtcttccc</td>
</tr>
<tr>
<td>EPHB6.Int12.1F</td>
<td>cctgcgtcctgaaggtcttccc</td>
</tr>
<tr>
<td>EPHB6.Int13.1R</td>
<td>gttcccccaaggtgctgctct</td>
</tr>
<tr>
<td>EPHB6.Int13.1F</td>
<td>gttaaaacctttggcatatatgacac</td>
</tr>
<tr>
<td>EPHB6.Int14.1R</td>
<td>gttgggaaaatcagcaggtgctcag</td>
</tr>
<tr>
<td>EPHB6.Int14.1F</td>
<td>ageccacacttcataccaataaa</td>
</tr>
<tr>
<td>Primername</td>
<td>Primersequence</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>MET.5UTR.F</td>
<td>tctccggetgtgctaaacctcagac</td>
</tr>
<tr>
<td>MET.Int1.1R</td>
<td>gtctctccttcgaagccctctta</td>
</tr>
<tr>
<td>MET.Int1.1F</td>
<td>tccagttggaagccattttctga</td>
</tr>
<tr>
<td>MET.Int2.2R</td>
<td>tgtgaatgtaaacagactcgaaha</td>
</tr>
<tr>
<td>MET.Int2.1F</td>
<td>gcttattacctccagctctgaaga</td>
</tr>
<tr>
<td>MET.Int3.1R</td>
<td>ggttttcttgccttttctgcctag</td>
</tr>
<tr>
<td>MET.Int3.1F</td>
<td>ggagggaacttgggtcttcagttg</td>
</tr>
<tr>
<td>MET.Int4.1R</td>
<td>aagggcacaagacactcactgtgta</td>
</tr>
<tr>
<td>MET.Int4.1F</td>
<td>ccttcgagaaacggtgctgaagctc</td>
</tr>
<tr>
<td>MET.Int5.1R</td>
<td>cagtgtagcagctgttagactcc</td>
</tr>
<tr>
<td>MET.Int5.1F</td>
<td>tgtcccatctttggtactcctgaa</td>
</tr>
<tr>
<td>MET.Int6.1R</td>
<td>tgtgacaagccagaccccttttta</td>
</tr>
<tr>
<td>MET.Int6.1F</td>
<td>tgttcttacacacacactatcag</td>
</tr>
<tr>
<td>MET.Int7.1R</td>
<td>tgtgagggagatatagagactcag</td>
</tr>
<tr>
<td>MET.Int7.1F</td>
<td>ttctccacacacgacacaaaaaaca</td>
</tr>
<tr>
<td>MET.Int8.1R</td>
<td>gtctccacacagagactcactgta</td>
</tr>
<tr>
<td>MET.Int8.1F</td>
<td>ccttcgagaaacggtgctgaagctc</td>
</tr>
<tr>
<td>MET.Int9.1R</td>
<td>ttggaaaaggaacctctcagagaca</td>
</tr>
<tr>
<td>MET.Int9.2F</td>
<td>tgtccctcctgctctaccacactg</td>
</tr>
<tr>
<td>MET.Int10.2R</td>
<td>tggaggggcttcagagctcagagag</td>
</tr>
<tr>
<td>MET.Int10.1F</td>
<td>tgtctagagcagactcactgta</td>
</tr>
<tr>
<td>MET.Int11.1R</td>
<td>tgtgagggagatagagactcagagag</td>
</tr>
<tr>
<td>MET.Int11.1F</td>
<td>acaacatggcctgttgtactctgta</td>
</tr>
<tr>
<td>MET.Int12.1R</td>
<td>aaggtgacttgctagcttgctgta</td>
</tr>
<tr>
<td>MET.Int.12.2F</td>
<td>tccctgagggccttgttgtctgta</td>
</tr>
<tr>
<td>MET.Int13.2R</td>
<td>cctaggcctccttgctccatcctc</td>
</tr>
<tr>
<td>MET.Int13.2F</td>
<td>attgtacctgttagctgtttctgta</td>
</tr>
<tr>
<td>MET.Int14.2R</td>
<td>tccgctttacttacatcctctgta</td>
</tr>
<tr>
<td>MET.Int.14.1F</td>
<td>gctctccctgttgtactctgta</td>
</tr>
<tr>
<td>MET.Int15.1R</td>
<td>tgtgagggagatagagactcagagag</td>
</tr>
<tr>
<td>MET.Int15.1F</td>
<td>ccgtagtgctacacagctctctctt</td>
</tr>
<tr>
<td>MET.Int16.1R</td>
<td>tgatctttccacaaggggaaggggtg</td>
</tr>
<tr>
<td>MET.Int16.1F</td>
<td>ctgcagtcacaaccttcaggacag</td>
</tr>
<tr>
<td>MET.Int17.1R</td>
<td>gacagagctatttgaaggggtg</td>
</tr>
<tr>
<td>MET.Int17.1F</td>
<td>caggcttgaccaattttacccaa</td>
</tr>
<tr>
<td>MET.Int18.1R</td>
<td>tggatgacagagctacaggata</td>
</tr>
<tr>
<td>MET.Int18.1F</td>
<td>aggccagatgaaataacttcctccag</td>
</tr>
<tr>
<td>MET.Int19.1R</td>
<td>tgaagaaaaactggagattggtggt</td>
</tr>
<tr>
<td>MET.Int19.1F</td>
<td>tggteacatctctcactctagctg</td>
</tr>
<tr>
<td>MET.Int20.1R</td>
<td>ccttgaagggcatctctag</td>
</tr>
<tr>
<td>MET.Int20.1F</td>
<td>gecctgctcagaggtctcttc</td>
</tr>
<tr>
<td>MET.3UTR.1R</td>
<td>ccaatagacaggtgatgctcag</td>
</tr>
</tbody>
</table>

**XXIII) PIK3CG primer sequences**

- **PIK3CG.Codons.542.5.1F**: 5’ taa gca tca ggc cac ecc tga 3’
- **PIK3CG.Codons 542.5.1R**: 5’ ctt gct gtc ccc att tca ctg a 3’
- **PIK3CG.Codon 1007.2F**: 5’ ggc tta tct agc cct tcg tca 3’
- **PIK3CG.Codon.1007.2R**: 5’ tgc aca gt c gat cct ttg tgg tct g 3’
- **PIK3CG.Codon.1047.1F**: 5’ cct act gat cat cct gtt ctg 3’
- **PIK3CG.Codon.1047.1R**: 5’ gca ttt gaa atc caa gtt cga tg 3’

**XXIV) c-CBL RING finger domain primers**

- **CBL.Int7.1F**: 5’ tgt ggt ttc act tta aac cct gga 3’
- **CBL.Int8.1R**: 5’ gc cag gcc acc cct tgt atc 3’
- **CBL.Int8.1F**: 5’ ggc ctg gct ttt ggg tta gg 3’
- **CBL.Int9.1R**: 5’ ca caa tgg att tgt cca gtc tcc 3’

**XXV) c-CBL cDNA primers**

- **CBL.Ex6/7.F**: 5’ ggc tc agg gaa ggc ttc ta 3’
- **CBL.Ex10/11.R**: 5’ gcc aga age age ctt aca age a 3’
XXVI)  **NRAS primers**

2  
  f: 5’gat gtg gct cgc caa tta ac 3’
  r: 5’tgc ata act gaa tgt ata ccc aaa 3’

3  
  f: 5’ttg cat tcc ctg tgg ttt tt 3’
  r: 5’cct cat ttc ccc ata aag att c 3’

4  
  f: 5’cct ccc aaa gtg ctg aga tt 3’
  r: 5’gca aac tct tgc aca aat gc 3’

5  
  f: 5’tgg aat ctt atg tcc cta aca taa aga 3’
  r: 5’tca gaa agg tgt tca tat gga 3’

These primers cover the entire coding sequence of the NRAS gene, however only primers to exons 2 & 3 were used to look at the mutation hotspots at codons 12, 13 and 61.

XXVII)  **c-CBL intronic primers to amplify exons 8 & 9**

CBL.i7F 5’ tgt ggt ttc act tta aac cct gga 3’

CBL.i9R 5’ca caa tgg att ttg cca gtc tcc 3’
List of Publications

Two novel imatinib-responsive PDGFRA fusion genes in chronic eosinophilic leukaemia.

A novel ETV6-PDGFRB fusion transcript missed by standard screening in a patient with an imatinib responsive chronic myeloproliferative disease.
Curtis CE, Grand FH, Waghorn K, Sahoo TP, George J, Cross NC.

The t(1;9)(p34;q34) and t(8;12)(p11;q15) fuse pre-mRNA processing proteins SFPQ (PSF) and CPSF6 to ABL and FGFR1.
Hidalgo-Curtis C, Chase A, Drachenberg M, Roberts MW, Finkelstein JZ, Mould S, Oscier D, Cross NC, Grand FH.

Frequent CBL mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms.
Blood. 2009 Apr 22. [Epub ahead of print]

Identification of a novel imatinib responsive KIF5B-PDGFRA fusion gene following screening for PDGFRA overexpression in patients with hypereosinophilia.
Score J, Curtis C, Waghorn K, Stalder M, Jotterand M, Grand FH, Cross NC.

Durable responses to imatinib in patients with PDGFRB fusion gene-positive and BCR-ABL-negative chronic myeloproliferative disorders.


Transient response to imatinib in a chronic eosinophilic leukemia associated with ins(9;4)(q33;q12q25) and a CDK5RAP2-PDGFRA fusion gene.

Detection and molecular monitoring of FIP1L1-PDGFRA-positive disease by analysis of patient-specific genomic DNA fusion junctions.

Imatinib sensitivity as a consequence of a CSF1R-Y571D mutation and CSF1/CSF1R signaling abnormalities in the cell line GDM1.

Widespread occurrence of the JAK2 V617F mutation in chronic myeloproliferative disorders.
Minimal molecular response in polycythemia vera patients treated with imatinib or interferon alpha.


Two novel imatinib-responsive PDGFRA fusion genes in chronic eosinophilic leukaemia

Claire E. Curtis,1 Francis H. Grand,1 Pellegrino Musto,1 Andrew Clark,3 John Murphy,2 Gianni Perla,4 Maria M. Minervini,4 Janet Stewart,5 Andreas Reiter5 and Nicholas C. P. Cross1
1Wessex Regional Genetics Laboratory, University of Southampton, Salisbury, UK, 2Haematology and Stem Cell Transplantation Unit, CROB – Centro di Riferimento Oncologico della Basilicata, Rionero in Vulture, Italy, 3Monklands Hospital, Airdrie, UK, 4Unit of Haematology, IRCCS ‘Casa Sollievo della Sofferenza’, S. Giovanni Rotondo, Italy, 5Institute of Medical Genetics, Yorkhill Hospital, Glasgow, UK, and 6III. Medizinische Universitätsklinik, Medizinische Fakultät Mannheim der Universität Heidelberg, Heidelberg, Germany

Received 16 February 2007; accepted for publication 26 March 2007
Correspondence: Prof. N.C.P. Cross, Wessex Regional Genetics Laboratory, University of Southampton, Salisbury District Hospital, Salisbury SP2 8BJ, UK.
E-mail: ncpc@soton.ac.uk

Summary
We identified two patients with a t(2;4)(p24;q12) and a t(4;12)(q23;p12), respectively, in association with BCR-ABL and FIP1L1-PDGFRA negative chronic eosinophilic leukaemia. Molecular analysis revealed a novel STRN-PDGFRA fusion for the t(2;4) and ETV6-PDGFRA for the t(4;12). The fusions were confirmed by specific amplification of the genomic breakpoints, reverse transcription polymerase chain reaction and fluorescence in situ hybridisation. Both patients were treated with imatinib and, following a rapid haematological response, achieved cytogenetic remission and a major molecular response. In conclusion, PDGFRA fuses to diverse partner genes in myeloid disorders. Identification of these fusions is important as they are particularly sensitive to imatinib.

Keywords: PDGFRA, ETV6, STRN, chronic eosinophilic leukaemia, imatinib.

Patients and methods

Patients
Case 1 had CEL and has been described in detail elsewhere (Musto et al, 2004). Briefly, a 64-year-old male presented with an eosinophil count of $12 \times 10^9/l$ and a bone marrow indicative of a myeloproliferative disorder (MPD). Cytogenetics revealed a t(2;4)(p24;q12) in 60% of bone marrow metaphases. Haematological and cytogenetic remission was achieved following treatment with 100 mg imatinib daily. This dose was continued for 9 months and then progressively reduced to 100 mg on alternate days followed by 100 mg twice a week. Following 24 months of imatinib, he refused further therapy despite not having any significant side effects. Elevated eosinophil counts were again detected 14 months after discontinuation of treatment.
Case 2 was a 51-year-old male who presented with late onset asthma and type 2 diabetes. He had an eosinophil count of 12.9 × 10^9/L and a hypercellular bone marrow suggestive of an MPD. Cytogenetics revealed a t(4;12)(q213;p112) in 100% of bone marrow metaphases. The patient was managed with hydroxyurea for over 7 years but with recurrent problems due to anaemia, thrombocytopenia and very poor control of the eosinophil count. Cytogenetic analysis at 6 years after diagnosis showed 46,XY,t(4;12)(q213;p112)[14]/46,XY[6]. Following the molecular diagnosis detailed below, hydroxyurea was stopped and imatinib started at 100 mg daily. The full blood counts completely normalised within 4 weeks, resulting in significant resolution of symptoms (apart from the asthma) and greatly improved quality of life. Cytogenetic analysis was performed 9 months later and indicated complete cytogenetic remission (30 metaphases).

**Polymerase chain reaction (PCR) and reverse transcription polymerase chain reaction (RT-PCR)**

Conditions for amplifying PDGFRA rearrangements by bubble PCR have been described in detail elsewhere (Score et al, 2006). Primers used to detect STRN-PDGFR DNA were STRN.Fusion.Ex6.1F (5'-TCTGAGTTAACAGATCTGCTC-CTC3') and PDGFR.A.Fusion.Ex12.2R (5'-CAAGCACTAGTCCATCCTTGG3') Primers for ETV6-PDGFRA mRNA from random hexamer reverse transcribed cDNA were ETV6.Fusion.n.Ex6.1F: 5'-ACTGTAGACTGCTTTGGGATTAC3' and PDGFR.A.Fusion.Ex12.2R. Primer sequences for analysis of reciprocal breakpoints and minimal residual disease (MRD) are available on request.

**Fluorescence in situ hybridisation (FISH)**

Fluorescence in situ hybridisation (FISH) was performed using bacterial artificial chromosome (BAC) clones selected from sequences flanking PDGFRA and ETV6 (http://www.ensembl.org) or commercially available probe sets. BAC DNA was grown, extracted and labelled by nick translation with digoxigenin or biotin-16-deoxy-uridine-5'-triphosphate (Roche, East Sussex, UK) and hybridised using standard procedures. Case 2 was analysed using BACs RP11-24O10 (3' of PDGFRA); RP11-434C1 (5' of ETV6).

**Results**

**Bubble PCR to identify PDGFRA rearrangements**

Both cases were negative for BCR-ABL and FIP1L1-PDGFR by RT-PCR or FISH analysis using standard procedures. The presence of a 4q translocation in conjunction with persistent eosinophilia suggested that PDGFRA might be involved. As the genomic breakpoints for all cases involving this gene thus far reported are tightly clustered within or close to PDGFRA exon 12 (Reiter et al, 2007) we investigated the possibility that PDGFRA was rearranged, by bubble PCR. As shown in Fig 1A, rearranged bands were detected for both cases 1 and 2.

**Identification and confirmation of the STRN-PDGFRa fusion**

Sequencing of the rearranged allele for case 1 identified a novel fusion between STRN (STRN; NM_003162) intron 6 and a truncated PDGFRA exon 12. To confirm the fusion, a 285 bp product was amplified from patient genomic DNA but not from control DNA. A reciprocal PDGFRASTRN fusion product was not detected with various primer combinations using either single step or nested PCR. Identification of the mRNA fusion was not possible due to lack of adequate patient material, however analogous PDGFRA breakpoints for FIP1L1-PDGFRa usually result in the use of cryptic splice sites within PDGFRA exon 12 downstream of the breakpoint. Two potential AG splice acceptor sites would produce in-frame mRNA fusions with STRN exon 6 and it is likely that either or both of these might represent the real mRNA fusion in case 1 (Fig 1B).

**Identification and confirmation of the ETV6-PDGFRa fusion**

Sequencing of the rearranged bubble PCR band from case 2 identified a second novel fusion between ets-variant gene 6 (ETV6; NM_001987) intron 6 and PDGFRA intron 11. To confirm the fusion a 210 bp product was specifically amplified from case 2 peripheral blood genomic DNA but not from control DNA using breakpoint-specific primers. A reciprocal PDGFRa-ETV6 fusion product was not detected with various primer combinations with either single or nested PCR. Expression of ETV6-PDGFRa was confirmed by single-step PCR from case 2, yielding a 276 bp product that was not observed in normal control cDNA. Sequencing of the product revealed an in-frame whole exon fusion between ETV6 exon 6 and PDGFRA exon 12 (Fig 2A). To further confirm the presence of the ETV6-PDGFRa fusion, two-colour interphase FISH analysis was performed. On normal cells BACs RP11-24O10 and RP11-434C1 gave four separate signals corresponding to the normal chromosomes 4 and 12. However, in 56% of cells for case 2, overlapping probe signals consistent with co-localisation of ETV6 and PDGFRA were observed (Fig 2B).

**Response to imatinib**

For cases 1 and 2, peripheral blood was analysed approximately 28 and 9 months after starting imatinib respectively. MRD was not detected using cDNA or genomic DNA for case 1, indicating the achievement of molecular remission. For case 2, analysis was only performed using genomic DNA. Of four replicate reactions, two were positive indicating low-level disease consistent with a major molecular response.
Discussion

This study has characterised the genomic breakpoints in two CEL cases with chromosome 4q rearrangements and identified STRN at 2p22 and ETV6 at 12p13 as novel PDGFRA fusion partners. This brings the number of known PDGFRA partner genes to six, with PDGFRA being the tenth tyrosine kinase reported to fuse to ETV6. As these fusions were referred specifically for investigation of 4q rearrangements and were not part of a series that underwent systematic cytogenetic investigation, it is not possible to accurately determine the frequency of these abnormalities. However, we estimate that they probably account for <1% of CEL cases.

STRN encodes striatin, a 780 amino acid member of the calmodulin-binding WD repeat family of proteins. It has four protein–protein interaction domains, a caveolin binding domain (55–63aa), a coiled-coil domain (70–116aa), a calcium dependent calmodulin-binding domain (149–166aa) and a series of WD repeat motifs (419–780aa) (Bartoli et al., 1998). Striatin is thought to be a scaffolding protein that targets STRN exon 6 ΔPDGFRA exon 12

Fig 1. Characterisation of STRN-PDGFRA. (A) Left panel: Bubble PCR strategy showing PCR primers and restriction sites in the region where all breakpoints have been reported to fall; right panel: Nested bubble PCR gel after digestion with HaeIII and Rsal. The normal controls show the expected germline bands of 1011 and 648 bp respectively (the distance between primer PDAI12-R4 and the relevant restriction site plus 33 bp contributed by the bubble oligo). Smaller bands are visible for cases 1 and 2 after HaeIII digestion and for case 2 after Rsal digestion derived from the rearranged allele. The normal allele is not always visible due to out competition in the amplification reaction. (B) Electropherogram of the genomic DNA junction amplified by bubble PCR, the two predicted in frame STRN-PDGFRA mRNA fusions and gel showing amplification of STRN-PDGFRA but not the reciprocal genomic DNA fusion specifically from case 1. (C) Structure of Pdgfrα, Striatin and the predicted fusion protein. Relevant domains are shown: transmembrane (tm), WW-like (WW) and tyrosine kinase domains for Pdgfrα; coiled-coiled region and WD40 repeats for striatin; partially deleted WW domain for striatin-Pdgfrα (ΔWW).
proteins, such as the oestrogen receptor to the cell membrane, facilitating the assembly of other proteins required for rapid receptor activation (Lu et al., 2004). In case 1, we identified a DNA fusion of STRN intron 6 to a truncated PDGFRA exon 12 and thus the large region of WD repeat motifs within striatin are not retained in the fusion. However, the N-terminal coiled-coil domain remains that may act as an oligomerisation motif, constitutively activating the Pdgfrα-derived tyrosine kinase moiety (Fig 1C). Alternatively, truncation of the WW-like domain may be the critical activating event (Stover et al., 2006).

Etv6 (ets translocation variant 6) is a member of the ets family of transcription factors and is frequently rearranged in both myeloid and lymphoid malignancies (Bohlander, 2005). ETV6 fuses to a second gene at 4q12, CHIC2/BTL in undifferentiated AML (Cools et al., 1999). Since PGFRA and CHIC2/BTL are only 200 kb apart, ETV6-PDGFRα and ETV6-CHIC2 would be expected to have a cytogenetically indistinguishable t(4;12)(q12;p13), however our case was reported to have a t(4;12)(q23;p11.2) which suggests that the rearrangement may have been more complex.

In common with other PDGFRA fusions, we found that STRN disrupts the region encoding negative regulatory WW-like domain of Pdgfrα. In contrast, for ETV6-PDGFRα the WW-like domain is intact (Fig 2C). Recently, Stover et al (2006) showed that truncation of the Pdgfrα WW-like domain is sufficient to activate the kinase in the absence of an oligomerisation domain provided by the partner protein. An artificial ETV6-PDGFRα fusion that retained the region encoding the WW-like domain transformed Ba/F3 cells to growth factor independence, indicating that autoinhibition by the WW-like domain can be overcome by enforced
homodimerisation. However, the growth of the transformed cells was slower than those transformed by similar fusions in which the WW-like domain was truncated (Stover et al., 2006). It remains unclear why the WW-like domain is truncated with other fusion partners that provide an oligomerisation domain such as Bcr, and also why truncation of this domain is not usually seen for fusions involving Pdgfrβ.

Acknowledgements

This work was supported by the Leukaemia Research Fund.

References


ETV6-PDGFRB results from the t(5;12)(q31–33;p13) and, although rare, is the most frequent of at least 15 known PDGFRB fusion genes. ETV6-PDGFRB is usually seen in patients with chronic myelomonocytic leukaemia, atypical chronic myeloid leukaemia (aCML) or chronic eosinophilic leukaemia (CEL), with most patients exhibiting peripheral blood or bone marrow eosinophilia. Although ETV6-PDGFRB disease may be controlled with conventional therapy (hydroxyurea, busulphan, interferon-x), many die from transformed disease or other causes. More recently, however, imatinib has emerged as the standard therapy for PDGFR-rearranged patients, with most experiencing sustained haematologic and cytogenetic remission, and many achieving molecular remission. Molecular confirmation of a PDGFR rearrangement is therefore important for appropriate clinical management and monitoring response to therapy. For the t(5;12)(q31–33;p13), confirmation of the presence of ETV6-PDGFRB is particularly important since many cases with this translocation have a distinct molecular aetiology, with upregulation of IL-3 expression and no involvement of PDGFRB.

In the great majority of positive cases, ETV6 exon 4 splices to PDGFRB exon 11 and is typically detected by reverse transcription PCR (RT-PCR) using primers directed to ETV6 exon 3 or 4 and PDGFRB exon 11. Out of 10 positive cases we have detected with these primer combinations, only one had a variant fusion in which ETV6 exon 4 was fused to PDGFRB exon 9 (unpublished data). Recently, a patient with idiopathic myelofibrosis and an ETV6 exon 7–PDGFRB exon 10 fusion was reported and it was suggested that this novel transcript may have accounted for the somewhat unusual clinical phenotype seen in this case. Importantly, both these variants are detectable using the standard primer sets above. Here we report a novel ETV6-PDGFRB fusion with breakpoints outside this region.

A 19-year-old male (E1026) presented to the outpatient department with a 2-month history of fever. On examination he was found to have generalized lymphadenopathy, hepatosplenomegaly and bony swellings over the left clavicle and right tibia. There were no bleeding manifestations. A peripheral blood smear showed a total leucocyte count exceeding 100 × 10^9/L with a left shift. The bone marrow aspirate was suggestive of a myeloproliferative disorder with an increase in the number of basophils and eosinophils. Fine needle aspiration of the lymph node showed infiltration by blasts which cytochemistry revealed to be negative for myeloperoxidase and periodic acid schiff. Flow cytometry of the peripheral blood and lymph node aspirate showed positive for myeloid markers (CD 11c, 13, 15 and 33) along with CD45 and myeloperoxidase. BCR-ABL was not detected by RT-PCR and cytogenetics showed a t(5;12)(q31–33;p13). A diagnosis of atypical CML in blast crisis was made. Despite the fact that ETV6-PDGFRB was not detected by RT-PCR, he was started on imatinib mesylate (300 mg twice a day) after taking informed consent. The patient had a complete haematological and cytogenetic response (including complete regression of the lymph nodes, bony swellings and hepatosplenomegaly) at 3 months post-treatment. He continues in remission on 600 mg/day imatinib, 1 year after diagnosis.

In view of the excellent response to imatinib, we performed further analysis to determine if PDGFRB was involved (material was not available for molecular cytogenetic analysis). We have previously reported a simple multiplex RT-PCR assay to identify PDGFRB fusions by looking for overexpression of sequence encoding the kinase domain (which is retained in any fusion) relative to that encoding the extracellular domain (which is expressed only by the normal, unarranged gene). Attempts to establish a similar assay for PDGFRB were only partially successful and therefore we developed an analogous strategy using multiplex ligation-dependent probe amplification (MLPA). We designed probes targeting the region encoding the kinase domain and upstream sequences of PDGFRB, PDGFRB1 and FGFR1. MLPA was performed on patient and control cDNA essentially as described. The cDNA MLPA assay detected relative overexpression of the PDGFRB kinase domain in 9 of 9 patients known to harbour PDGFRB fusion genes but not in 16 normal controls. Overexpression of the kinase domain was also seen for E1026, suggesting the presence of a PDGFRB fusion (Figure 1a).
Figure 1  (a) The cDNA MLPA assay. Probes are in the following order: FGFR1 tyrosine kinase domain, FGFR1 3' extracellular domain, PDGFRB tyrosine kinase domain, PDGFRB 3' extracellular domain, PDGFRB tyrosine kinase domain, PDGFRB 3' extracellular domain. Primer sequences and exact conditions are available on request. Under normal circumstances, expression of FGFR1, PDGFRB and PDGFRB is low in level in peripheral blood leukocytes. Overexpression of the PDGFRB intracellular domain is apparent in both the previously known ETV6-PDGFRB case and also E1026. (b) Electropherogram of the cDNA junction amplified by 5 rapid amplification of cDNA ends PCR. MLPA, multiplex ligation-dependent probe amplification.

Figure 2  Sequence of ETV6-PDGFRB fusions described to date, with the most common variant at the top and the fusion in case E1026 at the bottom.
We read with interest the article by Karlsson et al. described for FIP1L1-PDGFRB advanced phase. Good responses to imatinib have been reported. Case responded well to imatinib despite being treated in just two cases. It is noteworthy, however, that our case is reminiscent of that seen for FIP1L1-PDGFRB, for which breaks almost always fall within PDGFRB exon 12 and FIP1L1 intron-derived sequence is frequently incorporated into the mature mRNA. As a consequence, the PDGFRα autoinhibitory WW-like domain is disrupted resulting in partner protein-independent activation of the kinase moiety. The disruption of PDGFRβ WW-like domain in our case is unexpected in view of the fact that ETV6 encodes a dimerization domain that is required for transformation by normal ETV6-PDGFRB.

This unusual fusion in our case is reminiscent of that seen for FIP1L1-PDGFRB, for which breaks almost always fall within PDGFRB exon 12 and FIP1L1 intron-derived sequence is frequently incorporated into the mature mRNA. As a consequence, the PDGFRα autoinhibitory WW-like domain is disrupted resulting in partner protein-independent activation of the kinase moiety. The disruption of PDGFRβ WW-like domain in our case is unexpected in view of the fact that ETV6 encodes a dimerization domain that is required for transformation by normal ETV6-PDGFRB.

It is notable that our case presented with advanced phase disease. The fusion retained ETV6 exons 1–7, in common with that described by Tokita et al. in a case that also presented with advanced features. In contrast to the most common fusion involving ETV6 exons 1–4, these variants are predicted to encode a chimaeric protein that retains the ETV6 internal and ETS DNA-binding domains. It is possible that retention of these domains leads to a more aggressive phenotype; however, it would be premature to draw any clear conclusions from just two cases. It is noteworthy, however, that our case responded well to imatinib despite being treated in advanced phase. Good responses to imatinib have been described for FIP1L1-PDGFRB-associated CEL in transformation and this may be a general feature of diseases associated with PDGFRB fusions.

In summary, it appears that breakpoint diversity for the ETV6-PDGFRB fusion is more common than originally reported.

Acknowledgements

This work was supported by Leukaemia Research (UK).

CE Curtis1, FH Grand4, K Waghorn1, TP Sahoo2, J George2 and NCP Cross1

1Wessex Regional Genetics Laboratory, Salisbury District
2Hospital, Salisbury, UK and
3Department of Medical Oncology, Kidwai Memorial Institute of Oncology, Bangalore, India
E-mail: ncp@clive.soton.ac.uk

References

7 Stover EH, Chen J, Folens C, Lee BH, Mentens N, Maryn et al. Activation of FIP1L1-PDGFRAlpha requires disruption of the Nterminal membrane domain of PDGFRalpha and is FIP1L1-independent. Proc Natl Acad Sci USA 2006; 103: 8078–8083.

Treatment of severe autoimmune hemolytic anemia in B-cell chronic lymphocytic leukemia with alemtuzumab

Leukemia (2007) 21, 1841–1942; doi:10.1038/sj.leu.2404713; published online 26 April 2007

We read with interest the article by Karlsson et al., in Leukemia (2007), which reports five patients treated successfully with alemtuzumab (anti-CD52) for refractory autoimmune hemolytic anemia (AIHA) complicating B-cell chronic lymphocytic leukemia (B-CLL). AIHA occurs in about 5% of patients with B-CLL and requires treatment including corticosteroids, cytotoxic drugs, splenectomy or immunosuppressive drugs. Recently, rituximab (anti-CD20) has been used with efficacy in AHAI complicating B-CLL. But relapses are frequent and some patients become refractory.

We report the case of a 69-year-old woman who presented B-CLL complicated by AIHA. B-CLL was diagnosed in 1994 and was treated with different chemotherapy regimens (chlorambucil, cyclophosphamide and then fludarabine) between 1994 and 2002. In October 2003 chlorambucil was introduced for progressing B-CLL; the direct antiglobulin test (DAT) was positive and the patient was treated with alemtuzumab (anti-CD52) with a rapid response. The case is reminiscent of that seen for FIP1L1-PDGFRB, for which breaks almost always fall within PDGFRB exon 12 and FIP1L1 intron-derived sequence is frequently incorporated into the mature mRNA. As a consequence, the PDGFRα autoinhibitory WW-like domain is disrupted resulting in partner protein-independent activation of the kinase moiety. The disruption of PDGFRβ WW-like domain in our case is unexpected in view of the fact that ETV6 encodes a dimerization domain that is required for transformation by normal ETV6-PDGFRB.
The t(1;9)(p34;q34) and t(8;12)(p11;q15) Fuse Pre-mRNA Processing Proteins SFPQ (PSF) and CPSF6 to ABL and FGFR1

Claire Hidalgo-Curtis,1 Andrew Chase,1 Milton Drachenberg,2 Mark W. Roberts,2 Jerry Z. Finkelstein,2 Sarah Mould,3 David Oscier,3 Nicholas C. P. Cross,1 and Francis H. Grand 1*

1Wessex Regional Genetics Laboratory, Salisbury District Hospital and Human Genetics Division, University of Southampton, Southampton, United Kingdom
2Long Beach Memorial Medical Center, California, CA 90801
3Department of Hematology, Royal Bournemouth Hospital, Bournemouth, UK

We have investigated two patients with acquired chromosomal rearrangements, a male presenting with a t(1;9)(p34;q34) and B cell progenitor acute lymphoid leukemia and a female presenting with a t(8;12)(p11;q15) and the 8p11 myeloproliferative syndrome. We determined that the t(1;9) fused ABL to SFPQ (also known as PSF), a gene mapping to 1p34 that encodes a poly-pyrimidine tract-binding protein-associated splicing factor. The t(8;12) fused CPSF6, a cleavage and polyadenylation specificity factor, to FGFR1. The fusions were confirmed by amplification of the genomic breakpoints and RT-PCR. The predicted oncoproteic products of these fusions, SFPQ-ABL and CPSF6-FGFR1, are in-frame and encode the N-terminal domain of the partner protein and the entire tyrosine kinase domain and C-terminal sequences of ABL and FGFR1. SFPQ interacts with two FGFR1 fusion partners, ZNF198 and CPSF6, that are functionally related to the recurrent PDGFRα partner FIP1L1. Our findings thus identify a group of proteins that are important for pre-mRNA processing as fusion partners for tyrosine kinases in hematological malignancies.

INTRODUCTION

Diverse tyrosine kinase fusion genes have been identified in hematological malignancies, particularly myeloproliferative disorders (MPD), acute lymphoblastic leukemia (ALL), and certain subtypes of lymphoma. The 8p11 myeloproliferative syndrome (EMS) is a chronic MPD that commonly presents with eosinophilia, T-cell proliferation and acquired reciprocal chromosomal translocations with recurrent breakpoints at chromosome band 8p11 that disrupt the receptor tyrosine kinase FGFR1. To date, eight fusion genes have been identified by the molecular investigation of chromosomal rearrangements in patients with EMS; these are the t(8;13)(p11;q12), t(6;8)(q27;p11), t(8;9)(p11;q33), t(8;22)(p11;q22), t(8;19)(p12;q13.3), ins(12;8)(p11;p12), t(7;8)(q34;p11), t(8;7)(p11;q23), and which fuse ZMYM2, FGFR1OP, CEP110, BCR, HERV-K, FGFR1OP2, TRIM24, MYO18A, to FGFR1, respectively (Xiao et al., 1998; Popovici et al., 1999; Guasch et al., 2000, 2003; Demiroglu et al., 2001; Grand et al., 2004; Belloni et al., 2005; Walz et al., 2005). The FGFR1 fusion genes encode constitutively active proteins that are functionally and structurally similar to BCR-ABL. The ABL gene has also been described as being fused to multiple partner genes: BCR (Groffen et al., 1984), ETV6 (also known as TEL) (Golub, 1997), NUP214 (Graux et al., 2004), and EML1 (De Keersmaecker et al., 2005). Although uncommon, these fusions are important to recognize because of their potential to be targeted effectively by small molecular inhibitors (Cross and Reiter, 2006).

For most tyrosine kinase fusions, the partner proteins provide dimerisation motifs that are essential for transforming activity; however, accumulating evidence suggests that partner proteins may play additional roles in modifying intracellular signalling and disease phenotype. Fusions such as FIG-ROS (Charest et al., 2003) and FGFR1OP-FGFR1 (Delaval et al., 2005) are directed by the partner protein to specific intracellular compartments, which may be critical to their transforming ability. BCR fuses to multiple kinases including PDGFRα (Baxter et al., 2002), FGFR1 (Demi-
roglu et al., 2001), and ABL (Groffen et al., 1984). BCR-FGFR1 binds GRB2 via BCR Tyr177 and induces a CML-like leukaemia in mice, whereas BCR-FGFR1/Y177F (which lacks the GRB2 binding site) causes EMS-like disease. These results implicate different signalling pathways originating from both the kinase and the fusion partner in the differential pathogenesis of CML and EMS (Roumiantsev et al., 2004).

In this study, we have investigated an acquired t(1;9)(p34;q34), and identified SFPQ as a new fusion partner to ABL. The second patient presented with EMS and an acquired chromosomal rearrangement of 8p11, resulting in the fusion of a new partner gene, CPSF6 to FGFR1. Both partner genes share a common role associated with pre-mRNA processing and our findings thus identify a group of proteins that are important for pre-mRNA processing as fusion partners for multiple tyrosine kinases in hematological malignancies.

**MATERIALS AND METHODS**

**Patient Details**

**Case 1:** A 22-year-old male patient presented with CD34 positive ALL of pre-B immunophenotype. His full blood count at presentation was as follows: hemoglobin 7.8 g/L; WBC 39.7 × 10⁹ L⁻¹ (89% blasts); platelets 25 × 10⁹ L⁻¹. Bone marrow analysis revealed 94% blasts and cytogenetics showed a t(1;9)(p34;q34)[7]/47, idem, +21[6]/46, XY[7]. The patient was treated according to the Children’s Cancer Group Protocol, 1961, and subsequently received augmented BFM therapy with doxorubicin and double delayed intensification. He achieved complete remission but suffered extramedullary testicular relapse at 4.5 years. Following orchidectomy and intensive chemotherapy, he remains in complete remission more than 6 years after diagnosis.

**Case 2:** A 75-year-old female presented with cervical, left axillary, and inguinal lymphadenopathy. A CT scan showed right pleural effusion, ascites, splenomegaly, and para-aortic nodes. Axillary node biopsy showed effacement of normal architecture by a diffuse population of mononuclear cells. The majority expressed strong surface CD3, but there was a scattered population of cells expressing myeloid markers such as neutrophil elastase. The presenting blood count showed a hemoglobin of 140 g/L, WBC 16 × 10⁹ L⁻¹, with neutrophilia (12 × 10⁹ L⁻¹) and eosinophilia (0.8 × 10⁹ L⁻¹) plus platelets of 375 × 10⁹ L⁻¹. The marrow was hypercellular with an increase in eosinophils and eosinophil precursors but no increase in lymphoid cells or blasts. Cytogenetics analysis revealed 46,XX,t(8;12)(p11;q15)[19]/46, XX[1]. There was rapid clinical deterioration accompanied by pyrexia, lethargy, weight loss, and a widespread rash (biopsy was inconclusive on histology). Intensive chemotherapy was not offered and although there was a transient clinical response to steroids, she died 10 weeks later.

**5' Rapid Amplification of cDNA**

**Ends (5'RACE) PCR**

Partner genes were identified using the GeneRacer™ Kit (Invitrogen, Paisley, UK). Briefly, ~5 μg of total RNA extracted using the Qiagen RNeasy kit (Qiagen, Boundary Court, UK) was dephosphorylated, decapped, and ligated to the GeneRacer™ RNA oligo according to the manufacturer’s instructions. The ligated RNA was reverse transcribed using Superscript II™ reverse transcriptase (Invitrogen, Paisley, UK) and random primers (100 ng). Single step 5′RACE PCR was performed using the 5′ GeneRacer™ primer from the kit in combination with a reverse primer from ABL exon 4 (1R: 5′-CCA CGG TTG AAT GAT GAT GAT GAA CGG CC-3′) for Case 1. For Case 2, 5′RACE was performed using nested FGFR1 primers as previously described (Grand et al., 2004). The PCR cycles were designed to amplify fragments up to 3 kb, with an annealing temperature of 66°C using the High Fidelity PCR Master kit (Roche, Mannheim, Germany) according to the manufacturer’s instructions. Products were cloned with the TOPO TA Cloning Kit for Sequencing (Invitrogen, Paisley, UK) and sequenced.

**RT-PCR Methods**

**Case 1:** The presence of SFPQ-ABL mRNA was confirmed on random hexamer reverse transcribed cDNA using primers to SFPQ (Exon 5 1F: 5′-GAT GAA CC-3′) with the reverse ABL exon 4 1R (see above). Because of limited patient material the genomic breakpoint was initially amplified by long PCR using a forward genomic primer from intron 9 of SFPQ (2F: 5′-TG TGT TAG GCT GCT GTG TTT GAT GAT T-3′) and reverse primer ABL exon 4 1R. This product was cloned (Invitrogen, Paisley, UK) and sequenced using a reverse ABL primer from intron 4 (5′-TG TTT TTT GAC AGC CAC ACT GAC AG).

**Case 2:** The presence of CPSF6-FGFR1 fusion mRNA was confirmed on random hexamer reverse transcribed cDNA using primers to CPSF6 (Exon 8

*Genes, Chromosomes & Cancer* DOI 10.1002/gcc
1F: 5'-GGA GTG CTA TTG AGA CAC TGG TA-3' with the reverse FGFR1 RACE 1R (5'-AGA AGA ACC CCA GAG TTC ATG GA-3'). These primers were also used to amplify the genomic breakpoint.

**Fluorescence In Situ Hybridization Analysis**

To our knowledge, no other cases have been reported with a t(1;9)(p34;q34) or t(8;12)(p11;q15). For Case 1, FISH was performed using the Vysis ES BCR-ABL probes as described (Chase et al., 1997). Previous analysis suggested disruption of FGFR1 in Case 2 (Sohal et al., 2001).

**RESULTS**

**Characterization of the t(1;9)**

FISH using the Vysis ES probes that immediately flank the ABL gene hybridized to the normal copy of chromosome 9, also the der(9) and the der(1), indicating that the translocation targeted ABL; however, no BCR-ABL fusion was seen by FISH or RT-PCR. All ABL fusions reported to date result in the partner gene fusing to ABL exons a2 or a3. To identify the t(1;9) partner, we, therefore, performed 5'-RACE PCR from both these exons. These initial attempts failed despite the fact that normal 5' ABL sequence was readily obtained (data not shown). 5'RACE primers were subsequently designed in ABL exon 4 (accession number M14752). Sequencing of the products revealed several clones in which exon 9 SFPQ (accession number X70944) was fused in frame to ABL exon 4 (a4) (Fig. 1a).

**Confirmation of the SFPQ-ABL Fusion**

The presence of the SFPQ-ABL fusion was confirmed initially by RT-PCR. The reciprocal ABL-SFPQ product was also detectable by single step PCR. Cloning and sequencing revealed that the reciprocal product fuses ABL exon 3 (a3) to an alternative RNA isoform (see GenBank accession number 4685098) of SFPQ (Fig. 1a). As shown in Figure 1b, the SFPQ-ABL fusion was specifically amplified from patient cells by single step PCR, but was not detectable in normal controls. To further confirm the presence of the fusion the genomic breakpoint was amplified by long PCR, cloned, and sequenced. The genomic sequence near the breakpoint on the derivative chromosome 1 was 3' of exon 10 of SFPQ (GenBank accession number NM_005066.1) and was fused to sequence derived from intron 3 of ABL (Fig. 1c). The chimeric mRNA is formed through splicing to exon 9 (nucleotide 2072) of the SFPQ gene (NM_005066.1). The SFPQ-ABL fusion mRNA is predicted to encode a 1609 amino acid protein of 174 KDa that retains the N-terminal proline rich region and the two RNA recognition motifs (RRM) from SFPQ fused to part of the SH2 domain of ABL (Fig. 1d).

**Characterization of the t(8;12)**

FGFR1 had previously been reported as rearranged by FISH in this patient (Sohal et al., 2001). Because the breakpoints for all cases involving the FGFR1 gene thus far reported are tightly clustered within or close to exon 9, we investigated the possibility that FGFR1 was rearranged by 5'-RACE PCR. As shown on Figure 2a, sequencing of the rearranged 5'RACE band revealed a fusion between CPSF6 exon 8 (accession number NM_007007) and FGFR1 at nucleotide 1272 from ATG (accession number M34185).

**Confirmation of the CPSF6-FGFR1 Fusion**

To confirm the CPSF6-FGFR1 fusion, a 224 bp product was specifically amplified from patient cDNA but not from control DNA. The genomic breakpoint product of 543 bp was coamplified in the same reaction and sequenced (Figs. 2b and 2c). The CPSF6 breakpoint within intron 8 (Fig. 2c) results in the incorporation of part of the CPSF6 intron in the fusion mRNA; the CPSF6 intronic fragment maintains the reading frame with FGFR1 and encodes 11 amino acids.

A reciprocal FGFR1-CPSF6 fusion product was not detected with either single step or nested PCR (Fig. 2b). The CPSF6-FGFR1 fusion mRNA is predicted to encode a 895 amino acid protein of 97 KDa, that retains the N-terminal region and the RNA recognition motif (RRM) from CPSF6 fused to the entire tyrosine kinase domain of FGFR1 (Fig. 2d).

**DISCUSSION**

The SFPQ-ABL fusion is unique in that it is the first description of an ABL fusion to involve the fourth exon of ABL (exon a4). This is a structurally unusual ABL fusion protein as amino acids W127-K183 (accession number M14752) of the SH2 domain are not included whereas all previously described ABL fusion proteins retain the whole SH2 domain. Mouse models with BCR-ABL constructs in which the SH2 domain was inactivated have shown the fusion protein retained the ability to cause leukemia in mice. Under conditions in
which P210 and P190 forms of BCR-ABL cause fatal CML within 4 weeks, P210 SH2 mutants produced CML-like disease in some mice after a significant delay, with others developing a B-lymphoid leukemia (Roumiantsev et al., 2001).

SFPQ is a multifunctional protein able to bind single and double stranded DNA as well as RNA (Shav-Tal et al., 2002). SFPQ was first identified in a complex with polypyrimidine tract-binding protein (PTB), and is an essential factor for RNA splicing. SFPQ has previously been identified in papillary renal cell carcinoma as the fusion partner of the transcription factor TFE3. As with the SFPQ-ABL fusion, SFPQ-TFE3 includes exons 1–9 of SFPQ (NM_005066.1). SFPQ also binds to NPM-ALK (Crockett et al., 2004) and is
phosphorylated at Y293F by the NPM-ALK fusion. It has been postulated that SFPQ function is perturbed in NPM-ALK transformed cells (Crockett et al., 2004; Galietta et al., 2007).

As described above, constitutive activation of fusion tyrosine kinases is mediated by an oligomerization domain in the partner protein. The coiled-coil prediction program ncoils (http://www.sanger.ac.uk/cgi-bin/Pfam/swisspfamget.pl?name=P23246) (Lupas et al., 1994) indicated the presence of two coiled-coil domains between SFPQ amino acids 498–529 and amino acids 532–595. It is possible that one or more of these domains is responsible for oligomerization of SFPQ-ABL.

CPSF6 is the ninth partner gene found fused to FGFR1. All these fusions involve FGFR1 exon 9, and therefore disrupt the FRS2 binding domain (Hoch and Soriano, 2006). No recognisable oligomerisation motifs are identifiable in CPSF6, although it has been demonstrated previously that RNA recognition motifs (RRM), such as the one retained in CPSF6-FGFR1, may mediate homodimerization (Simpson et al., 2004; Handa et al., 2006) and it is possible that this domain causes the oligomerization of CPSF6-FGFR1.

CPSF6 has also been found in purified splicing factors with SFPQ and NonO, linking these proteins to both transcription and processing of pre-mRNAs (Rappsilber et al., 2002; Zhou et al., 2002; Dettwiler et al., 2004). During pre-mRNA processing, CPSF6 binds the highly conserved polyadenylation site AAUAAA that is present in almost every mRNA precursor, as well as interacting with the poly-A binding protein II (PABII) that binds the growing poly-A tail. Together CPSF6 and PABII activate poly (A) polymerase (PAP), by

Figure 2. A: Sequence of the CPSF6-FGFR1 mRNA fusion; (B) amplification of the CPSF6-FGFR1 fusion specifically from cDNA in the t(8;12) patient (E0003) but not the normal controls. The reciprocal fusion product was not detected (C) Sequences surrounding the genomic breakpoints in CPSF6 (underlined) and FGFR1; (D) domain structure of FGFR1, CPSF6, and the predicted fusion protein. Relevant protein domains are shown: RNA recognition motif (RRM), transmembrane domain, and tyrosine kinase domain for FGFR1.
directing the enzyme to the RNA allowing polyadenylation of the upstream cleavage product (Ruegsegger et al., 1998).

The U1 snRNP has been shown to function in polyadenylation and has been found in complex with SFPQ (Lutz et al., 1998; Liang and Lutz, 2006). Another FGFR1 fusion partner, ZNF198 (Xiao et al., 1998), has been found to immunoprecipitate with SFPQ (Kaysapa et al., 2005). ZNF198 was proposed as a possible scaffold protein bringing RNA-binding proteins together to facilitate pre-mRNA splicing (Kaysapa et al., 2005). In addition, FIP1L1, which fuses to PDGFRα in chronic eosinophilic leukemia (Cools et al., 2003), also plays a role in polyadenylation and pre-mRNA splicing (Preker et al., 1995; Kaufmann et al., 2004). Our findings, thus, identify proteins involved in pre-mRNA processing as a functionally related group that fuse to tyrosine kinase genes in hematological malignancies.

REFERENCES


HIDALGO-CURTIS ET AL.


Frequent CBL mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms

Francis H. Grand,1 Claire E. Hidalgo-Curtis,1 Thomas Ernst,1 Katerina Zoi,2 Christine Zoi,2 Carolann McGuire,3 Sebastian Kreil,1 Amy Jones,1 Joanna Score,1 Georgia Metzgeroth,4 David Oscier,5 Andrew Hall,6 Christian Brandts,7 Hubert Serve,7 Andreas Reiter,4 Andrew J. Chase,1 and Nicholas C. P. Cross1

1Wessex Regional Genetics Laboratory, Salisbury, and Human Genetics Division, School of Medicine, University of Southampton, Southampton, United Kingdom; 2Haematology Research Laboratory, Biomedical Research Foundation, Academy of Athens, Athens, Greece; 3Inflammation, Infection and Repair Division, School of Medicine, University of Southampton, Southampton, United Kingdom; 4III Medizinische Universitätsklinik, Fakultät für Klinische Medizin, Mannheim der Universität Heidelberg, Mannheim, Germany; 5Department of Haematology, Royal Bournemouth Hospital, Bournemouth, United Kingdom; 6Northern Institute for Cancer Research, Newcastle upon Tyne, United Kingdom; and 7Department of Medicine, Hematology & Oncology, University of Frankfurt, Frankfurt, Germany

Recent evidence has demonstrated that acquired uniparental disomy (aUPD) is a novel mechanism by which pathogenetic mutations in cancer may be reduced to homozygosity. To help identify novel mutations in myeloproliferative neoplasms (MPNs), we performed a genome-wide single nucleotide polymorphism (SNP) screen to identify aUPD in 58 patients with atypical chronic myeloid leukemia (aCML; n = 30), JAK2 mutation–negative myelofibrosis (MF; n = 18), or JAK2 mutation–negative polycythemia vera (PV; n = 10). Stretches of homozygosity, copy neutral SNP calls greater than 20Mb were seen in 10 (33%) aCML and 1 (6%) MF, but were absent in PV. In total, 7 different chromosomes were involved with 7q and 11q each affected in 10% of aCML cases. CBL mutations were identified in all 3 cases with 11q aUPD and analysis of 574 additional MPNs revealed a total of 27 CBL variants in 26 patients with aCML, myelofibrosis or chronic myelomonocytic leukemia. Most variants were nonsense substitutions in the RING or linker domains that abrogated CBL ubiquitin ligase activity and conferred a proliferative advantage to 32D cells overexpressing FLT3. We conclude that acquired, transforming CBL mutations are a novel and widespread pathogenetic abnormality in morphologically related, clinically aggressive MPNs. (Blood. 2009;113:6182-6192)

Introduction

Myeloproliferative neoplasms (MPNs) are clonal hematopoietic stem cell disorders characterized by overproliferation of one or more myeloid cell lineages in the bone marrow and increased numbers of mature and immature myeloid cells in the peripheral blood. Excess proliferation is frequently associated with splenomegaly and cardiovascular complications as well as increased risk of transformation to acute leukemia. MPNs are categorized into subtypes based on specific morphologic, hematologic, and laboratory parameters, the best characterized being the 4 so-called classic MPNs: polycythemia vera (PV), essential thrombocythemia (ET), primary myelofibrosis (MF), and chronic myeloid leukemia (CML). In addition, several atypical MPNs are recognized, some of which show both dysplastic and proliferative features, such as atypical, BCR-ABL negative CML (aCML).2

MPNs are associated with acquired, activating mutations or gene fusions of tyrosine kinases, abnormalities that are believed to be critical drivers of excess proliferation as a result of deregulated or constitutive signaling.3 The 2 most prominent examples are BCR-ABL in CML2 and the V617F JAK2 mutation in PV, ET, and MF5,5 but more than 40 variant tyrosine kinase fusions have been identified in MPNs as well as other mutations in JAK2 and FLT3.2 Activating mutations have been described in components that signal upstream (eg, MPL) or downstream (eg, NRAS) of tyrosine kinases;9,10 however, the molecular pathogenesis of the majority of atypical MPNs and approximately 50% of ET and MF cases remains obscure.

V617F JAK2 was initially identified by several different routes, one of which was based on the observation that many PV patients show evidence of acquired uniparental disomy (aUPD) at chromosome 9p.11,12 Regions of aUPD exhibit loss of heterozygosity (LOH) compared with constitutional DNA without change of copy number and arise by mitotic recombination followed by selection for one of the products. After the initial observations in PV, it has emerged that aUPD is common in both hematologic and epithelial malignancies, and is associated with known oncogenic changes in a variety of genes within the affected regions.13 In this study we set out to determine whether aUPD characterizes MPNs of unknown molecular etiology and, if so, whether it could be used as a tool to help identify novel driver mutations.

Methods

Patients

Peripheral blood or bone marrow samples were received from patients diagnosed with an MPN or other hematologic malignancy according to
standard morphologic, hematologic, and laboratory criteria. Clinical data were available from a subset of these cases. The study was approved by the Internal Review Boards from participating institutions and informed consent was obtained in accordance with the Declaration of Helsinki.

SNP array analysis

DNA labeling and hybridization to Affymetrix 50K XbaI chips was performed at the Deutsches Ressourcenzentrum für Genomforschung (RZPD, Berlin, Germany). Raw data were imported into the Affymetrix GeneChip Operating Software, analyzed using Affymetrix (High Wycombe, United Kingdom) GeneChip Genotyping Analysis Software (GTYPE 4.1) and copy number analysis tool (CNAT). Data were exported to custom-designed spreadsheets that display loss of heterozygosity and copy number changes in ideogram format. Data were also analyzed using Affymetrix Genotyping Console (version 2.1). Overall, a median of 98.2% (range, 91.5%-99.6%) of SNPs gave readable calls.

Mutation analysis

Detection of mutations by high resolution melting (HRM) analysis was performed as described using a Rotor-Gene 6000 (Corbett Life Sciences, St Neots, United Kingdom). Direct sequencing of polymerase chain reaction (PCR) products was performed by standard techniques on an ABI 3130 Genetic Analyzer (Applied Biosystems, Warrington, United Kingdom) and analyzed using Mutation Surveyor (SoftGenetics, State College, PA). CBL mutations are numbered relative to the ATG (A=1) in the Ensembl cDNA sequence ENSG00000110359. A dropping factor (relative intensity drop of the wild-type allele peak relative to that seen in a concurrently run normal sample) of 60% or more was considered indicative of a biallelic mutation whereas a dropping factor of less than 60% was considered as monoallelic. Primer sequences for CBL exon 8 were CBL_i7f (5'-tgtgtagcattctaaaaatcgga-3') and CBL_i8r (5'-gccgagcaggccctcttgc-3') and for CBL exon 9 were CBL_i8f (5'-ggccgcttgggttagg-3') and CBL_i9r (5'-cacaagttgggtttctc-3'). CBL reverse-transcriptase PCR (RT-PCR) was performed on random hexamer–reverse transcribed RNA using primers CBL_e7F: 5'-aatgctgtctttgatgccag-3' and CBL_e10R 5'-cgagccagcagtca-3'. Other primer sequences are available on request.

Microsatellite analysis

DNA samples were amplified with a series of fluorescently labeled primer pairs flanking highly polymorphic microsatellite markers on chromosome 11q as described. Affymetrix 50K XbaI was scored if 4 or more consecutive markers encompassing CBL were homozygous without change in allelic copy number (P<.05 of this occurring in the absence of aUPD based on published rates of heterozygosity). Copy number at 11q was estimated by multiplex ligation probe amplification (MLPA) using the Marfan probe set1 (P065; MRC Holland, Amsterdam, The Netherlands), a kit that contains four 11q control probes including one at 11q23.3.

Constructs

pMY-wtCBL and pMY-CBLR420Q express wild-type and R420Q CBL, respectively, along with green fluorescent protein (GFP). Other CBL mutant constructs were derived from pMY-wtCBL using the QuikChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA). pAL/FLT3 and pCMV–HA ubiquitin have also been described previously.

Cell lines

The IL-3–dependent murine myeloid cell 32Dcl3 (32D) was purchased from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (Braunschweig, Germany) and was maintained in 90% RPMI 1640 medium with 10% fetal bovine serum (FBS) and 10% WEHI-3B conditioned medium (WEHI). All cells were maintained in a humidified incubator at 37°C and 5% carbon dioxide. 32D cells were stably transduced by electroporation with wt-FLT3 in a pAL vector as previously described and maintained in 15 ng/mL blastocidin (Invitrogen, Paisley United Kingdom). The Platinum-E (Plat-E) retrovirus packaging cell line was transduced using the calcium phosphate method with 5 µg CBL construct DNA: 105 32Dwt/wtFLT3 cells were cocultivated with the Plat-E cells for 48 hours in a volume of 2 mL (50% RPMI, 50% DMEM, 10% FBS, 5% WEHI). GFP positive cells were flow sorted to highest possible purity using a BD FACSARia cell sorter (Becton Dickinson, Stanford, CA). If insufficient cells were obtained for experiments after flow sorting, the cells were maintained in culture and resorted to greater than 95% purity.

Growth factor withdrawal assays

Cell growth was compared with wt-FLT3, vector-only transduced cells and the parental 32D line in stably transduced cells. Cells (105/mL) were grown over 3 days in triplicate for each CBL mutant and a wt-CBL control in 96-well plates. Cultures were analyzed daily for proliferation using the Aqueous One Solution Cell Proliferation Assay kit (MTS assay; Promega, Southampton, United Kingdom). All experiments were performed at least 3 times independently and results were analyzed using GraphPad Prism 4 software (GraphPad Software, San Diego, CA).
Figure 1. Single nucleotide polymorphism array results and initial mutation analysis. (A) Array analysis of 2 control patients showing copy number (CN) and loss of heterozygosity (LOH) outputs. The x-axis for the CN plots show (copy number) - 2 as determined by CNAT, with values approaching or exceeding -1 indicating deletions and greater than -1 indicating amplifications. The x-axis for LOH shows log10 P values (ie, 20 indicates 10^-20). The 2 panels on the left are from a polycythemia vera (PV) case with a homozygous V617F JAK2 mutation. There is no gross copy number change (individual datapoints spread around the zero line), but a large block of homozygous SNP calls at 9p indicative of aUPD. The 2 panels on the right are from a chronic myeloid leukemia (CML) blast crisis case and show LOH accompanied by a 20q deletion. (B) For case UPN 5, a large block of 11q aUPD is apparent on analysis of granulocytes but almost completely absent from mononuclear cells (MNCs) extracted from the same sample. (C) Chromosome 1p aUPD in case UPN 27 is associated with a biallelic G>T MPL mutation that is predicted to result in a W515L substitution. (D) Whole chromosome 13 isodisomy in case UPN 28 associated with a homozygous FLT3 ITD. Two controls are shown: an acute myeloid leukemia (AML) patient previously known to have a heterozygous FLT3 ITD and a healthy individual.
consistent with an acquired abnormality (Figure 1B). Some of the affected regions harbor genes known to be relevant to hematologic malignancies and therefore we assessed their mutational status. Sequencing of the MPL gene in the case UPN 27 with lp aUPD revealed a homozygous G>T change (Figure 1C) that results in the previously reported W515L substitution.9 Case UPN 28 with whole chromosome 13 UPD had a homozygous FLT3 internal tandem duplication (ITD; Figure 1D). Candidate genes for the abnormalities at 17q and 21q include NF1 and RUNXI, respectively; however, these were not investigated.

11q aUPD is associated with CBL mutations

Next, we focused our analysis on the 2 recurrent regions of aUPD at 7q and 11q. Because of the strong association between MPNs and aberrant tyrosine kinase signaling, we analyzed genes encoding tyrosine kinases and associated signaling components. The minimally affected region at chromosome 7q contains 3 tyrosine kinases: MET, EPHA1, and EPHB6. We fully sequenced the coding exons of these genes in the 3 cases with 7q aUPD and 2 normal controls but found no abnormalities. Similarly, we found no sequence changes in exons 11 through 17 of BRAF, a gene encoding a downstream component of RAS signaling that is mutated in malignant melanoma and other cancers.20 There are no tyrosine kinase genes on chromosome 11q, and although there are several genes encoding signal transduction components, CBL stood out as a known oncogene that negatively regulates tyrosine kinase signaling.21 Sequencing of CBL in the three 11q aUPD cases revealed that UPN 2 and UPN 5 each had a T>C change at nucleotide 1277 in exon 8 that predicts a L380P substitution. Case UPN 3 had a C>G missense change in exon 9 at nucleotide 1387 that predicts a P417A change (Figure 2). Consistent with the array results, the mutations in cases UPN 3 and UPN 5 were predominant with the residual wild-type alleles only weakly visible. For case UPN 2 both the mutation and level of aUPD was less prominent, presumably due to a greater background of normal cells.

Prevalence and nature of CBL mutations

To determine the prevalence of CBL mutations, we sequenced exons 8 and 9 (that encode amino acids 366 through 409 and 410 through 477, respectively) in an additional 574 MPN cases. A total of 27 sequence variants were identified in 26 patients, of whom 3 had MF, 10 had chronic myelomonocytic leukemia (CMML), 12 had aCML and 1 had hyperesinophilic syndrome/chronic eosinophilic leukemia (HES/CEL; Tables 2 and 3). To ensure that we were not missing mutations elsewhere in the gene we analyzed all coding exons in 70 CBL mutation negative aCML or unclassified, CML-like MPN cases by high resolution melt (HRM) analysis but failed to detect any further sequence changes apart from known polymorphisms. Similarly, no mutations in CBLB exons 9 or 10 or CBLC exons 7 and 8 (the exons that correspond to CBL exons 8 and 9) were identified. The activity of BCR-ABL is known to increase during progression of CML from chronic phase to blast crisis22 and we speculated that abrogation of CBL activity might be at least partially responsible for this increase. HRM analysis of CBL exons 8 and 9 in blast crisis cases (n = 31), however, failed to reveal any mutations.

Of the 27 CBL variants, 21 (78%) were missense substitutions affecting 12 amino acids (15 different substitutions), 5 (19%) were candidate splicing abnormalities, and 1 (3%) produced a stop codon (Table 3). Most of the changes were novel, although 4 cases had R420Q as previously identified in a single case of AML.15 Apart from N454D, all residues affected by missense mutations were completely conserved in CBL orthologues as well as the 2 other human CBL family members (Figure 3). The mutation to a stop codon was found in a patient (UPN 9) who also had the Y371H variant; PCR across the affected regions and sequencing indicated that the 2 changes were on different alleles.

CBL mutations are acquired and rarely seen in conjunction with other known mutations

For 2 cases (UPN 5 and UPN 14) we were able to compare the mutational status of leukocytes and T cells or buccal epithelia cells. In both cases the L380P mutation was only present in leukocytes and was therefore acquired (Figure 4A). A third case (UPN 15) originally presented with V617F JAK2 negative ET in 1989 but progressed to MF in 2004. Analysis of DNA extracted from sequential bone marrow slides indicated that the R420Q mutation was absent at the ET phase but acquired on transformation (Figure 4B). The 3 CBL-positive MF cases tested negative for MPL 515 mutations and all CBL mutation positive cases were negative for cytogenetic indicators of tyrosine kinase fusions, FLT3 ITD, activating NRAS mutations and V617F JAK2, with the exception of 1 case (UPN 14) with MF who tested positive for both L380P CBL and V617F JAK2. Analysis of retrospective bone marrow slides for this case indicated the presence of V617F JAK2 at low level initially but absence of L380P CBL. The CBL mutation was detectable 7 years later and completely displaced V617F after a further 3 years, corresponding with a significant rise in peripheral leukocyte counts (Figure 4C). A further case (UPN 21) presented with systemic mastocytosis before evolving to CBL mutation-positive CMML but retrospective samples were unavailable.

Splicing abnormalities

RNA was available for 3 of the 5 cases with potential splicing mutations. Patients UPN 18 and UPN 20 both exhibited double bands by RT-PCR, sequencing of which showed one to be the normal product and the smaller band to result from the complete deletion of exon 8 in each case (Figure 4D), similar to that described in the cell line MOLM-13.23 Patient UPN 19 showed a single band by RT-PCR that was entirely normal on sequence analysis indicating that the exon 8+4 C>T change does not alter splicing. It seems likely, therefore, that the variant identified in case UPN 6 is also of no consequence. The pathogenicity of the intronic variant seen in case UPN 13 remains unknown.

### Table 1. Chromosomal regions and cases affected by aUPD

<table>
<thead>
<tr>
<th>Disease category</th>
<th>No. with aUPD / no. analyzed</th>
<th>Region affected</th>
<th>Size of affected region; Mb</th>
</tr>
</thead>
<tbody>
<tr>
<td>aCML</td>
<td>10/30</td>
<td>7q11.21-qter</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7q11.22-qter</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7q22.3-qter</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11q12.1-qter</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11q13.1-qter</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11q13.5-qter</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13pter-qter</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17cen-qter</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20q11.22-qter</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>21q21.1-qter</td>
<td>29</td>
</tr>
<tr>
<td>MF</td>
<td>1/18</td>
<td>1pter-1p21.3</td>
<td>95</td>
</tr>
<tr>
<td>PV</td>
<td>0/10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CML-BC</td>
<td>0/30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLL</td>
<td>0/20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aUPD indicates acquired uniparental disomy; aCML, atypical chronic myeloid leukemia; MF, myelofibrosis; PV, polycythemia vera; CML-BC, CML in transformation to blast crisis; and CLL, chronic lymphocytic leukemia.
The association between \textit{CBL} mutations and UPD

To examine the association between \textit{CBL} mutations and aUPD in more detail we performed microsatellite analysis on all mutated cases using 9 polymorphic markers spanning chromosome 11q13-qter. Of the 26 cases, 11 displayed patterns that indicated significant tracts of homozygosity, including all 3 with 11q aUPD detected by SNP arrays. Case UPN 9 with 2 mutations was heterozygous for all markers. Of the 11 cases with biallelic \textit{CBL} mutations as judged by sequence analysis, 5 showed complete homozygosity at all microsatellite loci tested and 6 had at least 4 consecutive homozygous calls encompassing the location of the \textit{CBL} gene at 11q23.3 (Figure 5). We performed multiplex ligation probe amplification (MLPA) analysis to determine whether the observed homozygosity was a consequence of a deletion but no copy number changes were detected (data not shown) and therefore it is likely that homozygosity arose by aUPD. The number of

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{\textit{CBL} mutations in the 3 cases with 11q aUPD. (A) Genotyping Console output indicating homozygous SNP calls in green, copy number state as estimated by a Hidden Markov Model in blue, and smoothed copy number log(2) ratio values in red. (B) Ideograms showing lack of large copy number changes and blocks of homozygosity at 11q. Abbreviations are as in Figure 1. (C) Sequence changes indicated by \textasciitilde in each case compared with normal controls. UPN 2 and UPN 5 each have the c.1139T>C: L380P mutation whereas UPN 3 has c.1249C>G: P417A.}
\end{figure}
Table 2. Cases analyzed for CBL mutations

<table>
<thead>
<tr>
<th>Disease category</th>
<th>Cases analyzed</th>
<th>Cases with CBL variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV</td>
<td>74</td>
<td>0</td>
</tr>
<tr>
<td>MF</td>
<td>53</td>
<td>3 (6%)</td>
</tr>
<tr>
<td>ET</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>SM</td>
<td>60</td>
<td>0</td>
</tr>
<tr>
<td>CMML</td>
<td>78</td>
<td>10 (13%)</td>
</tr>
<tr>
<td>aCML</td>
<td>152</td>
<td>12 (8%)</td>
</tr>
<tr>
<td>HES/CEL</td>
<td>96</td>
<td>1 (1%)</td>
</tr>
<tr>
<td>CNL</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>CML-BC</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>577</td>
<td></td>
</tr>
</tbody>
</table>

PV indicates polycythemia vera; MF, myelofibrosis; ET, essential thrombocytopenia; SM, systemic mastocytosis; CML, chronic myelomonocytic leukemia; aCML, atypical chronic myeloid leukemia; HES/CEL, hypereosinophilic syndrome/chronic myeloid leukemia; CNL, chronic neutrophilic leukemia; and CML-BC, CML in transformation to blast crisis.

Table 3. List of CBL sequence variants

<table>
<thead>
<tr>
<th>Patient identifier</th>
<th>Diagnosis</th>
<th>Exon 8 amplicon</th>
<th>Exon 9 amplicon</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPN 1</td>
<td>aCML</td>
<td>c.1127C&gt;T</td>
<td>S376F</td>
</tr>
<tr>
<td>UPN 2</td>
<td>aCML</td>
<td>c.1139T&gt;C</td>
<td>L380P</td>
</tr>
<tr>
<td>UPN 3</td>
<td>aCML</td>
<td>N</td>
<td>c.1249G&gt;G</td>
</tr>
<tr>
<td>UPN 4</td>
<td>aCML</td>
<td>N</td>
<td>c.1259G&gt;A</td>
</tr>
<tr>
<td>UPN 5</td>
<td>aCML</td>
<td>c.1139T&gt;C</td>
<td>L380P</td>
</tr>
<tr>
<td>UPN 6</td>
<td>aCML</td>
<td>int-5G&gt;A</td>
<td></td>
</tr>
<tr>
<td>UPN 7</td>
<td>aCML</td>
<td>c.1224G&gt;C</td>
<td>W408C</td>
</tr>
<tr>
<td>UPN 8</td>
<td>aCML</td>
<td>N</td>
<td>c.1254C&gt;A</td>
</tr>
<tr>
<td>UPN 9</td>
<td>aCML</td>
<td>c.1111T&gt;C</td>
<td>Y371H</td>
</tr>
<tr>
<td>UPN 10</td>
<td>aCML</td>
<td>N</td>
<td>c.1384C&gt;T</td>
</tr>
<tr>
<td>UPN 11</td>
<td>aCML</td>
<td>N</td>
<td>c.1259G&gt;T</td>
</tr>
<tr>
<td>UPN 12</td>
<td>aCML</td>
<td>N</td>
<td>c.1259G&gt;A</td>
</tr>
<tr>
<td>UPN 13</td>
<td>CEL</td>
<td>int-2A&gt;G</td>
<td></td>
</tr>
<tr>
<td>UPN 14</td>
<td>MF</td>
<td>c.1139T&gt;C</td>
<td>L380P</td>
</tr>
<tr>
<td>UPN 15</td>
<td>MF</td>
<td>N</td>
<td>c.1259G&gt;A</td>
</tr>
<tr>
<td>UPN 16</td>
<td>MF</td>
<td>N</td>
<td>c.1259G&gt;A</td>
</tr>
<tr>
<td>UPN 17</td>
<td>CMML</td>
<td>c.1142G&gt;A</td>
<td>C381Y</td>
</tr>
<tr>
<td>UPN 18</td>
<td>CMML</td>
<td>c.1227→1227+4</td>
<td>del ggtac</td>
</tr>
<tr>
<td>UPN 19</td>
<td>CMML</td>
<td>int-4C&gt;T</td>
<td></td>
</tr>
<tr>
<td>UPN 20</td>
<td>CMML</td>
<td>int-1G&gt;C</td>
<td></td>
</tr>
<tr>
<td>UPN 21</td>
<td>CMML</td>
<td>c.1186T&gt;G</td>
<td>C396G</td>
</tr>
<tr>
<td>UPN 22</td>
<td>CMML</td>
<td>N</td>
<td>c.1250C&gt;T</td>
</tr>
<tr>
<td>UPN 23</td>
<td>CMML</td>
<td>c.1151G&gt;A</td>
<td>C384Y</td>
</tr>
<tr>
<td>UPN 24</td>
<td>CMML</td>
<td>N</td>
<td>c.1360A&gt;G</td>
</tr>
<tr>
<td>UPN 25</td>
<td>CMML</td>
<td>c.1150T&gt;C</td>
<td>C384R</td>
</tr>
<tr>
<td>UPN 26</td>
<td>CMML</td>
<td>c.1192C&gt;T</td>
<td>H398Y</td>
</tr>
</tbody>
</table>

_UPN indicates unique patient number; aCML, atypical chronic myeloid leukemia; CEL, chronic eosinophilic leukemia; MF, myelofibrosis; CMML, chronic myelomonocytic leukemia; and N, normal.

Correlation between CBL mutations and clinical features

We compared clinical and hematologic features in patients with or without CBL mutations, restricting the analysis to cases with aCML, MF and CMML because these were the subgroups in which CBL mutations were most commonly found. Patients with CBL mutations had a shorter overall survival and progression-free survival compared with mutation negative cases (overall survival: 33 months vs 39 months; progression-free survival: 22 months vs 32 months) but the differences were not significant (Figure 6). Similarly, there was no significant difference in gender distribution, age and standard hematologic parameters between mutation positive and mutation negative cases and no differences emerged when individual disease entities were considered.

Transforming activities of CBL mutants

We determined whether selected CBL variants had oncogenic activity by testing whether they could transform the interleukin-3 (IL-3)-dependent cell line 32D to growth factor independence in conjunction with overexpression of wild-type FLT3, as has been described previously for R420Q.15 We found that S376F, H398Y, P417A, and R420Q alone each conferred a degree of growth factor independence to 32D cells, but this effect was considerably enhanced in 32Ds that overexpressed wild-type FLT3. In contrast, wild-type CBL conferred no growth advantage either with or without FLT3 overexpression. The N454D change was not transforming, indicating that it is probably a rare polymorphism or a pathogenetically unimportant passenger mutation (Figure 7). Reproducible differences in survival between the different mutants were seen, as illustrated in Figure 8A.

Abrogation of E3 ubiquitin ligase activity

It has been shown that FLT3 physically associates with CBL and that, upon ligand stimulation, FLT3 is rapidly ubiquitinylated by wild-type but not R420Q CBL.15 We found that transforming activity correlated directly with the ubiquitin ligase activity of the CBL variants, as assayed by their ability to transfer HA-tagged ubiquitin to FLT3 after stimulation with FLT3 ligand. The S376F, H398Y, P417A and R420Q mutants showed loss of ubiquitin ligase activity, whereas N454D was comparable to wild-type CBL (Figure 8B).

Discussion

The initial aim of our study was to identify large regions of aUPD and therefore we undertook a relatively low resolution genome scan using 50k SNP arrays. Because we did not have constitutional DNA available from most of our cases, we were not able to unambiguously determine whether the extended blocks of homozygosity that we observed were a consequence of aUPD, constitutional UPD or autozygosity due to consanguinity, however we used a very conservative definition of candidate aUPD regions (at least 20 Mb homozygous SNPs calls running to a telomere) and therefore we undertook a relatively low resolution genome scan using 50k SNP arrays. Because we did not have constitutional DNA available from most of our cases, we were not able to unambiguously determine whether the extended blocks of homozygosity that we observed were a consequence of aUPD, constitutional UPD or autozygosity due to consanguinity, however we used a very conservative definition of candidate aUPD regions (at least 20 Mb homozygous SNPs calls running to a telomere) and therefore expected that most or all would be acquired. We found that aUPD was relatively common in aCML, but was uncommon in other MPNs that were negative for known mutations, plus also CML blast crisis and CLL. Six distinct regions of aUPD were identified in aCML, indicating substantial genetic heterogeneity in the genesis of this disorder. However, we identified 2 recurrent regions of aUPD at 7q and 11q that were each seen in 10% of cases indicating the presence of common molecular abnormalities.
Our mutation screening focused primarily on tyrosine kinases and associated signal transduction components because of the known association between deregulated tyrosine kinase signaling and MPNs. Although we failed thus far to identify any abnormality on chromosome 7q, we found missense "CBL mutations in all three 11q aUPD cases. Screening of further MPNs revealed a total of 27 "CBL variants in 26 patients who had been diagnosed with aCML (n = 12; 8%), MF (n = 3, 6%), CMML (n = 10, 13%), or

---

**Figure 3. Location of missense mutations.** (Top panel) Schematic representation of the CBL protein showing the tyrosine kinase binding (TKB), linker, RING, proline-rich (P-rich), and ubiquitin-associated/leucine zipper (UBA/LZ) domains. (Bottom panel) ClustalW alignment of the entire linker and RING domains plus part of the distal sequence. The RING domain is shaded in gray with the defining C and H residues shaded pale blue. Residues affected by missense mutations found in this study are indicated by ♦.

---

**Figure 4. Acquisition of CBL mutations.** (A) Comparison of granulocytes (gran) with buccal epithelial cells (bucc) or T cells (T) demonstrates that the L380P mutations in cases UPN 5 and 14 are acquired. (B) Sequential leukocyte and platelet counts for case, who presented in 1989 with essential thrombocythemia (ET) but progressed to myelofibrosis (MF) in 2004. Homozygous "CBL R420Q, indicated by *, was detected on transformation but was undetectable in previous specimens. (C) Clinical course of case UPN 14, who presented with MF in 1998 but experienced elevated leukocyte counts in 2005, concomitant with the appearance of the CBL L380P mutation. The JAK2 V617F mutation, which had been present at low level since diagnosis, was still detectable when "CBL L380P first appeared but was undetectable in the 2008 sample, when the "CBL mutated clone predominated (♦ on sequence trace). (D) RT-PCR analysis. All cases and controls show the expected product from CBL exons 7-10. Cases UPN 18 and UPN 20 show smaller bands that result from complete deletion of exon 8.
HES/CEL (n = 1, 1%); that is, morphologically and clinically related diseases that generally exhibit a poor prognosis. We were unable to discern, however, any specific clinical or prognostic features specifically associated with CBL mutated cases.

Casitas B-lineage lymphoma (CBL) is a well characterized protein that plays both positive and negative regulatory roles in tyrosine kinase signaling. In its positive role, CBL binds to activated receptor tyrosine kinases via its N-terminal tyrosine kinase binding (TKB) domain and serves as an adaptor by recruiting downstream signal transduction components such as SHP2 and P13K. However the RING domain of CBL has E3 ligase activity and ubiquitinylates activated receptor tyrosine kinases on lysine residues, a signal that triggers internalization of the receptor/ligand complex and subsequent recycling or degradation.21,24,25 CBL was originally identified after the characterization of v-Cbl, the transforming component of the Cas NS-1 retrovirus, and thus its association with neoplasia is well established26; however, it is only very recently that CBL mutations were first identified in a human malignancy, specifically occasional cases of AML.15,23,27 We have found that CBL mutations are much more common in MPNs, and previous reports of 11q aUPD in MDS28 suggest that CBL mutations will also prove to be common in this disease. Indeed, 7 of 12 MDS cases with aUPD at 11q were recently shown to harbor CBL mutations.29 There are 2 other human CBL family members, CBLB and CBLC, but we did not identify aUPD in the regions containing these genes and did not detect any mutations of the linker/RING domain of CBLB.

We identified a total of 15 different CBL missense mutations affecting 12 residues (Table 2; Figure 3). Mutations of some of these residues (eg, Y371, C381, H398 and W408) have been
analyzed previously in a study that concluded that loss of E3 ubiquitin ligase activity and concomitant impairment of EGFR internalization by RING finger mutations was insufficient for oncogenicity, as assayed in NIH3T3 cells. However mutations of key residues within the α-helix of the linker region abolished ubiquitin ligase activity and were transforming. Although the precise amino acid substitutions we identified were different from those studied, we found that mutations in both the RING finger and linker regions conferred autonomous growth to 32D/wt-FLT3 cells, consistent with a causal relationship to the MPN phenotype. The fact that one mutation (N454D) was not transforming emphasizes the importance of functional analysis to distinguish driver mutations from irrelevant passengers or infrequent polymorphisms. Notably, this mutation was only weakly conserved between species, in contrast to all other missense mutations that affected highly conserved residues and are therefore likely to be functionally significant.

Enhanced autonomous growth in the presence of CBL mutations and overexpression of wild-type FLT3 may simply reflect the paucity of receptor tyrosine kinase gene expression in 32D cells or, alternatively, may indicate that the mutants are only weakly transforming and require the cooperation of other events to give rise to clinically manifest disease. Consistent with the latter hypothesis, we identified 2 cases in which CBL mutations were acquired during progression of a preexisting MPN (Figure 4). In both cases the identity of the cooperating change is unknown: although one presented with a (relatively low level) V617F JAK2 mutation, this disappeared on progression to CBL L380P positive disease, indicating that the 2 mutations must have arisen independently in different clones. This is reminiscent of the observation that the leukemic blasts of approximately half of V617F positive MPNs that evolve to AML are negative for the JAK2 mutation. It remains to be

![Graph A: Overall survival](image)

![Graph B: Progression-free survival](image)

**Figure 6. Clinical significance of CBL mutations.** Overall survival (A) and progression-free survival (B) for atypical chronic myeloid leukemia (aCML), chronic myelomonocytic leukemia (CMML), and MF cases with (n=19) or without (n=87) CBL mutations. The differences between mutation positive and negative cases was not significant as determined by Kaplan-Meier analysis.

![Graphs of S376F, H398Y, P417A, R420Q, N454D](image)

**Figure 7. Transforming activity of CBL mutants.** 32D cells or 32Ds overexpressing wild-type FLT3 were transfected with wild-type or mutant CBL constructs and assayed for IL-3 independent growth over 3 days. Results shown are the mean of at least 3 independent experiments, each of which were performed in triplicate. The y-axis shows proliferation in arbitrary units as determined by MTS assay (Promega) and the x-axis time in days.
A CBL mutation resulting in skipping of exon 8 has also been identified in a murine model of NUP98-HOXD13 on progression from MDS to AML, further corroborating the theory that acute leukemia arises from complementary mutations, one of which inhibits differentiation, and a second (in this case CBL) that enhances proliferation or inhibits apoptosis.

The fact that many CBL mutations are associated with aUPD strongly suggests that clones with homozygous mutations have a selective advantage over those that are heterozygous. This contrasts with the prevailing view that transforming CBL RING finger and linker region mutants act in a dominant negative fashion. Although it is conceivable that another, unknown locus on 11q is the target of aUPD, the finding of 1 case with a Y371H mutation and a stop codon on opposite alleles strongly suggests that homozygous CBL mutations do indeed confer a direct selective advantage. Because we did not screen the entire gene for mutations, it is possible that other cases with heterozygous mutations may also be functionally homozygous due to premature stop codons on the other allele. Alternatively, heterozygosity in these cases may only be apparent due to variable proportions of background normal cells that are not part of the malignant clone. Although the nontransforming N454D mutant was heterozygous, 2 cases with heterozygous R420Q were observed and thus absence of homozygosity cannot be used to infer the presence of an irrelevant, nontransforming sequence variant.

In summary, we have found that oncogenic CBL mutations are acquired in a subset of hematologically related, poor prognosis MPNs. Our findings further strengthen the notion that MPNs are primarily “tyrosine kinopathies,” that is, diseases caused by aberrant activation of proliferation and survival pathways as a consequence of mutations that either directly or indirectly promote excess tyrosine kinase signaling. In view of the fact that tyrosine kinases are such good drug targets, the possibility that CBL-mutated cases might be amenable to therapeutic inhibition is clearly attractive.

Figure 8. Survival and abrogation of E3 ubiquitin ligase activity. (A) Relative survival/proliferation of 32D cells expressing wild-type or mutant CBL constructs 48 hours after growth factor withdrawal as determined by MTS assay. (B) 32D cells expressing wild-type FLT3 and CBL constructs as indicated were transfected with HA-tagged ubiquitin and stimulated with FLT3 ligand. After lysis, FLT3 was immuno-precipitated, blotted, and probed with anti-HA. Total blots for FLT3 AND CBL are also shown.

Acknowledgments

This study was supported by Leukaemia Research (London, United Kingdom), Dr Mildred Scheel Stiftung für Krebsforschung (Bonn, Germany), the Lady Tata Memorial Trust (London, United Kingdom), Deutsche Krebshilfe eV (Bonn, Germany), and the Deutsche José Carreras Leukämie-Stiftung eV-DJCLS R06/02 (Munich, Germany).

Authorship

Contribution: F.H.G. and C.E.H.-C. designed the study and performed or assisted with the laboratory work and data analysis; T.E. and K.Z. performed or assisted with the laboratory work and data analysis; C.Z. provided clinical data and samples; C.M., S.K., A.I., and J.S. performed or assisted with the laboratory work and data analysis; G.M. and D.O. provided clinical data and samples; A.H. performed or assisted with the laboratory work and data analysis; C.B. and H.S. provided essential reagents and advice; A.R. provided clinical data and samples; and A.J.C. and N.C.P.C. performed or assisted with the laboratory work and data analysis. All authors contributed to the design of the study and the final manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Prof N. C. P. Cross, Wessex Regional Genetics Laboratory, Salisbury NHS Foundation Trust, Salisbury SP2 8BJ, United Kingdom; e-mail: ncpc@soton.ac.uk.
References


