The Prognostic Significance of

the Mixed Lineage Leukaemia Partial

Tandem Duplication in

Acute Myeloid Leukaemia

by

Stephen J Austin

2010

A thesis presented to Cardiff University in partial fulfillment of the

requirement for the degree of Doctor of Philosophy

Department of Haematology,

School of Medicine,

Cardiff University.

Supervisors: Dr Richard Darley and Professor K Mills

Head of Department: Professor Alan Burnett

UMI Number: U517329

UMI

Dissertation Publishing

ProQuest®

Declaration and Statements

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ....(candidate)
Date ....
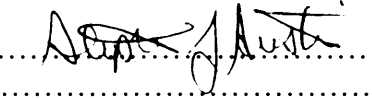
STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Signed ....(candidate)
Date ....
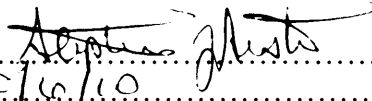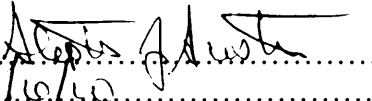
STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed ....(candidate)
Date ....

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ....(candidate)
Date..

Acknowledgements

I would like to acknowledge the invaluable contributions of several people:

Foreword

An accompanying CD is located in the back sleeve of this document and contains
supplementary material.

Dedicated to my wife Sarah,

for her tireless support and inspiration;

and to my children Daniel and Lily.

Posters and Presentations

Abstracts

Austin, S., Gilkes, M., Mills, K., Burnett, A. Gene Expression Profiling of AML Patients with *MLL*-PTD Identifies Differentially Regulated Genes. *20th Annual School of School of Medicine Postgraduate Research Day 2005* (Poster Presentation).

Austin, S., Gilkes, A., Holt, D., Mills, K., Pearce, L, Burnett, A. Internal Tandem Duplications of *FLT3* and Partial Tandem Duplications of *MLL* Can Provide Further Stratification of the Intermediate Risk Category of Acute Myeloid Leukaemia *Abstracts of the British Society of Haematology 45th Annual Scientific Meeting, 2004* (Poster Presentation).

Oral Presentations

"Mixed Lineage Leukaemia Partial Tandem Duplications in Acute Myeloid Leukaemia" *Department of Haematology Seminar Series*, February 2009

"Prognostic Significance of Mixed Lineage Leukaemia Partial Tandem Duplications in Acute Myeloid Leukaemia" *Department of Haematology Seminar Series*, November 2004

Abstract

Acute myeloid leukaemia (AML) is a heterogeneous disease, both in terms of biology and clinical outcome. Approximately 45% of all AML patients are cytogenetically normal at diagnosis and as such are placed in the intermediate risk group where their clinical outcome is unpredictable hence there remains a need for more prognostic markers. This study concerns a specific molecular genetic mutation, the mixed lineage leukaemia partial tandem duplication (*MLL* PTD). *MLL* is a transcriptional regulator that is known to be instrumental in both normal haematopoiesis and in leukaemogenesis. This study aimed to evaluate the prognostic importance of this abnormality through investigation of its influence on clinical outcome, its utility as a marker of minimal residual disease (MRD) and the downstream effects of its expression. A qualitative PCR assay was established and determined the frequency of *MLL* PTD to be 5.2% in *de novo* AML. *MLL* PTD was found to be a useful independent prognostic marker being associated with a greater risk of relapse and a reduction of relapse free survival and overall survival. A quantitative PCR (QPCR) assay was developed that reliably distinguished AML-related *MLL* PTDs from the low level of background *MLL* PTDs occurring normally. Despite a high rate of cytogenetic clonal evolution *MLL* PTD expression remained stable between diagnosis and relapse, making it a suitable marker for MRD. Results also suggested the QPCR assay could determine onset of remission and prediction of relapse. Gene expression profiling was used to identify a gene signature unique to the *MLL* PTD and was found to be distinct from that associated with *MLL* translocations. Analysis of the gene signature also identified three candidate chemotherapy compounds predicted to antagonise the effects of the *MLL* PTD. The results of this study could prove instrumental in improving the treatment of *MLL* PTD patients.

List of Abbreviations

ABL	Abelson tyrosine kinase
ALL	Acute lymphoid leukaemia
AML	Acute myeloid leukaemia
AML1	Acute myeloid leukaemia 1
AML1-ETO	Acute myeloid leukaemia 1- Eight twenty-one Oncoproteins
ANOVA	Analysis of variance
APL	Acute promyelocytic leukaemia
ARF	ADP-ribosylation factor
AT	Adenosine-thymine
ATRA	All-*trans* retinoic acid
BAALC	Brain and acute leukemia, cytoplasmic
BASP-1	Brain abundant, membrane attached signal protein 1
BCR	Breakpoint cluster domain
BCR-ABL	Breakpoint cluster region - Abelson tyrosine kinase
BIN	Biological interaction networks
BM	Bone marrow
BMP	Bone morphogenic protein
bp	Base pair
C5ar1	Complement component 5a receptor 1
CAV-1	Caveolin 1
CAV-2	Caveolin 2
CBF	Core binding factor
CBFα	Core binding factor alpha
CBFβ	Core binding factor beta
CBFβ-MYH11	Core binding factor beta- smooth muscle myosin heavy chain
CBX5	Chromobox homolog 5
CDKN2C	Cyclin-dependent kinase inhibitor 2C
cDNA	Complementary DNA
CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha
CFC	Colony forming cells
CFU	Colony-forming unit
CFU-GEMM	Colony-forming units – granulocyte-eyrthroid-macrophage-megakarocyte
CLP	Common lymphoid progenitors
CML	Chronic myeloid leukaemia
CMP	Common myeloid progenitors
c-MYC	Myelocytomatosis viral oncogene homolog
CN	Cytogenetically normal
Cp	Crossing point

CR	Complete remission
CRD	Complete remission duration
cRNA	Complementary RNA
CSF	Colony stimulating factor
dATP	Deoxyadenosine triphosphate
DAVID	Database for annotation, visualisation and integrated discovery
dCTP	Deoxycytidine triphosphate
DFS	Disease free survival
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DPP	Double peak positive
DS	Down's Syndrome
dsDNA	Double stranded DNA
dTTP	Deoxythymidine triphosphate
EDTA	Ethylenediaminetetracetic acid
ENPP2	Ectonucleotide pyrophosphatase/phosphodiesterase 2
EtBr	Ethidium bromide
ETO	Eight twenty-one oncoproteins
FAB	French-American-British
FDR	False discovery rate
FISH	Fluorescence *in situ* hybridisation
FL	Fms-like tyrosine kinase 3 ligand
FLT3	Fms-like tyrosine kinase 3
FOG	Friend of *GATA-1*
FYRC	F-y rich C terminus
FYRN	F-y rich N terminus
GAS-7	growth arrest-specific 7
GATA-1	GATA binding protein-1
G-CSF	Granulocyte - colony stimulating factor
GEP	Gene expression profiling
GM-CSF	Granulocyte/macrophage- colony stimulating factor GMP
GMP	granulocyte/macrophage progenitor
GOI	Gene of interest
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
hDOT1l	DOT1-like, histone H3 methyltransferase
HiDAC	High dose cytaribine
HMGB3	High-mobility group box 3
HOX	Homeobox

HR	Hazard ratio
HSC	Haematopoietic stem cell
HSP90	Heat shock protein 90
ID	Induction death
IL-3	Interleukin 3
ITD	Internal tandem duplication
IVT	*In vitro* transcription
JAK	Janus Kinase
JAK2	Janus Kinase 2
JMD	Juxtamembrane domain
KV	Kilivolt
LAIP	Leukaemia-associated aberrant immunophenotype
LIC	Leukaemia-initiating cells
LSC	Leukaemic stem cell
LT-HSC	Long-term haematopoietic stem cell
MAP	Mitogen-activated protein
MAPK	Mitogen-activated protein kinase
M-CSF	Macrophage colony-stimulating factor
M-CSFR	Macrophage colony-stimulating factor receptor
MDS	Myelodysplastic syndrome
MEIS1	Myeloid ecotropic viral integration site 1 homolog
MEN1	Multiple edocrine neoplasia type I
MEP	Megakaryocyte/erythrocyte progenitor
MgCl2	Magnesium chloride
MILE	Microarray innovations in Leukaemia
ml	Millilitre
MLL	Mixed lineage leukaemia
MLL-AF9	Mixed lineage leukaemia-ALL-1 fused to chromosome 9
MLL-ENL	Mixed-lineage leukaemia - eleven nineteen leukaemia
mM	Millimolar
MMuLV	Murine leukaemia virus
MOZ-TIF2	Monocytic leukaemia zinc finger - TGF-beta induced factor-2
MPD	Myeloproliferative disorder
MPN	Myeloproliferative neoplasms
MPP	Myeloid pluripotent progenitor
MRC	Medical Research Council
MRD	Minimal residual disease
MRDv	Minimal residual disease value
MT	Methyl transferase
MYB	Myeloblastosis viral oncogene homolog

MYH11	Smooth muscle myosin heavy chain
NCoR	Nuclear receptor corepressor
ng	Nanogram
NK	Natural killer cells
nm	Nanometre
NOD	Non-obese diabetic
NOS	Not otherwise specified
NPM1	Nucleophosmin 1
OD	Optical density
OR	Odds ratio
OS	Overall survival
p53	tumor protein p53
PB	Peripheral blood
PBS	Phosphate buffered saline
PBX3	Pre B-cell leukaemia transcription factor 3
PCA	Principle component analysis
PCR	Polymerase chain reaction
PHD	Plant homeodomain
PML	Promyelocytic leukaemia
PML-RARα	Promyelocytic leukaemia-retinoic acid receptor alpha
PTD	Partial tandem duplication
PU.1	Purine rich box-1
RAR	Retinoic acid receptor
RARA	Retinoic acid response element
RARE	Retinoic acid response elements
RAR-RXR	Retinoic acid receptor-retinoid X receptor
RAS	Rat sarcoma
RBC	Red blood cell
RCLB	Red cell lysis buffer
RFS	Relase free survival
RQ-PCR	Real-time quantitative-polymerase chain reaction
RR	Relapse risk
RT	Reverse transcription
RT-PCR	Reverse transcriptase-PCR
RXR	Retinoid X receptor
sAML	secondary AML
SCF	Stem cell factor
SCF-R	Stem cell factor-receptor
SCID	Severe combined immunodeficiency
SET	Variegation/enhancer of zeste/trithorax

SH2	Src homology 2
SLC5A8	Solute carrier family 5, member 8
SL-IC	SCID leukaemia-initiating cell
SMRT	Silencing mediator for retinoid or thyroid-hormone receptors
SNL	Speckled nuclear localisation sites
ssDNA	Single stranded DNA
STAT	Signal transducer and activator of transcription
STAT5	Signal transducer and activator of transcription 5
ST-HSC	Short term haematopoietic stem cells
TAD	Transactivation domain
TAM	Transient abnormal myelopoiesis
TBE	Tris borate EDTA
TFPI	Tissue factor pathway inhibitor
TKD	Tyrosine kinase domain
Tm	Melting temperature
TPD52	Tumour protein D52
TWIST1	Twist homolog 1
UK	United Kingdom
UPN	Unique patient number
WBC	White blood cells
WCC	White cell count
WT	Wild type
WT-1	Wilm's tumour 1
μg	Microgram
μM	Micromolar

List of Figures

List of Tables

Table of Contents

1 Introduction

1.1   Haematopoiesis

Blood is the fluid that sustains vertebrate life, performing a number of functions beyond nourishment and waste removal including protection of the body from infection by innate and adaptive immune responses, prevention of haemorrhage by coagulation and production of growth factors. The diverse functions performed by the blood are possible because of haematopoietic cell lineages that result in the production of specialised types of mature blood cells. The evolution of the lineages from a small population of haematopoietic stem cells (HSC) is called haematopoiesis.

1.1.1   Normal Haematopoiesis

Haematopoiesis is the ongoing process of differentiation, maturation and proliferation of mature blood cells from a renewable population of HSCs that occurs throughout the lifetime of an individual and gives rise to cell types that are specially adapted for a specific purpose within the body (Figure 1.1). Monocytes and granulocytes, (including neutrophils, eosinophils and basophils) represent the principal components of the innate immune system and co-operate to mount a non-specific defence against pathogens by phagocytosis, inflammation and B-cell signalling. The adaptive immune system, composed of B and T lymphocytes, is dependent on antigenic stimulation. The B lymphocytes undergo a process of education and maturation before they are capable of providing a defence through antibody-dependent cell-mediated cytotoxicity. The CD4$^+$ T helper lymphocytes are instrumental in this process, while the CD8$^+$ T suppressor-cytotoxic cells kill infected host cells. Additionally, T cells participate in the regulation of the immune system via cytokine production. The erythrocytes are responsible for the transport of oxygen and carbon dioxide

between tissues and the lungs. Platelets are instrumental in the initiation of the coagulation cascade to prevent haemorrhage, in addition to the release of cytokines, such as platelet derived growth factor, that promote wound repair through the regeneration of connective tissues.
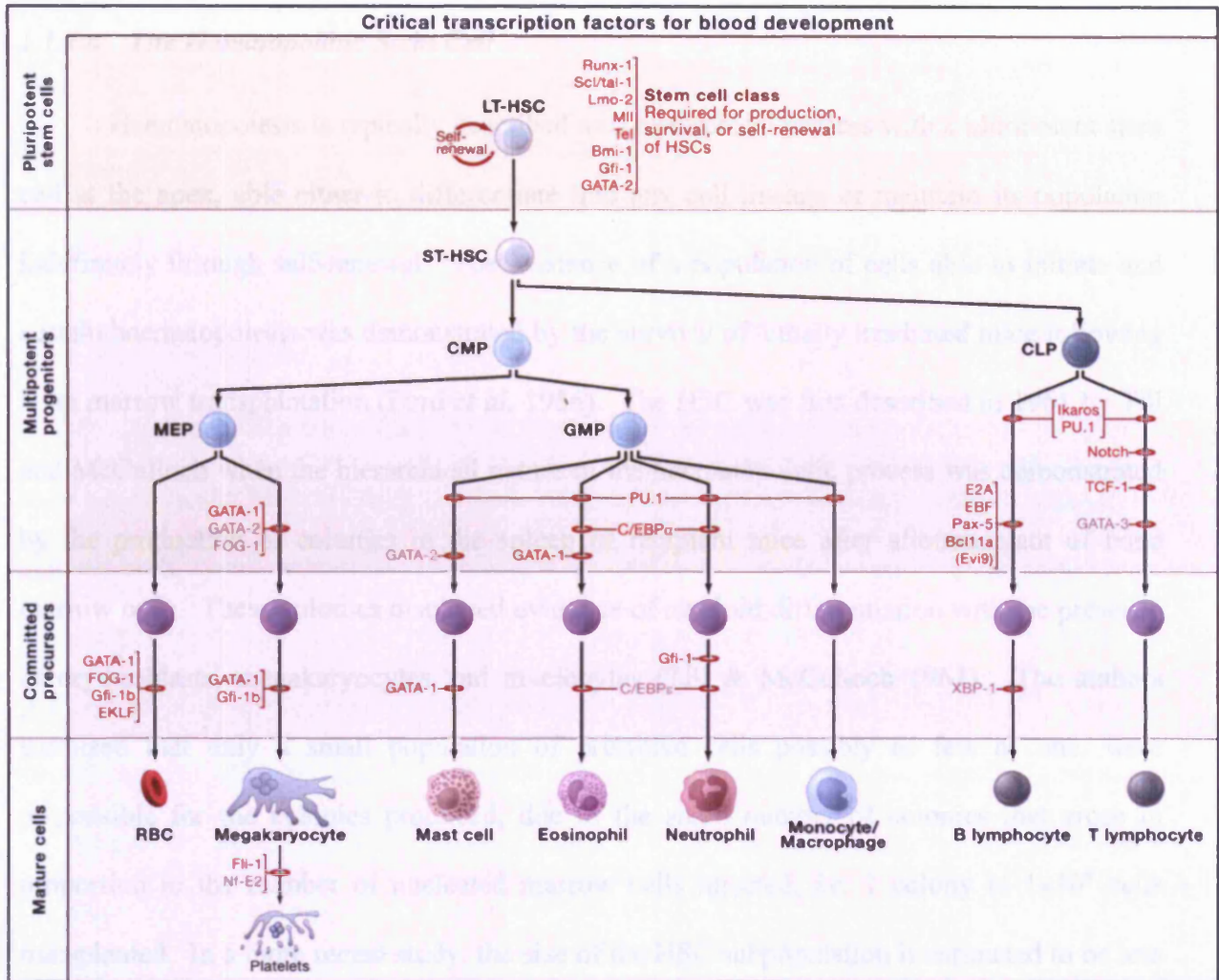
Figure 1.1 Haematopoietic Development
The curved arrows represent self-renewal, while the straight arrows represent developmental transitions in haematopoiesis. Red hoops represent points where developmental progress is blocked if the indicated transcription factor is absent. LT-HSC: long term haemtopoietic stem cell; ST-HSC: short term haematopoietic stem cell; CMP: common myeloid progenitor; CLP: common lymphoid progenitor; MEP: megakaryocyte/erythroid progenitor; GMP: granulocyte/macrophage progenitor; RBCs: red blood cells (Orkin & Zon 2008).

1.1.1.a   The Hematopoietic Stem Cell

Haematopoiesis is typically described as a hierarchical process with a pluripotent stem cell at the apex, able either to differentiate into any cell lineage or maintain its population indefinitely through self-renewal. The existence of a population of cells able to initiate and sustain haematopoiesis was demonstrated by the survival of lethally irradiated mice following bone marrow transplantation (Ford *et al.* 1956). The HSC was first described in 1961 by Till and McCulloch when the hierarchical nature of the haematopoietic process was demonstrated by the production of colonies in the spleen of recipient mice after allotransplant of bone marrow cells. These colonies displayed evidence of myeloid differentiation with the presence of erythroblasts, megakaryocytes and myelocytes (Till & McCulloch 1961). The authors surmised that only a small population of primitive cells possibly as few as one, were responsible for the colonies produced, due to the small number of colonies that arose in proportion to the number of nucleated marrow cells injected, *i.e.* 1 colony to $1x10^4$ cells transplanted. In a more recent study, the size of the HSC subpopulation is estimated to be less than 0.01% of the total cells in the bone marrow (Rizo *et al.* 2006).

With differentiation, the lineage potential of the cell grows increasingly restricted. HSCs are thought to become restricted in their lineage potential to form either a common myeloid progenitor (CMP) or common lymphoid progenitor (CLP). This represents a major divide in lineage commitment of the haematopoietic stem cell as the CLP gives rise to B and T lymphocytes as well as natural killer (NK) cells while the CMP can potentially develop into a megakaryocyte, erythrocyte, macrophage or granulocyte cells (Akashi *et al.* 2000; Kondo, Weissman, & Akashi 1997). To summarise, the HSC is defined by the features of indefinite self-renewal coupled with the ability to produce multiple lineages of differentiated progeny. Understanding the process of haematological regulation is of great importance as the

governing factors of regulation may suggest points within the haematopoietic network for therapeutic targeting when disrupted regulation results in leukaemic disease.

1.1.1.b Regulation of Haematopoiesis

1.1.1.b.i Overview

A regulated mechanism for the production of mature blood cells must maintain a steady state system by replacing old and damaged cells, and producing increased numbers of specific cell types in response to injury or infection. The process of haematopoietic differentiation is capable of producing 10^{12} mature cells every day from a relatively small population of HSC residing in the bone marrow (Ogawa 1993). The "traditional" hierarchy model where HSC become irreversibly lineage committed and progress to terminally differentiated progeny is likely an oversimplification; however, it has proved useful as a template for deciphering the stages of differentiation and the mechanisms by which haematopoiesis is regulated. A more accurate description of an HSC might be that of a cell with variable differentiation potential that is primarily controlled by the intrinsic influence of transcription factors and the extrinsic influence of haematopoietic growth factors (cytokines). Intrinsically, the transcription factors are required for HSC formation and lineage determination. The extrinsic influence of cytokines is critical in maintaining homeostasis by regulating the numbers of cells produced and can influence a variety of processes including quiescence, self-renewal, proliferation, differentiation, and apoptosis. Though the relative contribution of intrinsic and extrinsic regulation is unclear, it is understood as haematopoiesis proceeds there is a decrease in the capacity for self-renewal while the mechanisms promoting proliferation and differentiation increase (Figure 1.2). Additionally, negative regulation in the form of apoptosis contributes to overall homeostasis (De et al. 1999). Ultimately, the balance

of self-renewal, proliferation, differentiation and apoptosis regulates haematopoiesis so the function of supplying the necessary mature blood cells to the host is maintained.
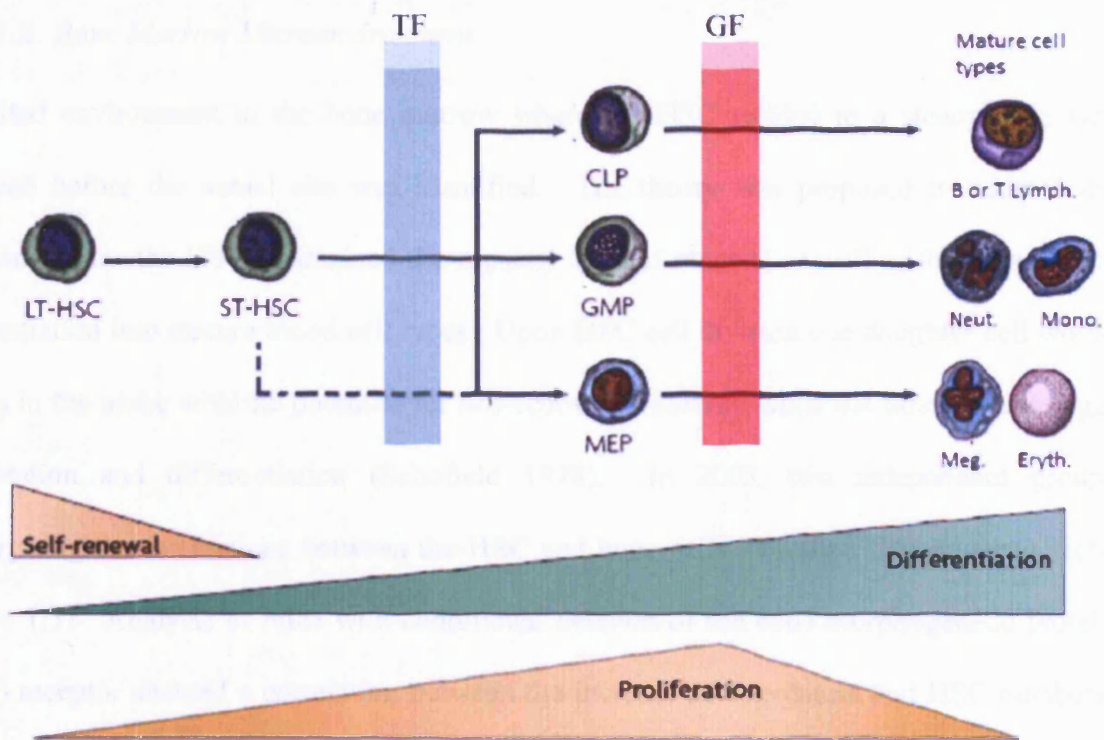
Figure 1.2 Self-Renewal and Differentiation in Haematopoiesis
A simplified schema demonstrating the balance between self-renewal and differentiation in normal haematopoiesis. As cells progress through the stages of lineage commitment, the capacity for self-renewal is lost while differentiation potential is increased. Early lineage decisions are prescribed by transcription factors (TF) while terminal differentiation is promoted by growth factors (GF) or cytokines. LT-HSC: long term heamtopoietic stem cell; ST-HSC: short term haematopoietic stem cell; CLP: common lymphoid progenitor; MEP: megakaryocyte/erythyroid progenitor; GMP: granulocyte/macrophage progenitor; RBCs:red blood cells adapted from (Rosenbauer & Tenen 2007).

1.1.1.b.ii Bone Marrow Microenvironment

A limited environment in the bone marrow where the HSC resided in a steady state was proposed before the actual site was identified. The theory was proposed to address the question of how the HSC maintained the capacity for self-renewal as well as the capacity for differentiation into mature blood cell types. Upon HSC cell division one daughter cell would remain in the niche with the potential for self-renewal preserved while the other would begin proliferation and differentiation (Schofield 1978). In 2003, two independent groups investigating the interactions between the HSC and bone cells identified the endosteal niche (Figure 1.3). Analysis of mice with conditional deletion of the bone morphogenetic protein (BMP) receptor showed a correlation between the increase of osteoblasts and HSC numbers. The attachment of HSCs to osteoblasts was thought to be mediated by N-cadherin (Zhang *et al.* 2003). The authors suggested that the osteoblasts lining the niche supported the HSCs while BMP regulated HSC numbers through control of the niche size. In another murine model, stimulation of osteoblasts with parathyroid hormone receptors generated raised levels of the Notch ligand jagged 1, which in turn promoted an expansion in the number of HSCs exhibiting Notch activation (Calvi *et al.* 2003). Furthermore, injection of parathyroid hormone in wild type mice caused an increase in stem cell production and enhanced post-transplant survival. This indicated osteoblasts within the endosteal niche promote an increase in HSC numbers. The receptor tyrosine kinase Tie2 and its ligand angiopoietin-1 promoted quiescence and anti-apoptosis in HSCs in addition to enhanced bone adhesion (Arai *et al.* 2004). However, the adherence function of osteoblasts, as well as the necessity of osteoblasts for the maintenance of HSC populations has been called into question (Kiel, Radice, & Morrison 2007). Recently, a vascular niche composed of sinusoidal endothelial cells has been proposed (Kiel *et al.* 2005). Whether there are two independent niches performing separate functions or one common niche that is a composite of the two has yet to be determined. The

functions of the niche are clinically relevant to bone marrow transplantation and resistance to chemotherapy as quiescent cells are resistant to the actions of cell cycle specific drugs.
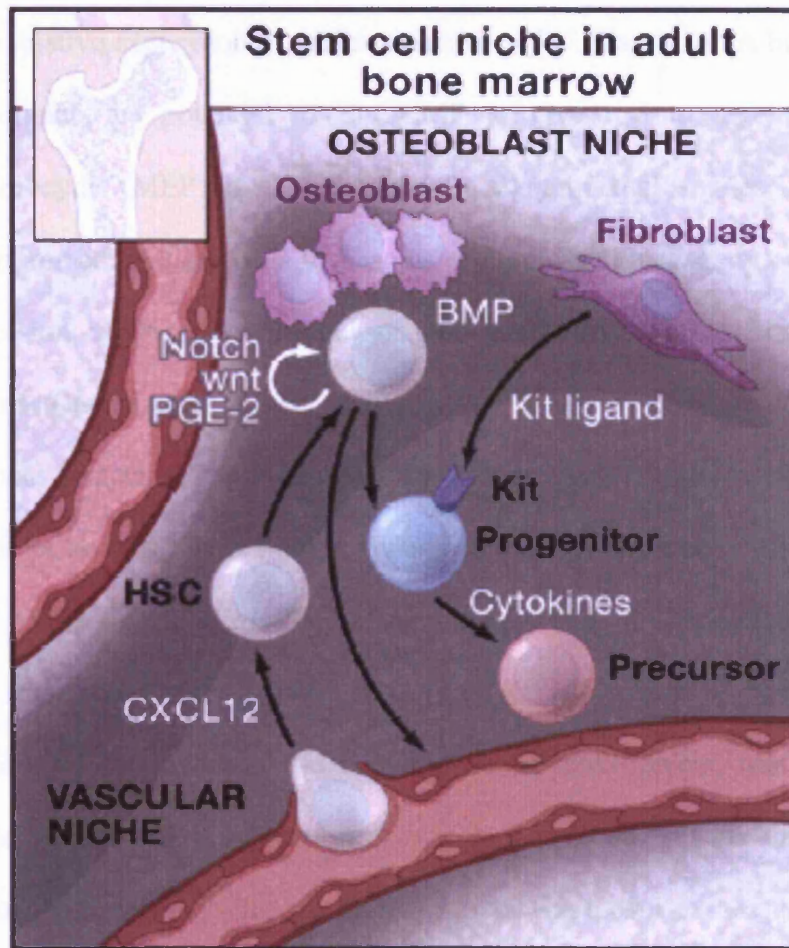
Figure 1.3 Haematopoietic Stem Cell Niche
Haematopoietic stem cells are next to osteoblasts in the endosteal niche and adjacent to blood vessels in the vascular niche. These niches may be intertwined within the bone marrow. Stromal cells support haematopoiesis through adhesion sites for HSCs and the secretion of cytokines that influence stem cell and progenitor function and survival. HSC: haematopoietic stem cell; BMP: bone morphogenetic protein (Orkin & Zon 2008).

1.1.1.b.iii Role of Transcription Factors

Haemopoietic transcription factors control the maintenance of HSCs, including self-renewal, and co-ordination of lineage specific development (Orkin & Zon 2008). Interestingly, many transcription factors have functions within the HSC in addition to the lineage restriction roles, which are important later in differentiation. Furthermore, there is often interplay of transcription factors in the regulation of cell fate. For example, over expression of *PU.1* promotes the commitment of HSC to a myeloid stem cell (or CMP).

Subsequently, the relative expression of purine rich box-1 (*PU.1*) and GATA binding protein-1 (*GATA-1*) determine the potential of the CMP to give rise to myeloid (GMP) or erythroid/megakaryocytic (MEP) differentiation. In experimental studies, *GATA-1* over expression reprogrammed myeloid cells to become erythroid and megakaryocyte cells while *PU.1* over expression repressed erythropoiesis and promoted myeloid differentiation in erythroid cell lines (Cantor *et al.* 2008; Iwasaki *et al.* 2003; Yamada *et al.* 2001). The proposed model was that the protein in excess prevented gene activation by blocking the binding sites of the competing factor's target genes while instigating a positive autoregulatory loop (Chen *et al.* 1995; Nerlov *et al.* 2000; Zhang *et al.* 2000) In another example, down regulation of the transcription factor friend of *GATA-1* (*FOG*), which is required to be co-expressed with *GATA-1* for erythroid and megakaryocytic development, was necessary for *GATA-1* to promote mast cell development (Orkin & Zon 2008). Thus the level of the transcription factor is important in lineage determination where over expression is shown to be antagonistic to the influence of competing factors (Orkin & Zon 2008).

The homeobox (*HOX*) gene family encodes transcription factors that have specific functions, including regulation of embryonal development and cell differentiation in haematopoiesis, and recent findings suggest a role in HSC self-renewal (Argiropoulos & Humphries 2007). *HOX* genes were originally identified as regulators of the anterior-posterior segmentation in *Drosophila* development and later discovered to have human homologues (Krumlauf 1994). The gene is organised into four genomic clusters labelled A-D, which are further divided into numbered members of the clusters termed paralogues. Many of the *HOX* gene subtypes in clusters A-C are expressed during haematopoiesis. The genes are primarily expressed in HSC and early progenitor cells and expression subsequently decreases as the lineage commitment of the cell progresses (Pineault *et al.* 2002; Sauvageau *et al.* 1994). Individual paralogues can increase the population of progenitors. Over expression

1-12

of *Hoxb4* and *Hoxa9* cause expansion of the murine HSC population and in the case of *Hoxa9*, leads to leukaemogenesis (Antonchuk, Sauvageau, & Humphries 2001; Thorsteinsdottir *et al.* 2002). The number of different gene paralogues suggests the importance of the gene may be protected through redundancy of gene function. Redundancy has been demonstrated between some paralogues but appears to be stage specific. For example, the recovery of HSC progenitor growth in cells lacking *Hox* expression was mediated by individual expression of *Hoxa9* or *Hoxb4* genes, however the subsequent differentiation was unique to the gene expressed (Ernst *et al.* 2004b). The mixed lineage leukaemia gene (*MLL*) maintains the expression of the *HOX* genes possibly through interaction with the *HOX* promoter region and is a key regulator of haematopoietic development.

Evidence that *MLL* is necessary for definitive haematopoiesis was shown in mouse studies using *Mll* deficient cells. $Mll^{/-}$ cells were unable to generate HSC or progenitor cell populations in foetal liver chimeras and transplantation of $Mll^{/-}$ embryonic cells into sublethally irradiated recipients failed to demonstrate HSC activity. Furthermore, a reduced number of recipients engrafted with $Mll^{+/-}$ were repopulated with donor cells indicating a critical level of *Mll* expression is needed for haematopoiesis to occur (Ernst *et al.* 2004a). Conditional deletion of *Mll* resulted in a loss of self-renewal capacity in the HSC population (Jude *et al.* 2007; McMahon *et al.* 2007). Quiescent HSC populations were maintained in *Mll* deficient cells, while there were contrasting effects on downstream cells, with proliferation of myeloid-erythroid progenitors reduced and no expansion of committed lineage cells observed (Jude *et al.* 2007). The role of *MLL* as a regulator of *HOX* expression is the probable reason for the decrease in proliferation, with re-expression of a single *Hox* gene capable of reinstating the haematopoietic process. There is less evidence for the role of *Hox* involvement in the quiescent status of the HSC population, as this result has not been reported in other *Hox*

knock-out studies (Brun *et al.* 2004; Scott *et al.* 1994). These studies demonstrate that *MLL* plays an instrumental role in haematopoietic development, with multiple functions dependent on the timepoint and cellular context of the expression. The *MLL* gene is discussed further in Section 1.2.3.a. The importance of transcription factors to haematopoiesis is perhaps best exemplified by the high number linked with leukaemia, which is discussed in Section 1.1.2.b.

.

1.1.1.b.iv Role of Cytokines

Cytokines are glycoproteins that influence haematopoiesis by binding to cytokine receptors on target cells to activate cell signalling pathways. The majority of cytokines are synthesised in T lymphocytes, monocytes and bone marrow stromal cells, apart from erythropoietin and thrombopoietin, which are produced in the kidney and liver respectively. Cytokines can remain membrane bound or be secreted allowing them to function in the local environment or circulate to act systemically. The function of individual cytokines is often difficult to ascertain due to cytokine pleiotrophy, redundancy and synergy in addition to cell type (context-specific) responses. The colony stimulating factors (CSF) for macrophage (M-CSF), granulocyte (G-CSF), granulocyte-macrophage (GM-CSF) and multiple colonies (multi-CSF, more commonly referred to as interleukin-3 (IL-3)), are so named due to the lineage of the colony produced by progenitors in *in vitro* assays; these factors play an important role in the development of the mature cells of the myeloid lineages. Straightforwardly, G-CSF and M-CSF are lineage specific cytokines that contribute to the proliferation and terminal differentiation of granulocytes and macrophages, whereas IL-3 modulates many lineages including macrophage, neutrophil, basophil, eosinophil, erythrocyte and mast cells. Cytokines that may also contribute to lineage commitment at earlier stages of haematopoiesis include IL-3, GM-SCF, fms-like tyrosine kinase 3 ligand (FL) and stem cell factor (SCF or c-kit ligand) (Blalock *et al.* 1999).

FL expression is involved with proliferation, differentiation, survival and mobilisation in haematopoietic precursor cells. FL has two forms, soluble and membrane bound, and both are able to promote proliferation of HSCs. Although FL alone was a weak stimulator of growth, it exhibits synergistic promotion of growth in CD34$^+$ BM cells when combined with cytokines SCF and IL-3 (Lyman *et al.* 1993; Rusten *et al.* 1996). Disrupting FL expression with murine knockout studies resulted in reduced overall leukocyte numbers in the BM, and significant reduction of myeloid and B-lymphoid progenitors indicating a role for FL in haematopoiesis (McKenna *et al.* 2000). FL is the cognate ligand for the fms-like tyrosine kinase 3 (*FLT3*) receptor, which is expressed on myeloid and lymphoid progenitor cells (though the level of expression decreases with differentiation). Activating mutations of *FLT3* occur in over 30% of all acute myeloid leukaemia (AML) patients and the *FLT3* ITD in particular is associated with poor prognosis (Section 1.2.4.b and 1.2.6.b) (Kottaridis *et al.* 2001).

SCF and FL are not only similar in size and overall protein structure, but they also share a number of functions. SCF expression promotes survival, mobility and proliferation of HSCs. SCF is involved with the prevention of apoptosis and promotes proliferation of HSCs through interaction with the cytokine thrombopoietin (Hassan & Zander 1996; Ku *et al.* 1996). Inhibition of the SCF receptor (c-kit) causes pancytopenia and decreased cellularity in the marrow suggesting that SCF is needed for normal haematopoiesis (Ogawa *et al.* 1991). Specifically, SCF promotes the expansion of myeloid progenitors throughout differentiation and enhances erythropoiesis (Lyman *et al.* 1994). The membrane bound SCF is associated with HSC adhesion in the bone marrow niche; cleavage of this form decreases adhesion resulting in HSC mobilisation (Heissig *et al.* 2002; Kollet *et al.* 2006). The cognate receptor for SCF (SCF-R) is the product of the c-kit oncogene. Mutations in the c-kit gene are

associated with core binding factor (*CBF*) AML, and appear to have a negative impact on these favourable risk cytogenetic groups (Paschka *et al.* 2006; Schnittger *et al.* 2006)

Cytokine signals are transmitted to the nucleus via binding to cognate receptors located on the membrane of the target cell. Activation of the receptor initiates a signalling pathway that ends in the nucleus with the activation or inactivation of a gene. This results in a change in the target cell programming which may include entering the cell cycle, commencement of differentiation, resistance to apoptosis or becoming functionally active (Hibbert & Johnston 2001; Olsson *et al.* 1996). All cytokine receptors are transmembrane glycoproteins with the common features of an extracellular amino terminal ligand binding domain, a hydrophobic transmembrane domain and a carboxy-terminal intracellular domain. Activation of the signalling pathways within the cell is generally regulated by phosphorylation of tyrosine residues which in turn can influence transcriptional processes. Within the cytokine receptor families this is accomplished either with a cytoplasmic domain that interacts with the Janus kinases (JAK) (class I) or by a tyrosine kinase domain in the receptor itself (tyrosine kinase superfamily). A cascade of phosphorylation events leads to the activation of transcription factors in the nucleus. For example, in the class I family of cytokine receptors, the IL-3 receptor binds to its cognate ligand and dimerises to enable binding and activation of JAK2. JAK2 mediates the intracellular phosphorylation of the tyrosine residues within the receptor creating SH2 binding sites. STAT5 signalling proteins are subsequently recruited to these sites and are phosphorylated thus allowing STAT5 to migrate into the nucleus where it initiates transcription. Creation of SH2 binding sites can also result in the recruitment of proteins leading to the initiation of other pathways including the mitogen-activated protein kinase and phosphoinositol-3 kinase pathways (Ihle & Kerr 1995). The tyrosine kinase receptors (*FLT3*, c-kit (SCF-R) and M-CSFR) act in a similar

fashion, with signalling proteins binding to the autophosphoylated tyrosines to promote proliferation and survival traits in the target cell.

1.1.2   Myeloid Leukaemogenesis

The hallmarks of leukaemia are uncontrolled proliferation of haematopoietic cells and a loss of terminal differentiation. Such disruption in the normal flow of haematopoiesis can result in several types of leukaemia, where developmental arrest can occur in different lineages with various extents of differentiation (Sachs 1986). Interestingly, myelodysplasia (MDS) and myeloproliferative neoplasms (MPN) are haematological disorders that demonstrate only a subset of the characteristics of leukaemia. MDS exhibits a dysregulated differentiation, whereas MPN is characterised by increased cell proliferation but with terminal differentiation; both disorders are able to progress to acute leukaemia. Leukaemias are grouped into acute and chronic categories, dependent upon the patient's symptoms (Sawyers, Denny, & Witte 1991). The subject of this study, the *MLL* partial tandem duplication (PTD), occurs primarily in AML therefore the remainder of the introduction will focus on this disease.

1.1.2.a   Multiple Mutations Needed for Leukaemogenesis

Despite the differences in the proposed progression of leukaemia, it is generally agreed that the characteristics of self-renewal and lack of differentiation are the result of an accumulation of multiple mutations within a cell causing an imbalance in the regulatory systems and resulting in a neoplastic phenotype. Mutations that confer the capacity for self-renewal, if not possessed by the cell originally, are essential for the development of cancer (Hanahan & Weinberg 2000). Deregulation of cell maintenance and DNA repair genes as

caused by AML fusion proteins promote further accumulation of abnormalities and contribute to multistep leukaemogenesis (Alcalay *et al.* 2003; Suela *et al.* 2007). As with all cancers, the balance of proto-oncogenes and tumour suppressors plays an important part in the onset of disease, with multiple mutations necessary before leukaemia is established (Cline 1994; Rabbitts 1991). A landmark paper published by Knudson (1971) demonstrated the multistep nature of carcinogenesis in a study of hereditary retinoblastoma. The study compared the onset time of retinoblastoma in hereditary and non-hereditary cases. The tumours in the hereditary cases that required only one additional mutation occurred much earlier than those in the non-hereditary cases that needed two co-occurring mutations (Knudson, Jr. 1971). Assuming a constant mutation rate, the shorter onset in the hereditary patients was in agreement with the multiple-hit theory of cancer. The mutations involved in the molecular pathogenesis of AML are discussed in more detail below (Section 1.2.3-5).

1.1.2.b   Leukaemia Initiating Cells

Models of leukaemogenesis are varied and provoke debate among the haematological community, with most models dependent upon the existence of the leukaemic stem cell (LSC), an HSC that developed leukaemic potential from the accumulation of oncogenetic insults. However, the existence of such a cell has been questioned with the proposal that more mature blood cells can regain stem cell characteristics thus enabling them to generate the disease (Passegue *et al.* 2003).

The hypothesis of leukaemia as a disease that requires an accumulation of mutations over time suggests that the most likely candidate cells would develop from a long-lived population like HSCs (Figure 1.4) (DePinho 2000). In support of this, long term (LT) HSCs in populations of older mice demonstrated increased self-renewal, decreased lymphoid potential and raised expression of genes associated with leukaemic transformation (Rossi *et*

al. 2005). A widely accepted paradigm of leukaemia is its derivation from a single cell and maintenance by a subpopulation of LSCs (Hope, Jin, & Dick 2004; McCULLOCH *et al.* 1979). Evidence of leukaemic abnormalities (eg. runt-related transcription factor 1 *(RUNX1)*-runt-related transcription factor 1; translocated to, 1 *(RUNX1T1)* breakpoint cluster region *(BCR)* c-abl oncogene 1, receptor tyrosine kinase *(ABL1)*) present in both lymphoid and myeloid cells indicate that the mutation can occur in a multilineage stem cell or HSC (Miyamoto, Weissman, & Akashi 2000; Secker-Walker & Craig 1993). Similarities between HSC and LSC include the state of quiescence. Populations of quiescent primitive progenitor cells have been detected in AML patients with up to 96% of the LSC in the G_0 phase of the cell cycle (Guan, Gerhard, & Hogge 2003; Guzman *et al.* 2001). Quiescent LSCs could explain the subsequent relapse of patients who had successfully achieved remission, as these cells are unaffected by cell cycle specific chemotherapy agents.
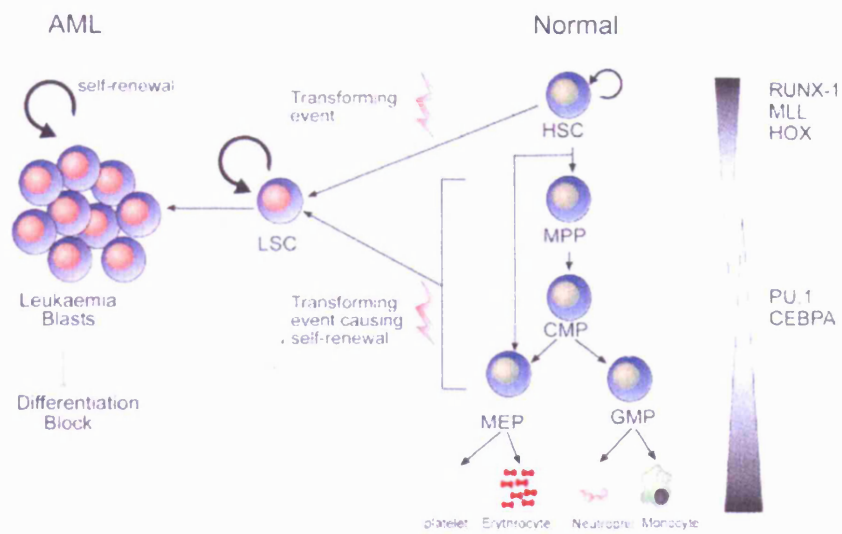
Figure 1.4 Leukaemia Stem Cell (LSC) Induction
The schematic diagram demonstrates normal myeloid differentiation and the transcription factors involved in self-renewal and differentiation. The lightning strikes indicate possible sites of transformation within the stem cell and progenitor stages of normal development. HSC: haematopoietic stem cell; MPP: myeloid pluripotent progenitor; CMP: common myeloid progenitor; MEP: megakaryocyte/erythroid progenitor; GMP: granulocyte/macrophage progenitor. (Gudgin.E. 2009)

The ability to test the LSC concept further came with the development of mouse

xenotransplant models of severe combined immunodeficiency (SCID) and non-obese diabetic

severe combined immunodeficiency (NOD/SCID). BM from AML patients was transplanted

into SCID and NOD/SCID mice and the resulting leukaemic grafts were found to be

representative of the patient's original disease. A population of cells termed SCID-leukaemia

initiating cells (SL-IC) had significant self-renewal capacity determined through serial

dilution studies. Purification of the transplanted fractions identified the $CD34^+/CD38^-$ as the

fraction containing the SL-IC. No engraftment was seen when the more differentiated

fraction $CD34^+/CD38^+$ was transplanted (Bonnet & Dick 1997; Lapidot et al. 1994). A recent

study has identified a potential problem in the NOD/SCID mice used for xenograft

experiments (Taussig et al. 2008). The mice were found to retain an immune response against

the CD38 antibody, which inhibited the engraftment of $CD38^+$ fractions. When the mice were

treated with immunosuppressive antibodies, $CD38^+$ fractions were successfully engrafted,

suggesting that the number of cell types capable of initiating leukaemia in SCID mice was

previously underestimated. The demonstration that LSCs were equivalent to colony forming

cells (CFCs) in murine studies using MLL-AF9 also led to the conclusion that LSCs, at least

in the MLL driven model, are much more abundant than estimated by xenograft models

(Somervaille & Cleary 2006). This new model suggested one in four cells are LSCs,

compared to one in 10^4 by xenograft modelling. Additionally, the impact of the

microenvironment upon the engrafted cells may have been underestimated. Extrinsic factors

needed for potentially leukaemic human cells to engraft may have been unavailable in the

mouse marrow microenvironment, leading to over estimation of the non-leukaemic population

(Rosen & Jordan 2009).

Data from other studies indicated that more committed cells may be able to initiate

leukaemia (Figure 1.4). An early study of AML patients suggested a heterogeneous origin of

the disease (Fialkow *et al.* 1981). Some patients exhibited differentiation restricted to the granulocyte-monocyte pathway, with normal erythroid development still possible, while in other cases, the disease affected stem cells with potential to differentiate into granulocytes, red cells and platelets. In experimental models, a number of translocations (*MLL-AF9* [ALL-1 gene fused to chromosome 9, also called *MLLT3*], *MLL-ENL* [eleven-nineteen leukaemia also called *MLLT1*] and *MOZ-TIF2* [monocytic leukaemia zinc finger-transforming growth-β induced factor 2]) introduced into CMP and GMP cells promoted leukaemia in murine models (Cozzio *et al.* 2003; Huntly *et al.* 2004; Krivtsov *et al.* 2006). In more recent examples, mice deficient in the translation of CCAAT/enhancer binding protein alpha (Cebpa) p42 demonstrated increased proliferation of the committed myeloid progenitors leading to AML. The GEP signature of the cells was shared with progenitors transformed with *MLL-AF9* suggesting a common pathway system for progenitor transformation (Kirstetter *et al.* 2008). Furthermore, Somervaille *et al.* used *MLL* fusion genes to identify a transcriptional program that results in the generation of myeloid leukaemia stem cells from myeloid progenitor cells. The *MLL* fusion genes exhibited significantly different LSC frequencies as determined by assessment of colony-forming cells in the spleen and bone marrow of leukaemic mice. The fraction of cells with the highest proportion of LSCs also contained the highest number of cycling cells indicating that they were proliferating, metabolically active cells (Somervaille *et al.* 2009). These high LSC frequency MLL cell populations were associated with the expression of embryonic stem cell-like program genes including the transcription/chromatin regulating genes high-mobility group box 3 (*Hmgb3*), myeloblastosis viral oncogene homolog *Myb* and chromobox homolog 5 (*Cbx5*). Co-expression of these three genes in BM progenitor cells enabled clonogenic activity in replating assays (Somervaille *et al.* 2009). Interestingly, the *MLL* transformation pathway of *Hoxa* and Meis homeobox 1 (*Meis1*) was

not utilised in the development of these clonal populations, suggesting a redundancy in the process of LSC development within *MLL* leukaemia..

In summary, mutations affecting both HSCs and committed progenitor cells have been demonstrated to give rise to myeloid leukaemia. As for the presence of an LSC, these cells may be better defined by an "LSC state" instead of a more strictly defined "LSC phenotype", which may not exist due to the diversity of the cell types involved in leukaemia (Ravandi & Estrov 2006).

1.2   Acute Myeloid Leukaemia

Leukaemias are a family of haematological disorders caused by mutations that alter the normal progression of haematopoiesis. In AML, accumulated mutations eventually give rise to a malignant clone of the myeloid lineage with altered proliferation, differentiation and survival characteristics; this results in a rapid increase of immature blasts in the bone marrow (BM) and peripheral blood (PB). AML is the most common form of acute leukaemia for adults, but it remains a rare disease with an incidence of approximately 2.5-3 cases per 100,000 per year worldwide (Swerdlow *et al.* 2008). However, the likelihood of developing AML increases with age, with a median age at presentation of 70 years, and an incidence of 18 per 100,000 per year (Estey & Dohner 2006).

Patients with AML usually present with symptoms resulting from either bone marrow failure and/or organ infiltration caused by an excess of blast cells. Bone marrow failure with loss of terminal myeloid differentiation can lead to anaemia, neutropenia, and thrombocytopenia causing fatigue, infection, fever and excessive bleeding. Patients with high white blood cell count may complain of bone pain caused by the increased pressure in the bone marrow and in extreme cases may suffer from respiratory distress and altered mental

state (leukostasis) requiring immediate medical attention. In rare cases, AML manifests as an extramedullary tumour mass consisting of myeloid blasts and is termed a myeloid sarcoma (Pileri *et al.* 2007).

1.2.1 Aetiology of AML

Although there have been a number of risk factors linked with the development of AML, the majority of *de novo* cases have no known aetiology (Sandler & Collman 1987). Cases of AML with known aetiology include patients with a previous history of exposure to occupational/environmental mutagens such as radiation or benzene; or cases of secondary AML arising from an existing haematological disorder (e.g. MPN or MDS), or exposure to chemotherapy (also termed therapy-related AML). Some therapy-related AML patients are believed to be predisposed to errors in the repair of mutations caused by cytotoxic therapy due to defects in the mismatch repair system, however, for most patients the method of pathogenesis remains unclear (Rund *et al.* 2005). Particular cytogenetic abnormalities have been shown to have an association with specific drugs. For example, alkylating agents are associated with monosomy 7 and monosomy 5; epipodophyllotoxins with balanced translocations at 11q23 and topoisomerase II inhibitors with balanced translocations t(8;21), inv(16), and 11q23 (Odero *et al.* 2001; Pedersen-Bjergaard *et al.* 2002) (Pui *et al.* 1989; Quesnel *et al.* 1993). The rate of cure for secondary AMLs (sAMLs) is considerably worse than *de novo* AML with a survival rate of less than 10% (Section 1.2.2 and 1.2.6.b) (Swerdlow *et al.* 2008).

Congenital conditions can also be associated with increased risk of AML. In Down's syndrome (DS), children under five, AML incidence has been estimated to be up to 150 fold greater than in children without DS, with acute megakaryoblastic leukaemia the predominant classification (Swerdlow *et al.* 2008). Additionally, children with DS often suffer from transient abnormal myelopoiesis (TAM), a disorder that is virtually indistinguishable from the

most frequent type of DS AML. TAM resolves spontaneously over a short time period but may recur as non transient AML 1-3 years later (Zipursky *et al.* 1999). Both the non transient form of AML and TAM are often associated with *GATA-1* mutations, and this mutation confers a better prognosis in children without DS who have AML (Greene *et al.* 2003; Lange *et al.* 1998).

1.2.2   Classifications of AML (WHO)

The classification of AML disease has changed tremendously in the past 35 years, with the basis for distinction progressing from morphological differences to cytogenetic aberrations and, most recently, including molecular genetic mutations.

Historically, the classification has been achieved using the dominant and most common method of cell classification; cell morphology. The heterogeneity of AML morphology was recognised in 1976 with the introduction of the French-American-British (FAB) classifications (Table 1.1) (Bennett *et al.* 1976). This system grouped AMLs according to the level of differentiation identified by observing the morphology of the leukaemic cells. The eight categories describe different types and extents of differentiation, starting with the primitive cells of M0 which show no differentiation, to the more differentiated groups such as M6 and M7 which contain blast cells derived from erythroid and megakaryocyte lineages respectively. The FAB system provided the clinician with a clear morphological picture of the patient's disease, yet the clinical response of patients was varied within individual FAB classifications. The discovery of consistent cytogenetic abnormalities such as t(8;21), 11q23 translocations, and more complex karyotypes demonstrated genetic diversity within the FAB groups and provided insight to the genetic pathogenesis of AML subtypes. The diagnostic cytogenetics of AML were recognised as independent prognostic

factors and led to the inclusion of AML subtypes defined by cytogenetic abnormalities in the 3rd edition of the World Health Organisation (WHO) Classification of Tumours of the Haematopoietic and Lymphoid Tissues published in 2001 (Section 1.2.6.b) (Vardiman *et al.* 2002, Grimwade *et al.* 1998). This was the first time that genetic information was incorporated into the WHO myeloid neoplasm classification system, which previously was the result of information compiled from morphology, immunophenotype, cytochemistry, and clinical features to create distinct clinically relevant disease entities (Vardiman *et al.* 2009). The importance of submicroscopic genetic mutations recently found to contribute to leukaemogenesis is such that the fourth edition of the WHO Classification of Tumours of the Haematopoietic and Lymphoid Tissues has introduced new provisional entities into the disease classifications based principally on genetic features.

FAB	AML
M0	Minimally differentiated myeloblastic leukaemia
M1	Myeloblastic leukaemia without maturation
M2	Myeloblastic leukaemia with maturation
M3	Hypergranular promyelocytic leukaemia
M4	Myelomonocytic leukaemia
M4Eo	Myelomonocytic leukaemia with eosinophils
M5	Monocytic leukaemia
M6	Erythroleukaemia
M7	Megakaryocytic leukaemia

Table 1.1 French American British (FAB) Classification of Acute Myeloid Leukaemia (Bennett *et al.* 1976)

The current WHO AML classification can be seen in (Table 1.2). The four main subgroups are: AML with recurrent cytogenetic abnormalities, AML not otherwise specified (NOS), AML with myelodysplastic-related changes and therapy related myeloid neoplasms. Existing groupings within "AML with recurrent genetic abnormalities" were modified and

two new "provisional entities", AML with mutated nucleophosmin (*NPM1*) and AML with mutated *CEBPA*, were added for the 4th edition of the WHO classifications. The current definition of AML is a *de novo* myeloid neoplasm containing more than 20% blasts in the PB or BM, where "myeloid" is classified as granulocytic, erythroid, monocytic/macrophage, megakaryocytic or mast cell lineages. The cell characteristics used to ascertain the lineage are determined by a combination of methods including morphology, immunophenotyping and cytochemistry.

Acute Myeloid Leukaemia and Related Myeloid Neoplasms

Acute myeloid leukemia with recurrent genetic abnormalities

    AML with t(8;21)(q22;q22); RUNX1-RUNX1T1

    AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); CBFB-MYH11

    APL with t(15;17)(q22;q12); PML-RARA

    AML with t(9;11)(p22;q23); MLLT3-MLL

    AML with t(6;9)(p23;q34); DEK-NUP214

    AML with inv(3)(q21q26.2) or t(3;3)(q21;q26.2); RPN1-EVI1

    AML (megakaryoblastic) with t(1;22)(p13;q13); RBM15-MKL1

    Provisional entity: AML with mutated NPM1

    Provisional entity: AML with mutated CEBPA

Acute myeloid leukemia with myelodysplasia-related changes

Therapy-related myeloid neoplasms

Acute myeloid leukemia, not otherwise specified

    AML with minimal differentiation

    AML without maturation

    AML with maturation

    Acute myelomonocytic leukemia

    Acute monoblastic/monocytic leukemia

    Acute erythroid leukemia

        Pure erythroid leukemia

        Erythroleukaemia, erythroid/myeloid

    Acute megakaryoblastic leukemia

    Acute basophilic leukemia

    Acute panmyelosis with myelofibrosis

Myeloid sarcoma

Myeloid proliferations related to Down syndrome

    Transient abnormal myelopoiesis

    Myeloid leukemia associated with Down syndrome

Blastic plasmacytoid dendritic cell neoplasm

Table 1.2 Acute Myeloid Leukaemia and Related Myeloid Neoplasms
World Health Organisation classifications for myeloid neoplasms (Swerdlow *et al.* 2008).

The number of subtypes of AML is reflected by the variability in the characteristics seen in Table 1.3. The strength of the association between AML and t(8;21)(q22;q22), inv(16)(p13.1q22), t(16;16)(p13.1;q22) and t(15;17)(q22;q12), dictates that patients with these translocations are diagnosed as having AML even when the blast count is <20%. Previously all *MLL* translocations were included in the group AML with 11q23 abnormalities, the most common *MLL* translocation t(9;11)(p22;q23); *MLLT3-MLL* has now been recognised as a distinct clinical entity with intermediate prognosis. However, all other balanced translocations of *MLL* should be specified in the diagnosis (Vardiman *et al.* 2009).

Characteristic Diagnostic Features of AML	
BM Cellularity	Usually increased
% Marrow Blasts	Increased, ≥20%, except in some cases with specific cytogenetic abnormalities or in some cases of erythroleukaemia.
Maturation	Varies, usually minimal
Morphology	May or may not be associated with dysplasia in one or more lineages
Haematopoiesis	Ineffective or effective
Blood Counts	WBC variable, usually anaemia and thrombocytopenia
Organomegaly	Uncommon

Table 1.3 Characteristic Diagnostic Features of AML
The biological heterogeneity of AML is apparent in the variability of the diagnostic characteristics displayed. WBC: white blood cells (Swerdlow *et al.* 2008).

Cases of AML that do not qualify for any other subgroups are placed under the heading AML NOS, where they are subdivided by the level of differentiation as in the original FAB system. As the number of mutations discovered increases and more genetic entities are created the proportion of patients assigned to AML NOS would be expected to decrease (Swerdlow *et al.* 2008).

Patients who do not present with *de novo* AML are divided into two groups: AML with myelodysplastic changes and therapy-related myeloid neoplasm (Section 1.2.2). Therapy-related myeloid neoplasms are no longer grouped by the type of drug administered *i.e.* alkylating agents, topoisomeras II inhibitors or therapeutic radiation, because many patients have received combinations of these agents (Swerdlow *et al.* 2008). Despite the frequent observation of genetic mutations also observed in *de novo* AML, patients with therapy-related AML (t-AML) are reported to have a significantly worse rate of survival than those with the identical mutation in *de novo* AML with the exception of favourable prognoses conferred with t(8;21)(q22;q22), inv(16)(p13.1q22), t(16;16)(p13.1;q22) and t(15;17)(q22;q12). This may indicate biological differences between t-AML and *de novo* AML and supports the maintenance of a t-AML subgroup (Rowley & Olney 2002; Smith *et al.* 2003).

The mutations associated with the new provisional diagnostic entities have had considerable impact on the basic understanding of AML; they also have valuable clinical use as prognostic indicators especially in the context of CN AML, thus justifying the subgroup status (Section 1.2.6.b). As provisional entities, the two mutations are classified in the scheme with the understanding that further characterisation is required. The difficulty with the new mutations is that they are not clearly identified with consistent morphological or clinical characteristics, nor are they mutually exclusive of cytogenetic abnormalities or each other. In fact, one of the most prognostically significant gene mutations, *FLT3*, is not

considered a distinct entity due to the number of associations with cytogenetic abnormalities in previously established groups such as t(6:9) and t(15;17), in addition to overlap with other molecular mutations (Thiede *et al.* 2006; Vardiman *et al.* 2009).

1.2.3 Cytogenetic Mutations in AML

1.2.3.a The MLL gene and MLL translocations

The *MLL* gene and translocations are covered in detail to familiarise the reader with the gene that is the focus of this thesis. The *MLL* gene is also known as *HRX*, *ALL-1* and *Htrxl* on account of its similarity with the Drosophila trithorax gene and due to its independent cloning in several different laboratories in 1991 and 1992 (Cimino *et al.* 1991; Ziemin-van der *et al.* 1991) (Tkachuk, Kohler, & Cleary 1992). The Drosophila trithorax homologue gene is involved in the regulation of developmental processes in the organism. The *MLL* gene is 100 kb long with 37 exons (Nilson *et al.* 1996); it is widely expressed in cell types throughout the body, including haemopoietic cells (Butler *et al.* 1997). The MLL protein has multiple regions that confer significant functional activity including DNA binding domains, transcriptional activation and repression domains, plant homeodomain (PHD) finger domain (associated with chromatin remodelling), self-association motifs and a C-terminal variegation/enhancer of zeste/trithorax domain (SET) (Figure 1.5). The ability to bind DNA is found in the adenosine-thymine (AT) hooks and methyl transferase (MT) region of the protein and could be involved in transcriptional regulation via direct deoxyreibonucleic acid (DNA) binding. MLL may mediate transcriptional repression through recruitment of repressor complexes such as histone deacetylases (HDAC) or polycomb group proteins via the MT domain (Xia *et al.* 2003).
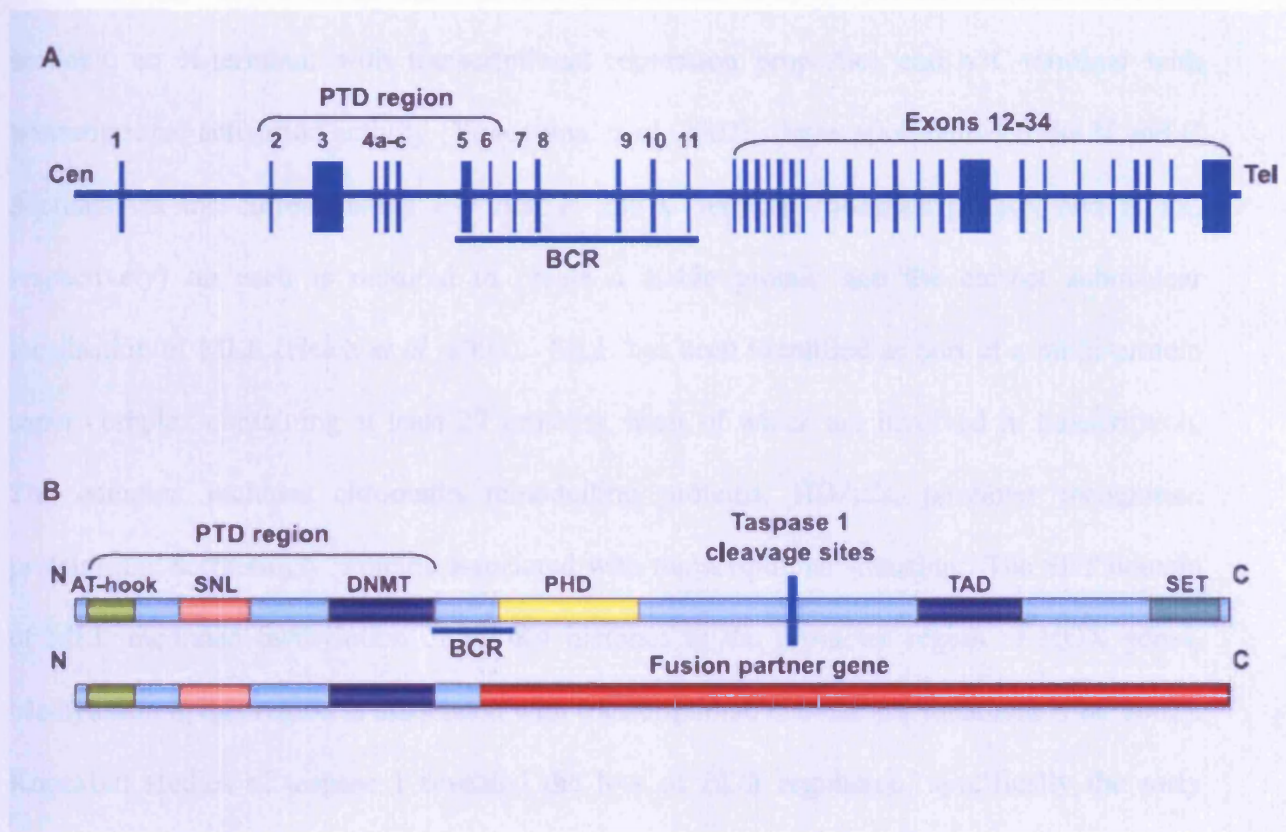
Figure 1.5 Structure of the *MLL* Gene and Protein
A. The schematic representation (exon and intron sizes not to scale) of the *MLL* gene shows
the location of the partial tandem duplication (PTD) and the breakpoint cluster region (*BCR*)
where translocations occur. B. The protein domains affected by the PTDs and translocations
are shown as well as the taspase 1 cleavage site. The MLL protein domains include AT-hook
(DNA binding motif which binds adenosine-thymidine rich DNA), speckled nuclear
localisation sites (SNL), DNA methyltransferase domain (DNMT), plant homeodomains
(PHD), transactivation domain (TAD), and suppressor of variegation/enhancer of
zeste/trithorax (SET) domain (Harper & Aplan 2008).

During maturation, the MLL protein can be cleaved by the protease taspase I into two sections, an N-terminal with transcriptional repression properties and a C-terminal with transcriptional activation activity (Yokoyama *et al.* 2002). Interaction between the N and C peptides via the corresponding F-y rich N and C terminus domains (FYRN and FYRC respectively) on each is required to create a stable protein and the correct subnuclear localisation of MLL (Hsieh *et al.* 2003). MLL has been identified as part of a multi-protein super complex containing at least 27 proteins, most of which are involved in transcription. The complex includes chromatin remodelling proteins, HDACs, promoter recognition proteins and surprisingly proteins associated with transcriptional silencing. The SET domain of MLL mediated methylation of H3-K4 histones in the promoter region of HOX genes. Methylation in this region is associated with transcriptional activation (Nakamura *et al.* 2002). Knockout studies of taspase 1 revealed the loss of *HOX* regulation, specifically the early expression of the 3' *HOX* genes, while the 5' genes were unaffected (Hsieh, Cheng, & Korsmeyer 2003). This is in contrast to the unilateral affects on *HOX* gene expression caused by *MLL* knock-out. Therefore, cleavage by taspase 1 appears to be a mechanism for the regulation of *MLL* and *HOX* genes which is lost in *MLL* fusion genes due the loss of the taspase cleavage site (Gilliland 2003). The amino terminal section of the *MLL* gene is retained after the translocation and can interact through a conserved binding motif with menin, the protein product of the tumour suppressor gene multiple edocrine neoplasia type I (*MEN1*) to promote transcription (Yokoyama *et al.* 2005). As with *MLL,* the down-regulation of menin disrupted the maintenance of *HOX* genes, stopped the differentiation arrest and abrogated the leukaemogenic properties of the *MLL* transformed leukaemic cells.

The *MLL* gene is very promiscuous, with over 70 translocation partners identified at the 11q23 locus; however, the mechanisms of leukaemogenesis promoted by these rearrangements are largely unknown (Harper & Aplan 2008). The frequency of *MLL*

translocations is around 10% in *de novo* AML and greater than 85% in sAML (Look 1997).

All of the translocations occur in an 8.5 kb breakpoint cluster region (bcr) between the exons

8 and 14 (new nomenclature according to Nilson et al.) (Nilson *et al.* 1996). The reported

prognostic significance varies, with some groups reporting an intermediate risk and others

adverse (Dimartino & Cleary 1999; Rubnitz, Behm, & Downing 1996). This may be the

result of differences in prognosis arising from different partner genes. For example, the

clinical outcome in t(9;11) translocation in *de novo* AML is significantly improved compared

to other 11q23 translocations in terms of complete remission duration and event-free survival

(Mrozek *et al.* 1997). For this reason the category of "Acute myeloid leukaemia with

t(9;11)(p22;2q23)" was established in the 2008 WHO classifications (Section 1.2.2).

Furthermore, the refined Medical Research Council (MRC) cytogenetic risk classification has

categorised t(6;11) and t(10;11) in the adverse risk group while leaving the other 11q23

mutations in the intermediate classification (Grimwade & Hills 2009). The translocations

occur in both acute lymphoblastic (ALL) and AML, however there is a bias with t(4;11) most

commonly identified in ALL and t(9;11) seen more often in AML, suggesting a potential

influence of the partner gene on the disease phenotype (Dimartino & Cleary 1999; Mrozek *et

al.* 1997; Tkachuk, Kohler, & Cleary 1992).

The role of the partner gene in *MLL* translocations is an important one, not only when

considering the domains lost by *MLL* but also the characteristics the new gene contributes.

The regulatory domains removed from *MLL* include self-association motifs, SET (a histone

methyltransferase) and transcriptional activation domains, one of which influences the

transcriptional regulation of *HOX* genes. It has been suggested that the MLL fusion proteins

be divided into two groups based on the proposed leukaemogenic mechanism. Class I fusion

proteins involve nuclear proteins that have intrinsic transcriptional activity and promote

aberrant recruitment of transcriptional complexes to downstream *MLL* targets (Prasad *et al.*

1995; Slany, Lavau, & Cleary 1998). This is illustrated by the *MLL* partner gene ALL-1 fused gene from chromosome 10 (*AF10*) (also known as *MLLT10*) which interacts with the histone methyltransferase, DOT1-like, histone H3 methyltransferase (*hDOT1L*). *hDOT1L* associates with AF10 via a conserved region of AF10 that is essential for its leukaemic potential. When MLL was directly bound to *hDOT1L,* the construct up-regulated *HOXA* genes and immortalised murine myeloid progenitor cells. Interaction of MLL with *hDOT1L* whether through association with (AF10*)* or MLL construct, leads to the hypermethylation of H3-K79 and subsequent up regulation of *HOX* gene expression and promotion of leukaemogenesis (Okada *et al.* 2005). On the other hand, class II fusions involve partners able to dimerise yet lacking intrinsic transactivating potential (Martin *et al.* 2003; So *et al.* 2003; So & Cleary 2004). The ability of the fusion gene to dimerise results in *MLL* gene dimerisation as well, which is thought to confer leukaemogenic potential (So *et al.* 2003). The partner genes may have a secondary role since even non-pathogenic genes may also promote MLL dimerisation as demonstrated by mice with an *MLL-LACZ* knock-in which showed up-regulation of *HOX* genes and developed myeloid and lymphoid leukaemias (Dobson *et al.* 2000). A gain of function through dimerisation could result in aberrant recruitment of coactivators or prevention of corepressor binding to cause upregulation of target gene expression (Martin *et al.* 2003; So *et al.* 2003).

The suggestion that *HOX* gene expression may not be necessary for *MLL* leukaemogenesis has been proposed. The class 2 fusion, *MLL*-growth arrest-specific 7 (*GAS7*), which activates MLL via a dimerisation mechanism was able to transform Hoxa9 and Hoxa7 deficient mice, though the resulting phenotype and penetrance was affected (So *et al.* 2004). Likewise, *HOX* genes were reported to influence the phenotype but not the incidence of *MLL-AF9* fusion gene leukaemia in a mouse model (Kumar *et al.* 2004). Additionally, a cell line derived from an *MLL-AF4* positive patient has functional *MLL* but lacks *HOX* gene

expression suggesting the transforming potential of the *MLL* gene can be independent of *HOX* (Bertrand *et al.* 2003). A likely explanation for the different results in relation to *HOX* expression and leukaemogenesis, is that the need for *HOX* is dependent upon the fusion gene and cell type being studied. An example of an aberration within the *MLL* gene that is not a translocation is the *MLL* partial tandem duplication (PTD) which is discussed in Section 1.2.4.a.

1.2.3.b   CBF mutations

The *CBF* is the most frequently affected site for chromosomal translocations in human leukaemia. Common *CBF* translocations: t(8;21)(q22;q22), inv16(p13;q22), and t(16;16)) are associated with a younger age group and favourable prognosis. The core binding factor mutations, t(8;21)(q22;q22), inv16(p13;q22), and t(16;16) prevent the dimerisation of the α and β protein components needed to initiate transcription of genes necessary for normal haematopoiesis. The t(8;21)(q22;q22) translocation involves the fusion of RUNX1 gene (also known as core binding factor α 2 (*CBFα2*)) with the *RUNX1T1* gene. It occurs in 10-15% of AML patients and is generally associated with FAB type M2 (Nucifora & Rowley 1995). The RUNX1 protein binds to the core enhancer sequences (TGGTGT) of genes responsible for haematopoiesis. Whereas the *CBFβ* subunit has no inherent DNA binding capacity, it increases the DNA binding affinity of the resulting heterodimeric unit (Cleary 1999; Kagoshima *et al.* 1993). The RUNX1 transcription factor up-regulates genes necessary for normal haematopoiesis including GM-CSF and M-CSFR which are involved in progenitor differentiation and proliferation (Rhoades *et al.* 1996; Takahashi *et al.* 1995). The translocation partner RUNX1T1 is a transcriptional repressor protein. Although it cannot physically bind DNA, it forms stable complexes with the repressor complex nuclear co-repressor (N-CoR) and the silencing mediator for retinoid or thyroid-hormone receptors

(SMRT) via the zinc finger domain and recruitment of histone deacytylases (HDAC). Retention of the ability of RUNX1T1 to bind stably with the repressor complex after translocation is thought to convert RUNX1 from an activator to a repressor, leading to a block in myeloid differentiation (Gelmetti *et al.* 1998).

The inversion of chromosome 16 (inv16(p13;q22)) also disrupts the *CBF* transcription complex by joining the *CBFβ* gene with the smooth muscle myosin heavy chain (*MYH11*) gene. The mutation is associated with a favourable prognosis and the FAB subtype M4Eo. The absence of *Cbfb* caused mice to die at midgestation, however the rescue of foetal liver haematopoiesis using a fusion protein restored early development although lymphoid and myeloid development failed, with the mice dying at birth (Miller *et al.* 2002). The data suggests that *Cbfb* is needed for the normal development of the myeloid and lymphoid lineages.

1.2.3.c PML-RAR

The t(15;17) *PML-RAR* translocation results in the fusion of the promyelocytic leukaemia protein (PML) to the retinoic acid receptor α (RARA) gene and is associated with a favourable prognosis and the FAB classification M3. A hypogranular morphological variant is sometimes observed, presenting with an elevated leukocyte count whereas the typical hypergranular cases are often leukopenic, but both exhibit a block in granulocytic differentiation (Puccetti & Ruthardt 2004). The retinoic acid receptor α (RARα) forms a heterodimer with the retinoid X receptor (*RXR*) to convey retinoid signalling within the cell. The heterodimer binds to specific retinoic acid response elements (RARE) on the target genes and initiates transcription via the recruitment of co-activators and histone acetyltransferases (HATs). When the ligand is unavailable, the co-repressor complexes NCoR/SMRT and

HDAC are recruited by the heterodimers at the RAREs, resulting in transcriptional repression and a block in granulocytic differentiation (Vitoux, Nasr, & de 2007). When translocated with the tumour repressor gene PML, the retinoic acid binding domain on the RARα protein is retained and pharmacological concentrations of retinoic acid can reinstate transcriptional activation and differentiation (Grignani *et al.* 1998). The ability to repair the haematopoietic pathway through application of retinoic acid is an example of a molecular-based targeted leukaemic therapy and is a goal for other leukaemia treatments (Section 1.2.6).

1.2.4 Molecular Genetic Mutations in Cytogenetically Normal AML

Cytogenetic analysis is a useful tool in the prognostic stratification of AML patients but it does have its limitations, specifically with regard to the large proportion of patients (~45%) that are cytogenetically normal (CN) and placed in the intermediate classification (Figure 1.6) (Byrd *et al.* 2002; Grimwade 2001). AML is typified by its diverse subgroups and broad range of patient prognosis. Nowhere is this more evident than in the CN cohort, with reported overall survival rates ranging from 24-42% suggesting underlying molecular heterogeneity (Farag *et al.* 2005; Mrozek, Heinonen, & Bloomfield 2001).
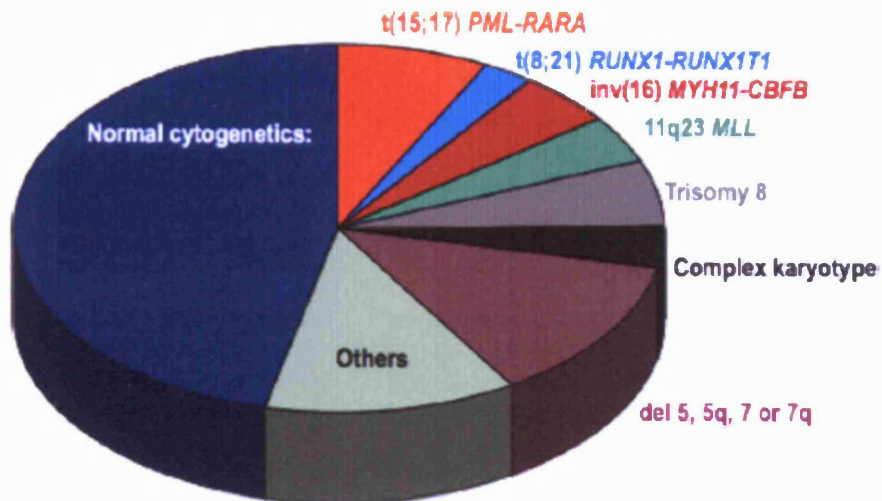
Figure 1.6 Cytogenetic Classification Frequencies
Relative frequencies of recurrent cytogenetic entities in AML. (Gulley, Shea, & Fedoriw 2009).

Recent findings have identified molecular genetic markers with prognostic significance in the CN group; these include *NPM1*, *FLT3*, *CEBPA* and *MLL* PTD. Such submicroscopic mutations are beginning to provide prognostic distinctions within the CN-AML patient group. As these mutations cannot be detected by normal cytogenetics they are typically detected by the sensitive method of polymerase chain reaction (PCR).

1.2.4.a   MLL PTD

The *MLL* PTD is an internal mutation of the *MLL* gene that often occurs in CN-AML. While previously discussed translocations in *MLL* (Section 1.2.3.a) implicated in the leukaemogenesis of AML all translocate with partner genes, the *MLL* PTD is the tandem

duplication of an internal sequence. The *MLL* PTD occurs in the bcr, with either exons 9 to 3 (e9/e3). 10 to 3 (e10/e3) or 3 to 11 (e11/e3) creating an in-frame fusion protein. The *MLL* PTD was discovered by association with trisomy 11 and normal cytogenetics (Caligiuri *et al.* 1996). A proposed mechanism for the PTD is the homologous recombination of Alu elements. These are a family of ~300 bp genes found throughout the genome which are flanked by Alu restriction sites. The reported reconstitution of an Alu site in the region where the *MLL* PTD inserts suggests that Alu elements are potential mediators for *MLL* PTDs (Strout *et al.* 1998) (Schichman *et al.* 1994b). By inserting 3` of the AT hooks and the methyl transferase and 5` to the zinc finger motifs, the distances between the DNA interacting regions are disrupted. This alteration in the pattern of the regulatory elements of MLL could be responsible for changing how the protein regulates haemopoietic genes (Figure 1.5) (Downing & Look 1996). The duplication of functional domains in *MLL* PTD may enable it to function in a manner similar to the class 2 *MLL* translocations, which dimerises to duplicate these regions (Section 1.2.3.a) (Ayton & Cleary 2001). The MLL PTD construct has transactivation activity comparable to that of a dimerised MLL translocation construct (Martin *et al.* 2003). *MLL* PTDs of various composition have been shown to exist in the PB and BM of healthy donors when a sensitive nested PCR technique is used (Schnittger *et al.* 1998). This suggests that careful optimisation of the PCR method would be required before the test could be made a routine diagnostic assay. This finding also suggests that the *MLL* PTD alone may not be enough to cause the onset of leukaemia.

Several links can be made between *MLL* PTD and *FLT3* internal tandem duplications (ITDs). Both *MLL* PTD and *FLT3* ITD have been reported to be associated with cytogenetically normal AMLs and appear to be susceptible to breakage by topoisomerase II inhibitors (Libura *et al.* 2003). There is also the evidence of the two duplications occurring simultaneously in AMLs (Steudel *et al.* 2003).

1.2.4.b FLT3

The *FLT3* gene is located on chromosome 13 and is a member of the class III tyrosine kinase receptor family. The gene encodes a membrane bound protein that is involved in haematopoiesis through the regulation of proliferation, differentiation and apoptosis of progenitor cells (Stirewalt & Radich 2003). *FLT3* is mutated in approximately one third of all AML patients, making it one of the most common mutations detected. There are two types of mutations that occur in the *FLT3* gene, the internal tandem duplication (ITD) and the tyrosine kinase domain (TKD) mutation. The more frequent ITD creates a variable length insertion in the juxtamembrane (JMD) region of the receptor and occurs in approximately 30% of CN-AML patients (Frohling *et al.* 2002; Kottaridis *et al.* 2001; Schnittger *et al.* 2002; Thiede *et al.* 2002). The TKD mutation is a missense mutation in the activation loop of the tyrosine kinase domain and occurs in 10% of CN-AML patients (Beran *et al.* 2004; Frohling *et al.* 2002; Thiede *et al.* 2002). Both mutations promote constitutive phosphorylation of the receptor in the absence of ligand and activate downstream signalling pathways including JAK2, STAT5 and mitogen activated protein kinase (MAPK) (Grundler *et al.* 2005; Lee *et al.* 2007). Furthermore, both *FLT3*-ITD and *FLT3*-TKD show a significant association with high bone marrow blast percentages and high white blood cell count at diagnosis (Frohling *et al.* 2002; Mead *et al.* 2007; Whitman *et al.* 2001). The *FLT3*-TKD generates conflicting clinical outcomes while studies have consistently shown that the *FLT3*-ITD has a negative impact on clinical outcome (Section 1.2.6.b) (Bacher *et al.* 2007b; Mead *et al.* 2007; Whitman *et al.* 2008b; Yanada *et al.* 2005). Differences between these two types of *FLT3* mutations are seen in experimental models. The TKD mutation promotes an oligoclonal lymphoid disorder in murine transplantation studies with a long latency period, whereas the ITD generates a myeloproliferative disease. Neither is capable of inducing AML (Grundler *et al.* 2005; Lee *et*

al. 2007). Together with the observation that *FLT3* mutations are often lost upon disease relapse, this indicates that the *FLT3* mutations are a secondary genetic event incapable of initiating disease (Gilliland & Griffin 2002; Kottaridis *et al.* 2002).

1.2.4.c NPM1

The *NPM1* gene has both tumour suppressor and oncogenic properties, dependent upon the context, and occurs in approximately 50% of CN-AML patients (Boissel *et al.* 2005; Falini *et al.* 2005; Schnittger *et al.* 2005a; Thiede *et al.* 2006). Over 40 mutations have been described in *NPM1*, the most common being a four basepair insertion that occurs in >75% of AML cases. All of the mutations result in aberrant localisation of the protein to the cytoplasm where its normal functions are disrupted; these include regulation of ribosomal protein assembly and the tumour suppressor pathways ADP-ribosylation factor 4 (*ARF*) and tumor protein p53 (p53) (Falini *et al.* 2005). The *FLT3*-ITD is detected in around 40% of all patients with *NPM1* mutations. The clinical outcomes of patients with the *NPM1* mutation but lacking the *FLT3*-ITD were significantly better than those with *FLT3* ITD alone (Section 1.2.6.b) (Dohner *et al.* 2005; Thiede *et al.* 2006). The suggestion that *NPM1* may be an early event in leukaemogenesis was supported by evidence that the *NPM1* mutations preceded mutations of *FLT3* and is a stable marker through relapse (Chou *et al.* 2006; Gorello *et al.* 2006; Thiede *et al.* 2006).

1.2.4.d CEBPA

CEBPA is a transcription factor belonging to the CCAAT/enhancer binding protein family and is involved in terminal differentiation and cell proliferation. The gene is instrumental in the early stages of myeloid differentiation and is expressed in myelomonocytic cells. The frequency of *CEBPA* gene mutations in *de novo* AML is ~9%,

with most of the mutations occurring in M1, M2 and M4 (Nerlov 2004). Support for the association with myeloid development is demonstrated by *CEBPA* expression beginning with the commitment of myeloid precursors and increasing during granulocyte differentiation. Additionally, mice without *CEBPA* expression lack mature granulocyte cells while other lineages are unaffected (Zhang *et al.* 1997). During haematopoietic development, *CEBPA* down regulates c-MYC to promote differentiation and up regulates genes involved in granulocytic differentiation (Johansen *et al.* 2001; Radomska *et al.* 1998). Despite evidence that *CEBPA* mutations cause dysregulation of granulopoiesis and myeloid progenitors, deletion of the gene fails to cause leukaemia (Heath *et al.* 2004; Zhang *et al.* 2004)(Kirstetter *et al.* 2008; Porse *et al.* 2001). There are two groups of mutations in *CEBPA*: N-terminal mutations which allow expression of the 30kDa protein while blocking translation of the larger 42kDa protein, and C-terminal mutations that disrupt the DNA binding domain (Nerlov 2004). Bereshchenko *et al.* investigated the influence of these two mutation groups found concomitantly in the majority of all AML patients with a *CEBPA* mutation (Bereshchenko *et al.* 2009). Homozygous mutations of either terminal are uncommon, whereas the combination of both mutations is detected in >90% of biallelic *CEBPA* mutations. This selection pattern indicates that each mutation contributes a particular function which combines to promote leukaemogenesis. The study determined that the C-terminal mutation promotes proliferation of the HSC population while the N-terminal, mutation promotes HSC differentiation to the stage of myeloid progenitor cells. The pro-leukaemic function of the C-terminal mutation takes place in a cell type upstream of the final leukaemic cell population, which has implications for the targeting of cells for therapeutic intervention (Section 1.2.6.c).

1.2.5 Pathogenesis of AML

A model of co-operative mutations has been proposed where two complementary categories of mutations are needed to generate AML. The two categories are mutations that enhance the proliferative and survival ability (Class I mutations) and mutations that impair differentiation and/or create the ability for self-renewal (ClassII) (Gilliland & Griffin 2002; Renneville *et al.* 2008). Although this is a reductionist model of co-operative mutations and confounding examples can be found (e.g. some mutations have been observed occurring simultaneously with mutations of the same class (Ishikawa *et al.* 2009)), it is a useful tool for understanding the onset of AML. Some class 2 AML fusion proteins including *RUNX1-RUNX1T1* and *PML-RAR* generate a "mutator" phenotype by the induction of processes that promote self-renewal while concomitantly down regulating genes involved in DNA repair (Alcalay *et al.* 2003). The idea of class 2 mutations preceding and promoting the class I mutations is supported by other mutator phenotypes associated with leukaemia including Down's syndrome and familial CEBPA mutations. Furthermore, the stability of class II mutations seen at relapse (*NPM1,CEBPa,* and AML translocations) as contrasted with the variability of class 1 mutations (*FLT3, c-KIT* and rat sarcoma (*RAS*)) supports this theory (Boissel *et al.* 2005; Kottaridis *et al.* 2002; Nakano *et al.* 1999; Shih *et al.* 2006b). The acquisition of a new class I mutation following remission is a possible explanation for the variability seen. Successful initial chemotherapy targeting cycling leukaemic cells with the class I mutation may return the disease to a previous stage of leukaemogenesis; from this stage, acquisition of a new class I mutation could cause relapse of the disease (Section 1.2.6.b).

1.2.6   Treatment of AML

1.2.6.a   Current treatments

The first goal of AML treatment is to induce complete remission (induction), defined as <5% blasts remaining in the bone marrow, followed by further treatment to extend the remission for as long as possible (consolidation). Patients who have maintained complete remission for >3 years have less than a 10% risk of relapse (Estey *et al.* 1997). The method of achieving complete remission has changed little in the past 30 years with the standard intensive chemotherapy remaining relatively unchanged. Currently, the decision of whether a patient receives standard treatment, allogeneic transplant, investigational treatment or palliative care is based on prognostic factors such as age and cytogenetic risk classification, as well as performance status (Estey & Dohner 2006) (Section 1.2.6.b). Recent advances in understanding the molecular pathogenesis of AML and predicting prognosis have led to the investigation of prognosis stratified treatment and therapies targeted to particular pathways. Such tailored treatments may offer improved clinical outcome, particularly to the elderly and poor-risk patients. (Section 1.2.6.c).

The standard treatment for AML patients consists of induction chemotherapy followed by a post remission therapy of further chemotherapy or a bone marrow transplant. The so-called "3+7" treatment is seven days of continuous intravenous infusion of cytarabine and three days of intravenous duanorubicin. Various combinations of alternative anthracyclines and doses, as well as other cytotoxic agents during induction have not improved patient outcome overall, although certain subgroups do benefit (Estey & Dohner 2006). Any benefits owing to higher doses in induction have to be balanced against the increase in toxicity and mortality. Generally, prognostic considerations dictate that favourable and intermediate risk patients proceed with consolidation chemotherapy after remission, usually consisting of multiple courses of high dose cytarabine (Section 1.2.6.b) (Jabbour, Estey, & Kantarjian

2006; Rowe 2008). The other alternative for patients with intermediate risk would be allogeneic bone marrow transplant. However, despite improved relapse-free survival there is generally no improvement in overall survival, due to treatment mortality together with a reduction in patient's quality of life (Burnett *et al.* 1998; Cassileth *et al.* 1998; Stone, O'Donnell, & Sekeres 2004)(Gale *et al.* 2005). Additionally, poor-risk *FLT3* ITD patients receiving a bone marrow transplant from a related donor show no improvement in outcome than over those without (Gale *et al.* 2005). This indicates that a less favourable prognosis does not necessarily benefit from more aggressive treatment. However, other studies indicate a positive clinical outcome for poor risk patients receiving an allogeneic transplant (Schlenk *et al.* 2008, Cornelissen *et al.* 2007).

1.2.6.b Factors Affecting Clinical Outcome

Prognostic factors including age, performance status and genetic characteristics are critical to determining the appropriate treatment strategy for the patient (e.g. standard treatment, molecularly targeted therapies or supportive care) and identifying those who might benefit from allogeneic transplant (Section 1.2.6.a and 1.2.6.c). One of the current goals in AML prognosis is to understand the relationships between the genetic mutations, including the more recently discovered molecular genetic aberrations, and the established indicators (e.g. age and white blood cell count) so that a stratified therapy can be administered with optimal benefits for all patients. In addition to the pre-treatment factors, the patient response to treatment as assessed by molecular and immunological minimal residual disease (MRD) methods contributes to the overall patient prognosis (Section 1.3) (Grimwade & Hills 2009). Prognostic scoring systems have been proposed including the HOVON/SAKK prognostic score which combines pre-treatment (including molecular genetic markers) with microarray

expression data to produce four risk categories with different outcomes for determining which patients would benefit from allogeneic transplant (Cornelissen *et al.* 2007).

The decision to consider alternatives to standard therapy for adverse risk karyotypes is due to the poor outcome; less than 50% respond to induction therapy and fewer than 10% of those who achieve remission will remain relapse free with either high dose cytarabine or autologous transplant (Bloomfield *et al.* 1998; Stone, O'Donnell, & Sekeres 2004; Suciu *et al.* 2003). This is in comparison to AML patients overall receiving standard therapy whose remission rates range from 50-85%, with 20-30% achieving the goal of long term disease free survival (Shipley & Butera 2009). Patients with secondary AML or who are elderly, are also associated with poor outcome. Refractory disease can occur in all three of these patient groups due to chemoresistance mediated by intracellular mechanisms (*e.g.* up-regulation of the multidrug resistance protein-1) that promote the removal of chemotherapeutic agents from the cell (Mahadevan & List 2004). Relapse following remission suggests that a small population of cells capable of expansion remain or arise via mutation despite the therapy administered. The high sensitivity of molecular techniques such as PCR and immunophenotyping enable detection of leukaemic cell populations well below the levels associated with the clinical definition of remission. The monitoring of MRD and its value to patient prognosis is discussed further below (Section 1.3). The rate of remission and disease free survival have a negative correlation with increasing age, with the rate of complete remission in the elderly ranging from 41-62% with dismal long term survival rates (Goldstone *et al.* 2001; Rowe *et al.* 2004). This is due to co-morbidity factors and an increased proportion of adverse cytogenetics. Analysis of cytogenetics is combined with assessment of performance status to determine if elderly patients are candidates for either intensive chemotherapy, investigational trial or would be better served by non-intensive or supportive

care (Section 1.2.6a). Despite conflicting reports of clinical outcome, allogeneic bone marrow transplant is the usual treatment for adverse risk patients (Shipley & Butera 2009).

While the root causes of many cases of AML remain unresolved, there are a number of acquired genetic abnormalities that are closely associated with the disease. Balanced translocations, complex karyotypes and sub-microscopic mutations are all examples of acquired mutations and most have a distinct prognosis.

The first impact of molecular biology on classification and prognosis arose from the association between diagnostic cytogenetics and outcome (Section 1.2.2) (Estey & Dohner 2006; Grimwade *et al.* 1998). Genetic abnormalities such as chromosomal deletions, additions, translocations and complex karyotypes have become the primary means for determining prognostic outcome of AML. Three clear groups, favourable, intermediate and adverse, were observed with markedly different outcomes. Compared to those patients with normal cytogenetics, the favourable group (including t(8;21) RUNX1-*RUNX1T1*, inversion 16 CBFB-*MYH11* and t(15;17) PML RARA), did significantly better in overall survival and the adverse group (including the monosomies 5 and 7, and complex karyotypes (five or more abnormalities) did significantly worse (Figure 1.7). Patients with *MLL* 11q23 translocations or other cytogenetic abnormalities that were not significantly different in clinical outcome than cytogenetically normal including the trisomies +11 and +21 were placed in the intermediate grouping (Grimwade *et al.* 1998). The largest subgroup in the breakdown of diagnostic cytogenetics is the cytogenetically normal group, with 40-50% of all AML patients in this class. These CN patients are placed in the intermediate prognostic group due to the lack of informative markers.

MRC/NCRI AML Trials: Overall Survival Ages 16–59
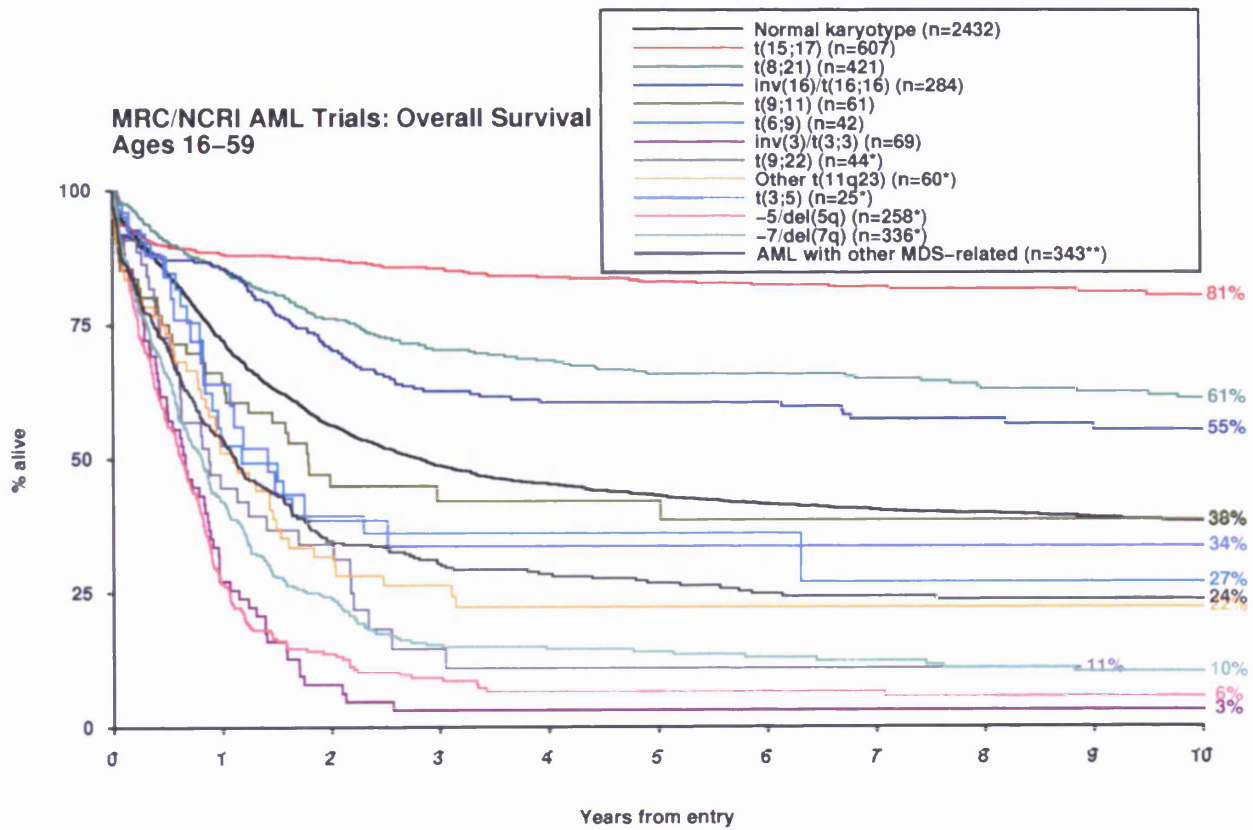
Figure 1.7 Overall Survival of AML Patients Grouped by Karyotype

Recognising the patients that may fall into the adverse risk group is particularly difficult in the cytogenetically normal group because the overall prognosis depends on the standard indicators as well as any molecular mutations acquired. New methods to assess CN-AML patient's risk category prior to initial treatment, as well as guidelines for mutational screening have been proposed which incorporate the prognostic impact of molecular genetic mutations (Mrozek *et al.* 2007; Renneville *et al.* 2008; Santamaria *et al.* 2009). Additionally, gene expression profiling (GEP) is capable of identifing subsets of CN-AML with a poor clinical outcome (Section 1.4.2) (Bullinger *et al.* 2004). There is general consensus that the *FLT3*-ITD confers lower disease free survival (DFS), overall survival (OS), and an increased risk of relapse (RR) (Figure 1.8) (Kottaridis *et al.* 2001; Sheikhha *et al.* 2003; Yanada *et al.* 2005). Studies specific to CN-AML found the *FLT3* ITD to be an independent adverse prognostic marker that conferred a lower OS and complete remission duration (CRD) (Baldus *et al.* 2006; Frohling *et al.* 2002; Whitman *et al.* 2001). On the other hand, the *NPM1* mutation confers a favourable prognosis, but only in the absence of a *FLT3* ITD. This improved outcome was demonstrated in better CR rates, relapse free survival (RFS), DFS and OS (Dohner *et al.* 2005; Gale *et al.* 2008; Thiede *et al.* 2006). In contrast, the patients with both an *NPM1* mutation and *FLT3*-ITD showed no improvement on intermediate poor prognosis conferred by the absence of both mutations (Gale *et al.* 2008). A prognosis determined without assessment for multiple mutations could be highly inaccurate and may prevent the patient from getting the best possible care. In fact, a recent study reported that *FLT3* TKD positive patients exhibit all possible prognoses dependent upon the identity of concomitant mutations. *FLT3* TKD as the sole abnormality has no effect, *FLT3* TKD with *PML-RAR* confers an unfavourable outcome and *FLT3* TKD with either *NPM1* or *CEBPA* confers a favourable impact greater than that seen with either mutation on its own (Bacher *et al.* 2007b). The identification of more molecular genetic mutations will surely discover more

complexities in the interactions between these and pre-existing mutations and their affect on

prognosis.
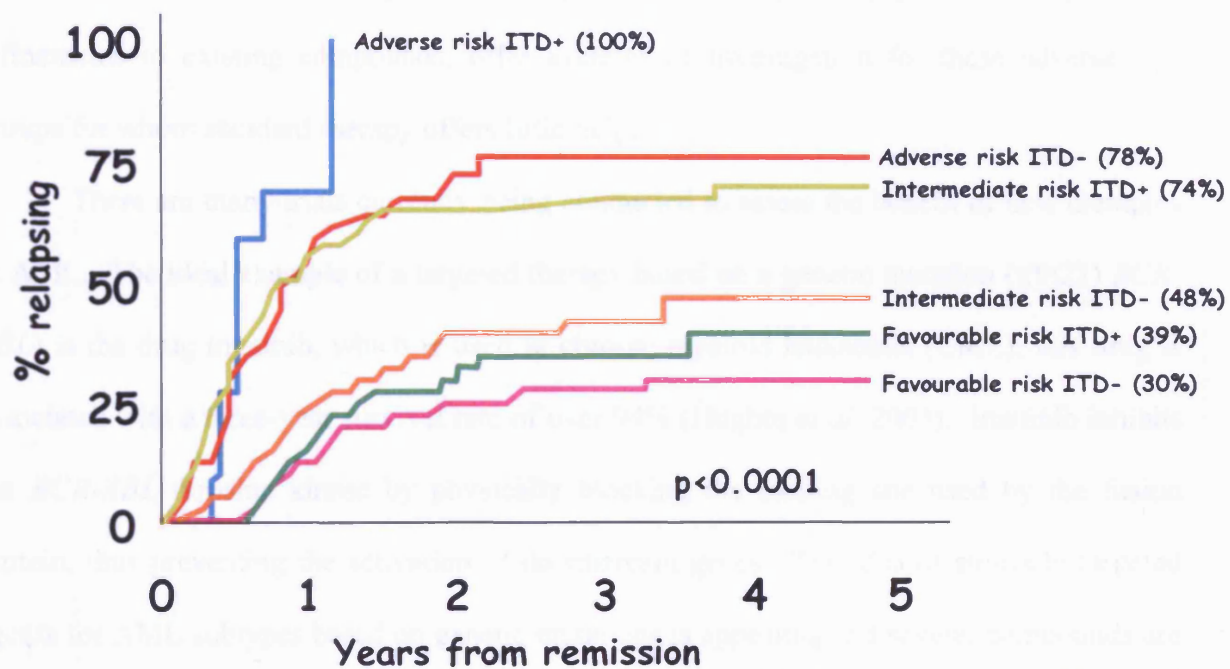
Figure 1.8 Impact of *FLT3* ITD on Risk of Relapse in Cytogenetic Risk Groups
The *FLT3* ITD is associated with a higher risk of relapse in every cytogenetic risk group. The
influence of the mutation is greatest in the intermediate and adverse risk groups. ITD:
internal tandem duplication.

1.2.6.c   New Treatments

Older patients and younger patients with adverse prognosis and no donor often opt for investigational therapy as the best option for improved clinical outcome (Estey & Dohner 2006). New treatment strategies including targeted therapies, epigenetic therapies and refinements to existing compounds, offer avenues of investigation for these adverse risk groups for whom standard therapy offers little help.

There are many trials currently being conducted to assess the benefit of new therapies in AML. The ideal example of a targeted therapy based on a genetic mutation (t(9;22) *BCR-ABL*) is the drug imatinib, which is used in chronic myeloid leukaemia (CML); this drug is associated with a three-year survival rate of over 94% (Hughes *et al.* 2003). Imatinib inhibits the *BCR-ABL* tyrosine kinase by physically blocking the binding site used by the fusion protein, thus preventing the activation of downstream genes. The idea of similarly targeted agents for AML subtypes based on genetic mutations is appealing and several compounds are under investigation including inhibitors of *FLT3* mutations such as the small molecule inhibitors PKC-412 and CEP-701. However, due to the genetic diversity of AML it is unlikely there will be a "one-size-fits-all" drug for AML. In initial trials as a single agent, these inhibitors appear to have a transient clinical response, however pre-clinical evidence suggests synergy between *FLT3* inhibitors and standard chemotherapy agents and trials in combination with chemotherapy are underway (Knapper *et al.* 2006; Knapper 2007; Shipley & Butera 2009). On a more global genetic scale, transcription modulators (DNA demethylating agents and histone deacetylase inhibitors) that affect epigenetic pathways have also been tested. DNA hypermethylation and histone deacetylation have been implicated in carcinogenesis via the suppression of tumour suppression genes; the reversal of these processes can reinstate tumour suppressor and pro-apoptotic gene function (Fandy, Carraway, & Gore 2007). Early studies have shown that AML patients have a positive response to

demethylating agents and histone deacetylating compounds, as well as to the combination of both and further trials are being conducted (Shipley & Butera 2009). As more genetic markers are discovered and the understanding of the existing markers expands there is increasing opportunity for discovering new genotype specific targets that could be exploited for therapeutic benefit. However, due to the small size of some genetic subgroups identified, large clinical trials are needed to assess their full contribution to prognostic stratification. Furthermore, to obtain a complete picture of the effectiveness of treatment modalities, all prognostic markers need to be taken into consideration so that independent prognostic variables can be identified.

1.3   Minimal Residual Disease

1.3.1   Background

The monitoring of minimal residual disease in AML is an important tool for determining the patient's initial response to treatment and for detection of impending patient relapse post-treatment. In acute promyelocytic leukaemia (APL), MRD evaluations are commonly used to enable the clinician to make a well-informed decision for further treatment. When a patient achieves the clinical definition of complete remission (<5% blasts in the bone marrow by morphological identification), the number of leukaemic cells still remaining can be as high as 10^{10} (Campana & Pui 1995). Between 50-70% of patients who achieve complete remission will relapse within three years, which implies a residual population of leukaemic cells was able to regenerate the disease (Liu Yin 2002b; Robak & Wierzbowska 2009). Patient response to induction therapy (gauged by the morphological detection of blasts) is an independent prognostic factor predicting for risk of relapse and overall survival, and carries recommendations for treatment (Cheson *et al.* 2003; Robak & Wierzbowska 2009). A

sensitive MRD assay would allow early identification of patients with either an inadequate response to treatment or at risk of relapse and would give clinicians an opportunity to alter the therapy at an earlier stage of disease progression, potentially improving the clinical outcome (Grimwade & Lo 2002; Liu Yin 2002b).

1.3.2 Methods of Detecting Minimal Residual Disease

The ability to detect leukaemic cells at a level below the level of morphological assessment has increased tremendously in the past 20 years. With the advent of new MRD methodologies, sensitivity has improved from approximately 1 cell in 20, detected by cytogenetic karyotyping/morphology to 1 cell in 500 by fluorescent *in situ* hybridisation (FISH) to 1 cell in 10^3-10^4 by flow cytometry and 1 cell in 10^3-10^6 by real time quantitative polymerase chain reaction (RQ-PCR) (Estey & Dohner 2006; Freeman, Jovanovic, & Grimwade 2008). The reported sensitivity is dependent on the methods used and the targets selected. Initially, conventional cytogenetic karyotyping was used to follow leukaemic disease, the leukaemic aberration present at the patient's diagnosis was absent during remission and then reappeared at relapse (Testa *et al.* 1979). FISH analysis is more sensitive than conventional cytogenetics and can detect some rearrangements not identified by karyotyping, however the higher sensitivities of flow cytometry and RQ-PCR make them the most effective MRD techniques.

One of the major advantages of flow cytometry for MRD analysis is its broad applicability to >80% of patients, through detection of aberrant expression of surface markers on leukaemic blasts (Al-Mawali, Gillis, & Lewis 2009; Kern & Schnittger 2003). While RQ-PCR measures MRD by the expression level of the marker gene, flow cytometry counts cell numbers using a leukaemia-associated aberrant immunophenotype (LAIP) that differentiates blast cells from normal cells. Although the sensitivity is lower than RQ-PCR, MRD analysis

using flow cytometry has predicted relapse free survival using cell numbers determined either after induction or consolidation (Kern *et al.* 2004; San Miguel *et al.* 2001). Disadvantages to this technique include lower sensitivity (compared to most RQ-PCRs) and possible inappropriate selection of markers due to poor specificity, sensitivity and stability in LAIPs (Al-Mawali *et al.* 2008).

The work presented in this study utilised RQ-PCR as a method of MRD detection so the associated methodology is discussed in greater detail. PCR is the most sensitive of the MRD methodologies, capable of detecting between 1 AML cell in 10^3 to 10^6 normal cells, however, MRD detection using PCR requires the identification of a leukaemic specific markers such as fusion genes (e.g. *PML-RAR*), mutations (e.g. *NPM1* mutation or *FLT3* ITD) or AML associated gene overexpression (e.g. Wilm's Tumour (*WT-1*). The main advantages of PCR as a method of MRD monitoring are that it is highly sensitive, quantitative and rapid, while disadvantages are risks of contamination and reduced sensitivity due to degraded RNA or inefficient reverse transcription. Additionally, a negative result must be considered not as the absence of disease but the absence of detectable disease by the methodology used (Liu Yin 2002a). This consideration is applicable to all types of MRD assays. Reasons for negative results include low or absent marker, low sensitivity in a particular sample or a change in the genotype of the patient (e.g. *FLT3* ITD). The PCR consists of multiple cycles of amplification of a target DNA specified by small complementary primer DNA sequences resulting in the exponential expansion of the target sequence. Reverse transcription PCR (RT-PCR) is a method where cDNA (following reverse transcription from RNA) is the template material for amplification. Qualitative RT-PCR is often used in diagnostic and follow up analysis of AML patients (Dolken 2001; Hokland & Pallisgaard 2000). A European consortium, "BIOMED-1 Concerted Action: Investigation of minimal residual disease in acute leukaemia", established standardised qualitative RT-PCR conditions for many

of the most common translocations in acute leukaemia (van Dongen *et al.* 1999). To increase sensitivity and reduce the risk of contamination, the Biomed-1 amplification used nested primers, with the second set internal to the first (van Dongen *et al.* 1999). However, differences in the kinetics of decreasing disease meant that haematological remission was found to be possible despite persistent positive results in core binding factor leukaemias making the results of qualitative tests uninformative if positive (Costello *et al.* 1997; Tobal & Yin 1996b). Therefore, despite being generally less sensitive, the method of RQ-PCR is often more informative than a qualitative RT-PCR due to the determination of an accurate quantitation of residual disease.

One of the early methods of quantitative PCR was competitive PCR. Competitive PCR is a technique where a known concentration of an internal control is serially diluted and amplified with the same primers as the unquantified target gene. The internal control is a different size to the target so the intensity of the amplification endpoints can be compared to determine the quantity of the target sample. Despite the success of this method at detecting relapse in RUNX1-RUNXT1 patients, it was labour intensive and was superseded with the advent of fluorescent based RQ-PCR technology (Tobal & Yin 1996a).

Using RQ-PCR technology, detection of fluorescence from internal probes or a fluorescent dye enables the calculation of the amount of target template present in an amplified sample. The fluorescence produced is proportional to the level of target being amplified allowing quantitation to be based on the accumulation of target instead of the endpoint amount, thus preventing misinterpretation due to "plateauing" of the PCR. The accumulation of fluorescence is measured on every successive cycle with the quantitation of the product based on the cycle at which the fluorescence is first detected. The number of cycles needed for detection of a sample is called the crossing point (Cp) or cycle threshold and this value is used to calculate the amount of target produced. Further discussion on RQ-

PCR can be found in Section 2.6.2 of materials and methods and in Chapter 4. Advantages to RQ-PCR are the speed of the testing, the reduction of contamination risk due to the elimination of post-PCR processing and the capability for automation and standardisation. The main disadvantage to RQ-PCR is the general lack of sensitivity in comparison to the qualitative testing (Liu Yin 2002a; Takenokuchi *et al.* 2004).

1.3.3 Assessment of Treatment Response and Prediction of Relapse

The two main goals of RQ-PCR MRD detection are assessment of the patient's initial treatment response as a prognostic guide to further therapy and prediction of post-treatment relapse. As examples of treatment response assessment, in RUNX1-*RUNX1T1* and CBFB-*MYH11*, the level of MRD following induction and consolidation was found to have a positive correlation with relapse risk with one study reporting RUNX1-*RUNX1T1* MRD to be an independent prognostic factor after adjustment for other pre-treatment factors (Krauter *et al.* 2003; Perea *et al.* 2006; Weisser *et al.* 2007). Several of the groups reported that achieving a 2-log reduction of disease related transcripts was prognostically significant in *CBF* leukaemias and leukaemias with *WT-1* overexpression (Krauter *et al.* 2003; Lapillonne *et al.* 2006; Stentoft *et al.* 2006). An example of where prediction of post-treatment relapse by RQ-PCR has been used as a prognostic guide is in the management of APL: an increasing level of *PML-RAR* transcript, detected by sequential RQ-PCR testing of the bone marrow successfully predicted relapse; preceding clinical relapse by 3 months (Lo-Coco & Ammatuna 2006; Reiter, Lengfelder, & Grimwade 2004). Furthermore, pre-emptive treatment studies to assess the dangers of associated coagulopathy in APL demonstrated a benefit to early treatment intervention on the basis of MRD analysis (Esteve *et al.* 2007; Lo *et al.* 1999a).

In addition to the different kinetics of disease reduction in different leukaemias, the kinetics of relapsing disease also have to be considered when analysing the effectiveness of MRD. The

length of time from first molecular detection of MRD to haematological relapse can differ

from four months for APL to 1-2 months for CBFB-*MYH11* (Ommen *et al.* 2008). Relapse

kinetics coupled with different sensitivities between peripheral blood and bone marrow make

the sample type and time points used for MRD crucial to success (Figure 1.9).
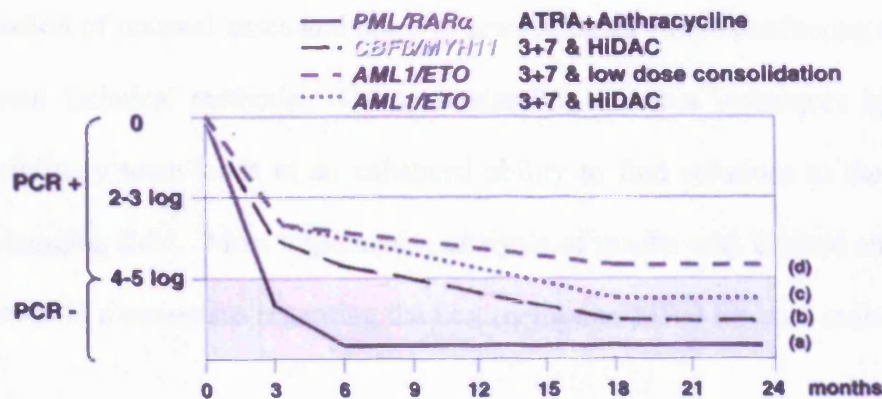
Figure 1.9 Minimal Residual Disease Kinetics
Different MRD kinetics in the bone marrow of long term survivors of three favourable
cytogenetic translocations are shown following standard chemotherapy. a) PML/RARA
becomes undetectable after induction/consolidation, whereas high dose treatment of
CBFB/MYH11and *AML1/ETO* (RUNX1/*RUNX1T1*) take around 12 months. The effect of
different treatment methods are shown with low dose consolidation of *AML1/ETO*
(RUNX1/*RUNX1T1*) remain low but detectable with the patient in complete remission.
ATRA: all trans retinoic acid; HiDAC: high dose cytarabine (Jaeger & Kainz 2003).

Further study is needed to confirm the effectiveness of MRD monitoring in identifying

patients whose suboptimal response to initial treatment indicates they may benefit from an

alternative treatment. Additionally, the effectiveness of early intervention in post-treatment

patients on the basis of increasing MRD needs to be studied in groups other than APL. These

studies should consider how MRD analysis could supplement the prognostic factors already in

use to improve clinical outcome. An RQ-PCR standardisation methodology has been

reported for MRD detection in leukaemia (Gabert *et al.* 2003), however, an international

adoption of common methodology is needed to address variations in sample type, monitoring

schedule, PCR technique, efficiency and reporting methods. Preferably, methodologies would be established for the different AML genotypes tested, similar to the approach being used for MRD in CML (Hughes *et al.* 2006).

Ideally, all of the techniques mentioned above are utilised and discussed in a multidisciplinary setting so that diagnostic results can be compared. This provides a forum for discussion of unusual cases and helps to resolve cases where conflicting results are found by different technical methods. The understanding of other techniques by members of a multidisciplinary team leads to an enhanced ability to find solutions to the challenges of a rapidly changing field. Most importantly, analysis of results with clinical and scientific staff present leads to a consensus regarding the best method of MRD for each individual patient.

1.3.4 Additional Minimal Residual Disease Targets

It appears likely that MRD analysis will be beneficial to improving patient outcome (Section 1.3.3), however the most commonly tested translocations (RUNX1-*RUNX1T1*, CBFB-*MYH11* and *PML-RAR*) occur in only 25% of all AML patients, therefore additional MRD targets are needed. Molecular genetic aberrations, mutations and gene over-expression, have been tested as targets for MRD. The use of *FLT3*-ITD appears promising due to the high incidence in all AMLs, but is hampered by instability at relapse and the need for patient-specific primers to increase the sensitivity (Schnittger *et al.* 2002; Schnittger *et al.* 2004). Initial studies of *NPM1* and *WT-1* as MRD markers are encouraging. *NPM1* is the most frequent mutation in CN-AML, is stable upon relapse, and requires only three primer sets to encompass >90% of all mutations (Figure 1.10) (Falini *et al.* 2005; Gorello *et al.* 2006). The expression level of *NPM1* has been shown to be predictive of relapse and a 3-log reduction was associated with a reduced risk of relapse (Chou *et al.* 2007; Gorello *et al.* 2006). *WT-1* is over expressed in the majority of AML cases and despite misgivings regarding detection of

1-60

over-expression in a background of normal expression, it has shown promise as an MRD marker (Garg *et al.* 2003b). Patients with the greatest reduction of *WT-1* expression after induction had a significantly lower risk of relapse after adjustment for WBC and cytogenetics; conversely continuing to express *WT-1* after consolidation had a significantly higher risk of relapse (Cilloni *et al.* 2009). In summary, the use of new molecular aberrations for MRD monitoring appears to be prognostically useful, although the assays are often less straightforward than translocation testing.
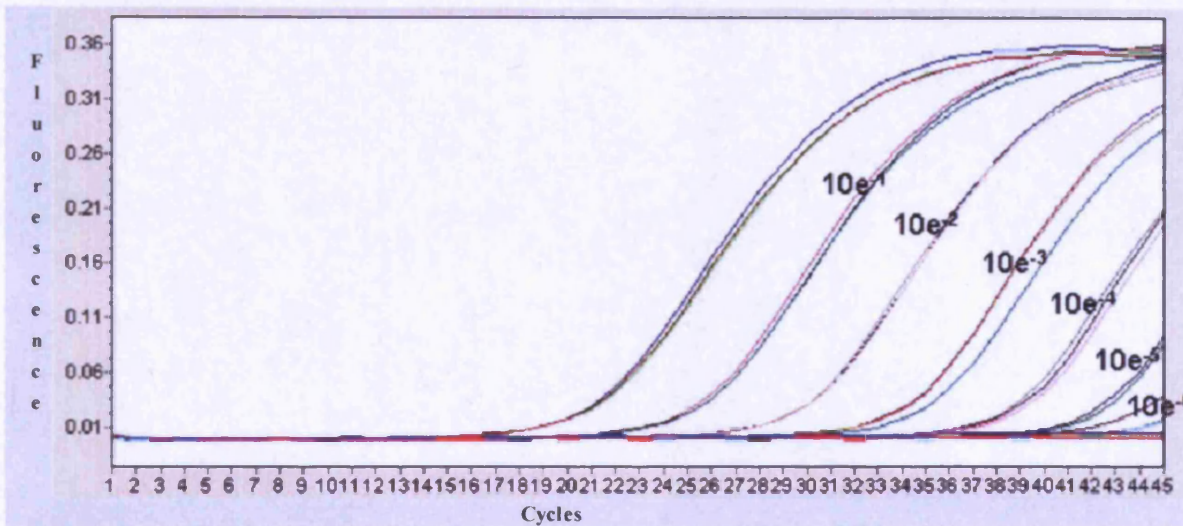
Figure 1.10 Sensitivity of *NPM1* by RQ-PCR
The sensitivity of *NPM1* is shown by a limited dilution series of *NPM1* positive sample into a *NPM1* negative sample. The mutation is nearly detectable to 1:1,000,000. (Kern *et al.* 2008).

1.4   Gene Expression Profiling

Gene expression profiling provides a new technique for the analysis of the heterogeneity of AML by measuring the expression level of thousands of gene transcripts simultaneously. Genetic and biological diversity is a hallmark of AML and is used to determine prognosis and predict the response to different therapies (Section 1.2.6.b). Given that a large proportion of AML are cytogenetically normal and multiple mutations appear to be needed to promote leukaemogenesis, the heterogeneity of AML is likely to increase with improved test resolution (Dash & Gilliland 2001). The ability of GEP to evaluate multiple gene expressions in a single analysis offers a potential solution to this dilemma. Researchers have used GEP investigations to gain insight into the complexity of AML with the aim of predicting AML subsets, predicting prognosis and identifying new leukaemia related pathways.

1.4.1   Prediction of AML Subgroups

GEP analysis has been proposed as a method of clinical diagnosis for AML due to its success as a predictor of AML subgroups. Studies such as the Microarray Innovations in Leukaemia (MILE), aim to standardise GEP for use in a routine diagnostic setting (Kohlmann *et al.* 2008). Class prediction is used to predict groups based on known sample information (e.g. cytogenetic abnormality) whereby a unique signature is identified for a particular cytogenetic aberration and subsequently used to predict that group from the expression data of unknown samples. This method was first used to study acute leukaemia in 1999, when Golub *et al.* analysed 38 samples from acute leukaemia patients using a set of 50 predictor genes to successfully differentiate AML and ALL samples (Golub *et al.* 1999). Unique gene

signatures attributed to the favourable cytogenetic translocations t(8;21) RUNX1-*RUNX1T1*, t(15;17) *PML-RAR* and inv(16) CBFB-*MYH11* have been used to predict the subgroups with up to 100% accuracy (Debernardi *et al.* 2003; Valk *et al.* 2004)(Ross *et al.* 2004). Results for the more diverse 11q23 rearrangements were conflicting, with reports of 90% accuracy and no prediction value at all (Haferlach *et al.* 2005; Verhaak *et al.* 2009). Furthermore, the investigation of other cytogenetic abnormalities such as complex karyotype, trisomy 8 and 3q have failed to demonstrate accurate prediction (Bullinger *et al.* 2004; Verhaak *et al.* 2009; Virtaneva *et al.* 2001). Prediction of the molecular genetic mutations has met with similarly mixed success. Prediction accuracy of *CEBPA* biallelic mutations was 100%, while prediction of *NPM1* was 66% and *FLT3* ITD was either low prediction accuracy or none at all (Verhaak *et al.* 2009). Generally, the mutations with more direct roles in transcription modulation such as the *CBF* translocations were more accurately predicted while non-transcription factor abnormalities (*e.g. FLT3* ITD) fared less well (Wouters, Lowenberg, & Delwel 2009a). In summary, the success of GEP subgroup prediction was limited to certain groups of diagnostic markers; the identification of additional signatures able to classify the non-transcription factor abnormalities would give GEP greater potential as a diagnostic assay.

1.4.2   Prediction of Prognostic Subgroups

GEP analysis is a potential solution to the difficulties of prognostic classification of patients due to the rising number of genetic markers being identified and the complexity of combinatorial effects. The strength of GEP is not primarily in the prediction value of singular definitive markers, but in the overall signature of expression, which has been investigated with the aim of predicting prognosis. In fact, while *FLT3* ITD was poorly predicted as an entity, the gene signature was a significant indicator of clinical outcome. The poor prediction may have been due in part to the identification of similar signatures that arose from other

mutations or epigenetic changes without involvement from *FLT3* ITD (Baldus & Bullinger 2008). Several studies have reported the identification of new prognostic subgroups using GEP analysis. A set of 133 predictor probe sets was determined to be an independent prognostic factor, successfully predicting the OS of AML patients including those with CN-AML. The CN-AML patients were divided into two groups with significantly different overall survival (Bullinger *et al.* 2004). A laboratory using a different GEP methodology independently validated these results. The presence of FTL3 ITD contributed to some of the prognostic effect (Radmacher *et al.* 2006). Furthermore, a gene signature composed of 86 probe sets was an independent predictor of OS in CN-AML patients, and maintained its prognostic value after multivariable analysis considered other prognostic variables including age, *FLT3* ITD and *NPM1* status (Metzeler *et al.* 2008). Recently, GEP was used to identify a gene expression ratio capable of predicting the chemotherapeutic response of the patient to tipifarnib, a fanesyl transferase inhibitor (Raponi *et al.* 2008). The ability to predict the response of the patient's leukaemic cells to chemotherapy agents would allow clinicians to make a better informed decision on treatment options. Whether the ability of GEP to predict prognosis will ultimately identify subgroups or possibly novel markers that are independent of the prognostic markers already established remains to be seen.

1.4.3   Pathway Identification

Another potential application of GEP is the discovery of new biological pathways related to leukaemogenesis. Introduction of leukaemic mutations into animal or cell line models followed by GEP analysis may identify downstream pathways or targets instrumental to leukaemogenesis. For example, ablation of Cebpa p42 translation in a knock-in mouse study demonstrated a common gene signature between Cebpa and *MLL-AF9*, indicating the utilisation of a common pathway specific to the transformation of committed progenitor cells

(Kirstetter *et al.* 2008). A similar methodology was used to identify the use of an embryonic transcriptional pathway in myeloid progenitors with *MLL* translocations (Section 1.1.2.b) (Somervaille *et al.* 2009). The method of defining an AML subgroup, identifying differentially expressed genes and selecting genes for further study is described as the supervised class comparison strategy (Wouters, Lowenberg, & Delwel 2009b). The interpretation of gene expression data in relation to pathway analysis is made possible by a number of bioinformatic analyses (*e.g.* KEGG, DAVID and CMAP). These utilise literature and gene signature databases to identify connections between genes, gene pathways and drug treatments, further detail on these analyses can be found in Chapter 5 (Section 5.3.7) (Lamb *et al.* 2006) (Kanehisa *et al.* 2010b) (Dennis, Jr. *et al.* 2003). For example, a subgroup of *CEBPA* negative patients was identified with a GEP similar to that of patients with the *CEBPA* mutation. The silencing of the *CEBPA* gene in the subgroup was often due to promoter hypermethylation (Wouters *et al.* 2007). Pathway analysis was used to reveal that the leukaemic cells in this group of patients expressed a combination of myeloid and T-lymphoid genes; these findings were subsequently confirmed by the expression of T-lymphoid genes in mice lacking expression of *CEBPA*.

The next step in genome wide studies is the integration of GEP with other whole genome investigations including microRNAs, epigenetic changes (eg. methylation status and histone acetylation) and ultimately whole genome sequencing (Magic, Supic, & Brankovic-Magic 2009; Mardis & Wilson 2009; Mrozek *et al.* 2009; Nervi, Fazi, & Grignani 2008).

1.5 Aims of Study

Despite improvements in treatment and care, many AML patients fare poorly, with only 20-30% of patients achieving long term disease free survival (Shipley & Butera 2009). The predominant reason for this is treatment failure due to refractory disease and relapse particularly in adverse risk patients. Identification of adverse risk patients who are unlikely to benefit from standard chemotherapy treatment is one method of improving clinical outcome. Investigation of different treatment modalities such as allogeneic transplant or targeted therapy (e.g. *FLT3* inhibitors) through clinical trials has the potential to improve these patients overall survival. The principal means of determining prognosis is through the patient's diagnostic cytogenetics, however, there is still variation in survival rates within these prognostic classifications, particularly within the CN-AML patients due to the absence of a cytogenetic marker. Additionally, minimal residual disease monitoring can contribute to the prediction of prognosis through initial treatment response and the detection of impending relapse post-treatment in many types of AML. The recent identification of molecular genetic mutations such as *FLT3*, *NPM1* and *CEBPA* has enabled a refinement of prognostic classification, especially in the CN-AML group where many of the mutations are detected. Another recurrent molecular genetic mutation identified is the partial tandem duplication of the *MLL* gene. Early studies suggest the *MLL* PTD has an association with CN-AML and poor prognosis. The *MLL* gene plays an instrumental role in normal haematopoiesis and its disruption through fusion genes has been shown to be leukaemogeneic, though the mechanism for this is unclear. The general aim of this thesis is to investigate the prognostic significance of the *MLL* PTD in AML and provide information for determining appropriate treatment for this subgroup of patients.

Specifically three aims were set:

1) Chapter 3. Evaluate the use of *MLL* PTD as an indicator of poor prognosis at diagnosis

2) Chapter 4. Investigate the usefulness of the *MLL* PTD as a marker for MRD

3) Chapter 5. Identify a gene signature unique to *MLL* PTD using gene expression profiling to discover potential targeted therapeutic options and to determine if the signature can be used for class prediction (and if the validated genes are candidate genes for MRD markers)

2 Materials and Methods

2.1 Aims

The purpose of this chapter is to explain in detail how the experiments employed in this thesis were performed and why certain techniques were used. For more uncommon techniques, additional explanation is given to explain the underlying processes within the experiment. The techniques described include: PCR, RQ-PCR, RNA isolation, sequencing and gene expression analysis. Any experimental modifications of the specific methods and all technical optimisations are cross referenced to a more detailed description in the relevant chapters.

2.2 Materials

2.2.1 Chemicals

Unless otherwise indicated general chemicals were purchased from Sigma-Aldrich, (Dorset, UK) or Fisher Scientific, (Loughborough, UK) and were of analytical grade. The water used was water for injections from Hameln Pharmaceuticals (Gloucester, UK).

2.2.2 Enzymes

All enzymes related to reverse transcription and PCR were purchased from Applied Biosystems, (Warrington, UK). RQ-PCR enzymes were obtained from Roche, (West Sussex, UK).

2.2.3   Nucleic acids

DNA size markers were purchased from Abgene, (Epsom, UK) or previously 1kb ladder from Life Technologies, (Warrington, UK). Deoxyribonucleoside triphosphates were purchased from Applied Biosystems, (Warrington, UK). All PCR and RQ-PCR primers were ordered through Thermo Electron, (Epsom, UK), where they were purified by high performance liquid chromatography and verified for concentration by spectrophotometry. Trizol, used in RNA isolation was purchased from Invitrogen Life Technologies, (Paisley, UK).

2.2.4   Photography

Black and white Polaroid film 667, ISO 3000, was purchased from Fahrenheit, (Milton Keynes, UK) and was used for all gel photography.

2.2.5   Capillary Electrophoresis

All labelled DNA markers, capillaries and associated reagents for the ABI Prism 310 Genetic Analyzer were purchased from Applied Biosystems, (Warrington, UK).

2.2.6   Miscellaneous

Thin walled 0.5ml PCR tubes were purchased from Abgene, (Epsom, UK). Tubes used for RNA isolation and storage were obtained from Bioquote Ltd, (York, UK). All tubes used were certified Rnase/DNase, DNA and pyrogen free. PCR reagents were all purchased from Applied Biosystems, (Warrington, UK). Aerosol resistant tips were bought from Anachem, (Luton, UK). 50ml polypropylene tubes used for cell lysis were purchased from

Sarstedt, (Leicester, UK). Further laboratory consumables were obtained from Bioquote Ltd (York, UK), Eppendorf UK limited (Histon, UK) Fisher Scientific UK Ltd (Loughborough, UK), Bioquote Ltd (York, UK), Eppendorf UK limited (Histon, UK) Fisher Scientific UK Ltd (Loughborough, UK).

2.2.7 Commercial Kits

Compositions of some of the solutions provided by commercial kits were not provided. ABI Big Dye Sequencing Kit 3.1 was purchased from Applied Biosystems, UK. Qiagen Dye Ex 2.0 Spin Kit, Qiagen Gel purification Kit and Qiagen PCR Purification Kit were bought from Qiagen, UK. LightCycler FastStart DNA Master Syber Green was purchased from Roche, (West Sussex, UK).

2.2.8 Service Providers

The Cardiff University Central Biotechnology Services (CBS) genomics section electrophoresed all sequencing products using the ABI 3130xl.

2.2.9 Cell Lines

RNA from the cell lines tested for use as positive controls in gene validation studies were kindly provided by Cardiff University Haematology Department and included HL-60, Kasumi, K562, KG-1, NB4, NB4-R2 and U937.

2.3   Patient Samples

2.3.1   Patient Material

A total of 386 patients and 20 healthy controls were tested for the presence of *MLL*

PTD in this study. The healthy control samples were obtained from donor samples collected

for bone marrow transplant, no treatment was given prior to collection. The patient samples

were derived from two groups; all healthy donor samples and 130 of the Medical Research

Council AML Trial-United Kingdom (MRC-UK) patients were obtained from Cardiff

University, University Hospital Wales and 256 patients were assayed by the Haematology

department's RT PCR AML screen assay.

All samples were tested originally as "query new AML" from the clinicians' initial

findings. As the full blood count results and observational clinical details converged on the

possibility of AML, a bone marrow was taken to determine the state of haematopoiesis. The

diagnosis of AML was determined using a combination of cell morphology, flow cytometry,

cytogenetic analysis and PCR, culminating in a multidisciplinary meeting where the diagnosis

was decided along with the optimal method of MRD detection. The FAB classification

(Bennett *et al.* 1976) and/or WHO categorisation (2002) was also determined (Vardiman,

Harris, & Brunning 2002). A complete list of all patients tested and characteristics can be

found in Appendix A and a full list of AML patients' clinical characteristics in Section 3.2.1.

The majority of the samples were from bone marrow aspirates taken at diagnosis, yet in some

cases where a bone marrow was unavailable peripheral blood was used. Bone marrow was

the preferred sample as it is the site of haematopoiesis containing stem cells and their

differentiated descendents. Occasionally, the origin of the sample was not able to be

determined, in these instances it was labelled "not specified". To optimise conditions for

isolating RNA, BM and PB samples were collected into ethylenediaminetetraacetic acid

(EDTA) tubes and stored at room temperature until lysing. EDTA is a common anti-coagulant used to prevent samples from clotting in the time between sample collection and when the red blood cells are lysed.

2.3.2 Patient Treatment

All data used for clinical outcome analysis was obtained from intensively treated trial patients from MRC AML trial 11, 3.5% (8/225); from trial 12, 7.1% (16/225); from trial 14, 18.2% (41/225) and from trial 15 71.1% (160/225). The patients were treated according to trial protocols with the particular therapy given dependent upon their randomisation to a specific therapeutic arm of the study. Details of the treatments can be found at the following websites or are given in brief in (Grimwade *et al.* 2010; Wheatley *et al.* 2009).

AML 11 http://www.cancer.gov/clinicaltrials/MRC-LEUK-AML11

AML 12 http://www.cancer.gov/clinicaltrials/MRC-LEUK-AML12

AML 14 http://www.download.bham.ac.uk/bctu/aml14/

AML 15 http://www.aml15.bham.ac.uk/

2.4 RNA Isolation

RNA was the basis of all the assays undertaken. The advantage of RNA versus DNA was that it gave evidence of expression, expression levels and avoided the problematic amplification of large products caused by intervening intronic areas. The disadvantage of RNA was the problem of degradation due to enzymes which degrade RNA called RNases.

The major difficulty with RNA work was preventing degradation via RNases. Great care was taken to reduce the possibility of RNase contamination at all stages of the testing

including the use of RNase and DNase free tubes, tips and water. All tips used contained filters to prevent samples being transferred from one tube to another through aerosols in the barrel of the pipette chamber. Gloves were always worn and every effort was taken to keep the work area as free from dust as possible.

The method of isolation used was dependent upon the denaturant guanidine isothiocyanate and phenol to both lyse the cells and denature RNases to prevent RNA degradation. Chloroform was added to create an organic phase containing the DNA and protein while the RNA remained in the upper aqueous phase. Isopropanol precipitated the RNA which was then washed with 75% ethanol and resuspended in molecular grade water. All of the experiments conducted used RNA as starting material. Total RNA (ribosomal, transfer, nuclear and messenger) was isolated using RNAzol Bee or Trizol, in both instances the manufacturer's protocols were followed. The change from RNAzol Bee to Trizol was made to unify with the protocol required for the Affymetrix assay. Bone marrow samples were used preferentially; PB was used if BM was not available.

Red blood cell lysis was undertaken in order to derive a pellet of white blood cells (WBCs) for RNA isolation not contaminated with anucleated RBCs. The buffer contained ammonium chloride, which lysed red cells with minimal effect on lymphocytes. To minimise the risk of exposure to aerosols the processing of samples was done in the fume hood. The red cell lysis buffer (RCLB) was 0.15M NH_4Cl, 0.1mM EDTA and 10mM $KHCO_3$ which was autoclaved and stored at 4°C.

2.4.1 Red Cell Lysis

The PB or BM sample was poured into a 50 ml polypropylene tube. To maximise the amount of sample collected, especially with small marrow samples, 2ml of 1x RCLB was

pipetted into the EDTA tube, inverted and then gently poured into the 50ml tube. For efficient lysing no more than 3-4ml of PB was used in a 50ml tube.

The tube was filled with 45ml of 1x RCLB and rotated until the solution changed from opaque to clear indicating lysis of the majority of RBCs, usually 10-15 minutes. PB often took longer than BM due to the greater number of RBCs present. Samples with unusually high WBC counts never became completely clear due to the high number of unlysed WBC present.

The samples were spun at 1559xg for 10 minutes to pellet the WBC. The supernatant was gently decanted so the pellet remained in the bottom of the tube. 40ml of 1x RCLB was added to the WBC pellet and vortexed until resuspended. Step 3 was repeated (centrifugation for 10 minutes at 1559xg, decant gently.). 40ml of PBS was added to the pellet and vortexed until resuspended. Once again, step 3 was repeated (centrifugation for 10 minutes at 1559xg, decant gently.) The small amount of remaining PBS was pipetted off being careful not to disturb the cell pellet.

1ml of Trizol stored at 4°C was added to the pellet and resuspended by a combination of repeated pipetting with a 1ml pipette and vortexing. If the WBC pellet was very small (not opaque) or not visible at all then 500µl was used. If the solution was not easily pipetted after adding 1ml, usually due to a large WBC pellet, an additional 1ml of Trizol was added to the sample. This was indicated when the cell pellet volume was more than 10% of the Trizol volume.

The sample was stored at –70°C for up to 1 week before RNA isolation. The sample can be stored for up to 6 weeks before degradation begins. Some methods such as RNA later allow for longer storage times.

2.4.2   RNA Isolation

Samples were removed from the −70°C and left to thaw at room temperature. Chloroform was added to an empty 1.5 ml eppendorf tube to the concentration of 0.1 volume of chloroform per ml of Trizol (usually ~130 µl).

The Trizol solution containing the WBCs was added to the chloroform, capped tightly, inverted for 45 seconds and left at room temperature for 5 minutes. Tubes were then centrifuged for 15 minutes at 16,000xg at 4°C. Using a 1ml pipette, the upper RNA containing aqueous layer was carefully pipetted off without disturbing the protein interface below and transferred to a second tube containing 0.6 volumes of isopropanol (usually ~600µl). The samples were inverted for 20 seconds and left to sit at room temperature for 10 minutes. Samples were then centrifuged for 15 minutes at 16,000xg at 4°C. The supernatant was decanted, 1ml of 75% ethanol was added to the pellet, and inverted several times. Tubes were again centrifuged for 15 minutes at 16,000xg, at 4°C. The supernatant was discarded and the tubes inverted to let the ethanol evaporate from the RNA pellets. Pellets were not allowed to over dry as the RNA can become very difficult to resuspend. Dependent upon size, the pellet was resuspended in 15µl-200µl of "molecular genetic" grade water.

2.4.3   RNA Quantitation and Storage

To ensure that the same amount of RNA was used as starting material in each experiment, the concentration and quality of RNA was determined using an ultra violet spectrophotometer to measure optical density (OD). A solution of RNA with an OD of 1.0 at 260nm in a 10mm path length has a concentration of 40µg/ml. The absorbance of the purified RNA was measured using either a Gene Quant pro or Nanodrop spectrophotometer. The Gene Quant spectrophotometer was purchased from Amersham Biosciences, UK, and the

Nanodrop from Labtech International, UK. The change between the spectrophotometers was made in September 2005 as equipment was upgraded. Checks were made to verify that the results derived from the new machine were comparable with those obtained previously.

RNA was measured on the Gene Quant by using a quartz micro cuvette. Molecular grade water was used as a reference and then $2\mu l$ of RNA was added into $70\mu l$ molecular genetic grade water. Measurements were made at 260nm and 280nm and the concentration was calculated using the formula:

Concentration $(\mu g/\mu l) = (OD_{260nm})$ x conversion factor x dilution factor x $(1ml/1000\mu l)$.

Where the conversion factor is $40\mu g/ml$.

The purity of the sample was determined by dividing the A260 by the A280 absorbance values. Pure RNA should give a ratio between 1.8 and 2.0. The purity was calculated by the formula:

RNA purity $= OD_{260}/ OD_{280}$

To measure OD with the Nanodrop, $2\mu l$ of the sample was loaded directly onto the instrument. The instrument produces a column of liquid between two "posts" and then reads the OD through the sample column. The instrument was set to zero using $2\mu l$ of water as a reference and then the RNAs were measured neat using $2\mu l$ of the total sample. The Nanodrop uses the following formula to calculate the concentration of RNA. To prevent degradation during storage all RNA samples were kept at -70°C.

Concentration $(ng/\mu l) = (OD_{260nm})$ x conversion factor x $(1000\ ng/\mu g)$ x $(1ml/1000\mu l)$.

2-77

Where the conversion factor is 40µg/ml.

2.5 Reverse Transcription

RNA was then reverse transcribed into complementary DNA (cDNA) using the reverse transcriptase enzyme isolated from Moloney murine leukaemia virus (MMuLV) using reagents for reverse transcription purchased from Applied Biosystems. The manufacturer's guidelines were followed with minor modifications. All assays, except for the t(8;21), used random hexamers to prime the reverse transcription. For the t(8;21) PCR, the outer sense primer was used as a specific primer to prime the reverse transcription reaction. This was necessary to achieve the sensitivity needed to detect MRD in follow-up patient samples. All assays contained a negative control containing water with no RNA as a check for contamination.

One µg of RNA was used as template for each reaction. Water was added to make the total volume up to 20µl after all of the reagents were added. The following reagents were used at a final concentration of: 2.5µM random hexamers, 1X PCR buffer pH 8.3 (10mM Tris-HCl, 50mM KCl), 1Unit/µl RNase inhibitor, 5mM $MgCl_2$, 2.5Units/µl MuLV reverse transcriptase and 1mM each deoxynucleoside triphosphates (dNTP). The tube was spun briefly and run on the ABI 9700 (Applied Biosystems, Warrington, UK) using the PCR programme: 25°C for 10 minutes, 42°C for 30 minutes, 95°C for 5 minutes and 4°C indefinitely. cDNA was either used immediately or stored at -20°C until needed for PCR.

For the t(8;21) reaction, an additional reverse transcription (RT) was done using the specific *ETO* primer from the first round. The outer primer from the nested PCR replaces the random hexamers in the RT reaction at a concentration of 123nM. The reaction was then

completed as stated above. This protocol modification was used to increase the level of sensitivity to that comparable with the other tests.

2.6 Polymerase Chain Reaction (PCR)

PCR is a method where millions of copies of a specific region of DNA can be produced from a small amount of starting material. PCR was conceived in 1986 by Kerry Mullis (Mullis *et al.* 1986). Since then PCR has revolutionised research by reducing the amount of time and starting material needed to produce large amounts of DNA.

PCR is a technique that utilises the ability of the DNA polymerase enzyme to recognise and extend double stranded DNA in a 5' to 3' direction. By creating two small stretches of DNA (primers) specific to the area of amplification, one complementary to the sense strand and the other to the antisense strand, a unique area of DNA can be amplified. The reaction is repeated through cycles of DNA separation from double to single stranded (denaturation), binding of complementary primer DNA (annealing) and elongation of the template at optimum temperature (extension) so that the amplicon becomes the most abundant DNA in the reaction. With near doubling of product at every cycle the amplicon is exponentially amplified with the reaction capable of creating up to 1 million copies of target sequence. The discovery of heat stable Taq DNA polymerase meant that the enzyme did not have to be added after every denaturation step and made the technique universally accessible. There are a number of areas in which a PCR can be optimised in order to improve sensitivity and productivity. The majority of PCRs need to be optimised to find the combination of factors that will make the reaction as robust and productive as possible. The annealing temperature of the primers needs to be tested empirically to find the temperature that reduces non-specific products and gives good yield. If the temperature is too low the primers may bind to sites that are slight mismatched causing non-specific bands, and if it is too high they

may not even bind to the matched site and thereby reduce the overall yield. The requirement for magnesium ions is another important component to the reaction as it affects the stringency of primer binding and is a catalyst for Taq activity. The length of time for the different stages of the PCR may need to be adjusted to allow for complete extension depending on the length of the product to be amplified.

Additional advances in the efficiency of PCR machines and the Taq enzyme have made PCR reactions faster, more specific and the reactions more robust. Specifically, the ability to prevent Taq enzyme from being active until the reaction has been heated to the denaturation temperature of 94°C, has produced a significant improvement in the specificity and yield of PCRs. These so called "hotstart" enzymes prevent the Taq enzyme from amplifying non-specific products at sub-optimum conditions during the PCR set up thereby inhibiting the production of incorrect duplicates at a very early stage. All of the PCRs described in this study used the hotstart enzyme Taq Gold.

The speed, sensitivity and efficiency of recognising target sequences make PCR well suited for doing mutational testing. Turnaround times are shortened using PCR, and for haematology testing the increased sensitivity meant it could be used to detect minimal residual disease (MRD) at a much lower level than previous methods allowed (cytogenetics, FISH, immunophenotyping) (Section 1.3). The disadvantage of PCR is the risk of contamination. The increased sensitivity is coupled with the danger of a false positive resulting from even the smallest amount of contamination or non-specific amplification during the reaction. For this reason negative controls are always used from the RT reaction all the way through to the PCR reaction itself and very strict procedures are observed to prevent contamination from post amplification samples.

An area dedicated for the preparation of PCR reactions was used for all assays. Dedicated pre PCR pipettes were used and all tubes and filter tips were quality assured to be

RNase and DNase free. To prevent contamination after amplification, all post PCR manipulations were carried out in a separate room.

2.6.1 Qualitative PCR

Qualitative PCR was used as a sensitive method for finding mutations in diagnostic samples and the subsequent detection of the marker in follow-up samples for MRD. All qualitative PCRs were done on the ABI 9700. cDNA was used as template for all of the AML screening PCRs including t(8;21), Inv 16, t(15;17), *FLT3* ITD (semi quantitative) and *MLL* PTD. All of the original qualitative AML screening assays used between 1.5 and 5µl cDNA starting material and were amplified in two successive (nested) PCRs to increase the sensitivity. The exceptions were the *FLT3* ITD, the *MLL* PTD and the *ABL* PCR which were amplified for only one round. The PCRs were run using standard laboratory protocols with a positive control consisting of a known positive from either a cell line or confirmed patient sample and a negative control consisting of water included from the reverse transcription step. Every sample had an *ABL* PCR run as a control for both RNA quality and the presence of PCR inhibitors. Any sample lacking an adequate *ABL* result was not included in the study.

The Biomed-1 PCR protocol was established in order to create standardisation between laboratories for molecular testing by PCR (van Dongen *et al.* 1999). As most laboratories have set up their own PCR tests from publications or in-house methods there is a considerable amount of variation form one lab to another, preventing accurate comparisons of test results between centres. In 1999, the Biomed-1 primers were designed to detect the most common mutations occurring in AML. The primers were selected specifically so that all tests could be amplified under the same cycling conditions and reagent concentrations, enabling all of the tests to be run simultaneously on the same PCR machine. As the tests needed little or no optimisation, laboratories using Biomed-1 primers would be able to compare results. The

t(15;17), t(8;21) and inversion 16 PCRs are all Biomed-1 based and use the same cycling conditions and reagent concentrations as given:

The first round cDNA was added to make a total volume of 25μl per sample containing: 2mM $MgCl_2$, 200μM dNTPs, 1X PCR Buffer II ([1mM] Tris-HCl, pH8.3, [5mM] KCl), 0.4μM of each primer, and 0.025Units/μl Taq Gold. (Inversion 16 dNTP and $MgCl_2$ concentrations were slightly less as only 1.5μl of cDNA was used and was compensated for by adding more reagent buffer.)

Conditions for amplification were 94°C for 10 minutes, 35 cycles of 94°C for 30 seconds, 65°C for 60 seconds, 72°C for 60 seconds and then the block temperature was reduced and held at room temperature.

The second round master mix was composed of: 200μM dNTPs, 1mM $MgCl_2$, 1X PCR Buffer II ([1mM] Tris-HCl, pH8.3, [5mM] KCl), 0.4μM of each primer in a total volume of 24μl (inversion 16 dNTP concentrations were slightly less). The master mix, minus the first round template, was made up in the pre PCR hood. When preparation was completed and 24μl was aliquotted to each sample tube, the tubes were taken to post PCR for the addition of 1μl of first round product. Cycling conditions were the same as the first round. Samples were then electrophoresed on an agarose gel for analysis.

The only optimisations required for the biomed-1 primers were varying amounts of cDNA to be used in the first round PCR or $MgCl_2$ concentrations in the second round; these are noted in the individual PCR sections.

2.6.1.a   t(8;21) with RUNX1-RUNX1T1

The t(8;21) PCR detects the fusion of the RUNX1 gene on chromosome 21 to the RUNX1T1 gene on chromosome 8 using a nested primer set. There are no known variant

fusion transcripts from the area of this translocation. The estimated log sensitivity of the PCR was between 1×10^{-4} and 1×10^{-5}. This means the assay is able to detect 1 t(8;21) transcript in a total of between 10,000 and 100,000 transcripts.

For the first round amplification 5µl of cDNA was added to a Biomed-1 master mix to equal 25µl. The first round primers were:

AML-1A: 5'- CTA CCG CAG CCA TGA AGA ACC-3'

***ETO*-B:**   5'- AGA GGA AGG CCC ATT GCT GAA-3'

One microlitre of the first round amplification was added to the Biomed-1 second round master mix to give a total volume of 25µl. The second round primers are shown below. The expected sizes seen on an agarose gel were: t(8;21): 260bp (395bp first round).

***AML1*-C:**   5'- ATG ACC TCA GGT TTG TCG GTC G-3'

***ETO*-D:**   5'- TGA ACT GGT TTC TGG AGC TCC T-3'

2.6.1.b   t(15;17) with PML-RAR

The t(15;17) PCR detected the fusion transcript of the PML gene on chromosome 15 and the RARα gene on chromosome 17 using a nested set of primers. There are three different breakpoint regions in PML, bcr 1, bcr 2 and bcr 3, which can create a range of PCR product sizes. The bcr 1 and bcr 3 breakpoints make up the bulk of the transcripts seen with frequencies of 55% and 40% respectively. Both bcr 1 and bcr 3 are "stable" breakpoint regions and always produce the same size for their respective products. Bcr 2 on the other

hand, has a range of 168bp where the PML break can occur and therefore the product can vary in size from patient to patient. To produce efficient and robust PCRs two different sets of PCR primers were designed, one for bcr 1 and 2 and another for bcr 3. By using these primer sets the products were kept under 500bp and could be amplified more efficiently. Initial patient analysis entailed two PCRs done using both of the primer sets so that any of the potential mutations could be detected. Subsequent PCRs were done using the primer set appropriate for the breakpoint seen in the primary analysis. The estimated log sensitivities of both PML-RARα PCRs were between $1x10^{-3}$ and $1x10^{-4}$. This means the assay is able to detect 1 t(15;17) transcript in a total of between 1,000 and 10,000 transcripts.

For the first round amplification 5μl of cDNA was added to a biomed-1 master mix to equal 25μl. The first round primers were:

Bcr 1 and bcr 2:

PML-A1:	5'- CAG TGT ACG CCT TCT CCA TCA-3'
RARα-B:	5'- GCT TGT AGA TGC GGG GTA GA-3'

Bcr 3:

PML-C1:	5'- TCA AGA TGG AGT CTG AGG AGG-3'
RARα-B:	5'- GCT TGT AGA TGC GGG GTA GA-3'

One microlitre of the first round amplification was added to the biomed-1 second round master mix to give a total volume of 25μl. The concentration of MgCl₂ in the second round was adjusted to a final concentration of 2mM. The second round primers were:

Bcr 1 and bcr 2:

        PML-A2:   5'- CTG CTG GAG GCT GTG GAC-3'

        RARα-D: 5'- CTG CTG CTC TGG GTC TCA AT-3'

Bcr 3:

        PML-C2:   5'- AGC GCG ACT ACG AGG AGA T-3'

        RARα-D: 5'- CTG CTG CTC TGG GTC TCA AT-3'

The expected sizes seen on an agarose gel were:

Bcr1: 214bp (381 bp first round)

Bcr2: 43-211 bp (210- 378 bp first round)

Bcr 3: 289 bp (376 bp first round)

2.6.1.c *Inversion (16) with CBFB-MYH11*

This PCR reaction was designed to detect either the pericentric inversion of chromosome 16 or the t(16;16). There are three main "types" of aberration, A, D and E that occur at a rate of 88%, 5% and 5% respectively. The type A PCR was to be run as a first level of detection. If the type D or E product was present it was seen as a band at 1500-1600bp, if detected the PCR was repeated using the D and E primer set to determine which of these two mutations had occurred. The D and E primers are positioned to create a more efficiently sized PCR product of between 300 and 600bp. On follow-up samples the primer set appropriate to the type identified was used. The estimated log sensitivity of the assay for all types was between 1×10^{-4} and 1×10^{-5}. This means the assay is able to detect one inversion

16 transcript in a total of between 10,000 and 100,000 transcripts. For the first round amplification 1.5µl of cDNA was added to a biomed-1 master mix to equal 25µl. The first round primers were:

Type A:

      CBFB-A: 5' –GCA GGC AAG GTA TAT TTG AAG G - 3'

      MYH11-**B2:**   5'-TCC TCT TCT CCT CAT TCT GCT C -3'

If the patient was known to be type D or E, the primer *MYH11*-B1 was used in place of *MYH11*-B2.

      Type D or E: *MYH11*-B1   5' –TGA AGC AAC TCC TGG GTG TC -3'

One microlitre of the first round amplification was added to the biomed-1 second round master mix to give a total volume of 25µl. The second round primers were:

Type A:

      CBFB-C: 5' - GGG CTG TCT GGA GTT TGA TG -3'

      MYH11-**D2**:   5'- CCT GAG CGC CTG CAT GTT-3'

If the patient was known to be type D or E, the primer *MYH11*-B1 was used in place of *MYH11*-D1.

      Type D or E: *MYH11*-D1:   5' –TCC CTG TGA CGC TCT CAA CT - 3'

The expected sizes seen on an agarose gel if using the Type A primer set were:

Type A: 271 base pairs (1st round - 418bp)

Type D or E: ~1500bp

The expected sizes seen on an agarose gel if using the Type D and E primer set were:

Type D: 338 base pairs (1st round - 155bp)

Type E: 545 base pairs (1st round - 362bp)

2.6.1.d FLT3 ITD PCR

The *FLT3* PCR was used to detect ITDs occurring in exon 14 of the *FLT3* gene on chromosome 13. The protocol was used as described by (Kiyoi *et al.* 1999) with modifications. The original PCR was designed to amplify DNA spanning the region where ITDs occur, this was adapted to work with cDNA to streamline the AML screening process by using the RT step to produce cDNA for use as template. The reagent concentrations of the PCR were optimised for use with cDNA template instead of genomic DNA. The product size of the *FLT3* ITDs was variable, however, the mutation was rarely homozygous and the wild type (WT) product was present at a constant size of 237bp. To make the PCR product visible on a fluorescent detection system the forward primer was labelled on the 5' end with the fluorescein dye 6-fam. Using a semi quantitative method of analysis on the ABI Genescan 310, the *FLT3* product can be expressed as a ratio of the amount of ITD expressed over the total amount of *FLT3* expression. This enabled a level of sensitivity of 1-2% of ITD to total *FLT3* expressed to be reproducibly detected. A result was recorded as positive if the ITD level was higher than 1%. This level of sensitivity also allowed for the use of the *FLT3* ITD PCR for MRD detection in follow-up samples.

Five microlitres of cDNA was used in a reaction of total volume 25µl with 2mM MgCl₂, 50mM KCl, 10mM Tris-HCl, 200µM dNTPs, 0.625 Units Taq Gold, and 0.8µM of each primer. Samples were incubated at 95°C for 10 minutes followed by 30 cycles of: 94°C for 1 minute, 56°C for 1 minute, 72°C for 1 minute, and a final extension at 72°C for 10 minutes. The primer sequences were:

FLT3 F: 5' –6-fam GCA ATT TAG GTA TGA AAG CCA GC - 3'

FLT3 R: 5' – CTT TCA GCA TTT TGA CGG CAA CC - 3'

The expected sizes seen using capillary electrophoresis were:

FLT3 WT: 237bp

FLT3 ITD: the size of the ITDs ranged from 3-480bp, which was seen as a product added to the size of the wt (240- 717bp).

2.6.1.d.i   Expression of results

The percentage of *FLT3* ITD was the amount of *FLT3* ITD RNA expressed relative to the total amount of *FLT3* RNA expressed:-

FLT3 ITD % = (Area of *FLT3* ITD/(Area *FLT3* ITD+ Area *FLT3* WT))*100

2.6.1.e   MLL PTD

For analysis of the *MLL* PTD, occurring from exon 3 to exon 9,10, or 11 in the *MLL* gene, the region from exon 9 to exon 4 (Section 1.2.3.a), was amplified as previously described with modifications (Schnittger *et al.* 1998). The amplification detected only the PTD, no product was produced if the mutation was absent. Due to *MLL* PTDs being found in

samples from healthy, non-leukaemic individuals (Marcucci *et al.* 1998; Schnittger *et al.* 1998), the optimisation of the PCR sensitivity was critical. A number of modifications were systematically tested and the results are given in Chapter Three (Section 3.3.1). cDNA (0.5μl) was used in a reaction with 1mM $MgCl_2$, 50mM KCl, 10mM Tris-Cl, 200μM dNTPs, 0.625 Units Taq Gold, and 0.2μM of each primer. Samples were incubated at 95°C for 10 minutes followed by 35 cycles of: 94°C for 1 minute, 63°C for 1 minute, 72°C for 1 minute, and a final extension at 72°C for 10 minutes. The results were analysed by gel electrophoresis. 10μl of product was run on a 2% agarose gel containing ethidium bromide (EtBr). The primer sequences were:

MLL **6.1**:5'-GTC CAG AGC AGA GCA AAC AG-3'

MLL **E3AS**: 5'-ACA CAG ATG GAT CTG AGA GG-3'

2.6.1.f   ABL PCR

The *ABL* PCR was run for every sample to confirm that the RNA was amplifiable and determine whether there was contaminating DNA present. Any failure of the *ABL* PCR was attributed to improper sample handling, excessive time between collection and isolation or inhibitors present in the PCR resulting in degradation of RNA.

The *ABL* PCR from (Lion 1996a; Lion 1996b) was adapted by modifying the primer sequences to improve the PCR efficiency. 5μl of cDNA was used in a reaction with 2mM $MgCl_2$, 50 mM KCl, 10mM Tris-Cl, 200 μM dNTPs, 0.625 Units Taq Gold, and 0.8μM of each primer. Samples were incubated at 95°C for 10 minutes followed by 36 cycles of: 94°C for 1 minute, 58°C for 1 minute, 72°C for 1 minute, and a final extension at 72°C for 10 minutes. The results were analysed by gel electrophoresis. 10μl of product was run out on a 2% agarose gel containing EtBr. The primer sequences were:

ABL S:5' – AGC ATC TGA CTT TGA GCC - 3'

ABL-A:5' – TTA TAG CCT AAG ACC CGG - 3'

The expected sizes seen on an agarose gel were (overpage):

RNA 183bp

DNA ~600bp

2.6.2   Real Time Quantitative PCR (RQ-PCR)

Real time quantitative PCR is a technique whereby the amount of starting template
can be calculated from the crossing point of the sample's amplification compared to the
crossing points of a standard curve. By normalising against a recognised control gene (*ABL*)
samples are said to have a relative copy number calculated. This is in comparison to an
absolute number which is calculated from a standard curve of a sample with a known value.
Whereas in qualitative PCR the end point values are read as results that give a qualitative
result, the RQ-PCR can not only determine whether a transcript was present but also how
much of it was present. The disadvantage of this methodology was that some sensitivity was
lost as the RQ-PCR was only a single round test compared to the nested protocol of many
qualitative tests. Another danger when looking for an absolute number was that minor
differences in the efficiency of the amplification can cause skewing of results between
different samples and/or the sample and the control gene, leading to misinterpretation of the
final copy number. This can be protected for with the use of a properly selected internal
control (Sambrook & Russel 2001).

All of the QPCRs were performed using the LightCycler version 1.2 and the LightCycler Faststart DNA Master Syber Green kit. In this system, the RQ-PCR assay determines the amount of DNA produced after every round of amplification by measuring the fluorescence emitted by Syber Green, a dye specific for double stranded DNA. The crossing point (Cp), the cycle number at which the level of fluorescence exceeds the background, was calculated. Samples with a greater amount of template will reach the Cp sooner and have a lower cycle number than samples with less template. It is important to note that the reason that this method is necessary is due to the plateau reached during most PCR reactions. As a PCR reaction proceeds with exponential expansion the availability of reagents becomes rate-limiting and hence the amount of product plateaus. Theoretically, this means that the same product endpoint can be achieved with a sample beginning with 200ng of DNA and a sample that started at 1 µg DNA.

The amount of product was determined from the quantity of fluorescence produced from double stranded (dsDNA) labelled with Syber Green. As any dsDNA would be labelled including non-specific amplification and primer dimers, the melting temperature (Tm) of the product was used to determine the specificity of the assay. To create specificity a melting curve analysis needed to be run where the Tm of the desired amplicon was determined. At the end of the amplification the products are allowed to renature and fluoresce as the temperature is slowly increased. The increasing temperature causes the dsDNA to denature as its characteristic Tm is reached. Features such as length, nucleotide composition, secondary structure and mismatches can all affect the Tm of DNA. When the Tm of a product is reached, the DNA denatures into single stranded (ssDNA) and ceased to fluoresce. This drop in fluorescence was calculated to be the Tm of the product. Any non-specific amplification can be identified by the different Tm exhibited, in the same way as size was used to differentiate between products run out on an agarose gel.

2.6.2.a   MLL PTD RQ-PCR

A two-step quantitative PCR (the RT and the PCR done in two independent reactions) was done to determine the gene dosage of the *MLL* PTD and S14 in AML patients. The benefits of performing a two-step RT-PCR procedure versus a one-step are: 1) the original cDNA from the AML screen could be used in the RQ-PCR for both the *MLL* PTD and the *ABL* assay and 2) the sensitivity of the assay was increased.

The *MLL* PTD RQ-PCR was performed according to the manufacturer's protocol using the original cDNA from the AML screen. In brief, $0.5\mu l$ of cDNA was added to a glass capillary that contained 1mM $MgCl_2$, $0.2\mu M$ of each primer Light Cycler Buffer and $2\mu l$ of Syber Green for fluorescent detection in a total volume of 20 μl. The glass capillaries used by the light cycler allows heat to transfer efficiently to the sample and the amount of fluorescence to be measured as the reaction takes place in the machine. The primers *MLL* 6.1 and *MLL* E3AS from the qualitative PCR were used for the RQ-PCR. In the initial conditions samples were incubated at 95°C for 10 minutes followed by 45 cycles of: 95°C for 10 seconds, 65°C for 5 seconds, 72°C for 10 seconds with fluorescent acquisition performed after the 72°C extension step. A melt programme of 95°C for 0 seconds, 65°C for 15 seconds, then increasing to 95°C at a rate of 0.1°C/seconds was used to determine the Tm. Crossing point and Tm were calculated using LightCycler software. A detailed description of the optimisation and quantitation to achieve the final RQ-PCR protocol is described in Chapter 4.

2.6.2.b S14 RQ-PCR

The expression data generated from the real-time PCR was normalised to a housekeeping gene, S14 (gene encoding for the 40S ribosomal subunit protein S14) (Rhoads, Dixit, & Roufa 1986). The control gene was recommended by Roche for relative gene expression studies and used by the department in other publications (Guinn *et al.* 2005) et al, 2005;(Walsby *et al.* 2008). The S14 RQ-PCR was performed using 1µl of cDNA, 0.5µM primers and 4mM $MgCl_2$ in a total volume of 20µl. The amplification conditions were 95°C for 10 minutes; 40x 95°C for 3 seconds, 60°C for 5 seconds, 72°C for 10 seconds; followed by a melt step of 70°C to 95°C at 0.1°C/second. Using the Cp as a measure of quantity the *MLL* PTD was then normalised against the Cp value for S14.

2.6.2.c RQ-PCR for Validation Genes

The details for the design and optimisation of primers used to amplify the validation genes in are given in the Materials and Methods of Chapter 5. *ABL* was used as a normalisation gene in the validation RQ-PCRs. The RQ-PCR conditions for all validation genes as well as *ABL* were 1µl of cDNA, 3mM $MgCl_2$, 0.5µM forward and reverse primer and 1µl of Syber Green in a total volume of 10µl. The amplification conditions were 95°C for 10 minutes; 45x 95°C for 3 seconds, 60°C for 5 seconds, 72°C for 12 seconds; followed by a melt step of 60°C to 95°C at 0.1°C/second for all genes.

2.7   Electrophoresis

2.7.1   Gel Electrophoresis

Gel electrophoresis analysis was performed by loading PCR products onto an agarose gel and applying an electric current. The electric field generated causes the negatively

charged DNA molecules to migrate towards the positive pole separating by size as it moves. Larger fragments move more slowly and smaller fragments move more quickly through the agarose matrix. Comparison to a DNA marker of known size allows for accurate sizing of PCR products. The DNA was visualised by staining the gel with EtBr, a fluorescent dye that intercalates with DNA so that it can be seen when exposed to ultra violet light. All of the PCRs and the QPCRs, with the exception of the *FLT3* ITD assay, were run out on agarose gels containing EtBr to separate and visualise the DNA products.

A solution of 2% agarose (weight to volume) was made with 100ml 0.5x tris borate EDTA (TBE) (final concentration of 44mM Tris-borate and 1mM EDTA, pH 8.3). The solution was heated to until the agarose was dissolved and then cooled to 55-60°C. EtBr (0.5μg/ml) was added and the agarose was poured into a casting tray followed by the insertion of a comb to created wells. After the gel had solidified, 10μl of sample was combined with loading dye, loaded into the well and an electric field of between 100 and 170 volts was applied for 20 to 40 minutes. DNA sizes were determined by comparison to a DNA size ladder. The DNA was then be visualised on an ultraviolet light box and photographed using Polaroid 667 film.

2.7.2 Capillary electrophoresis

The *FLT3* ITD products were amplified with a 5' fluorescently labelled primer and were then analysed by running 1μl of the denatured product via capillary electrophoresis using an ABI prism 310 Genetic analyzer according to the manufacturers protocol. Briefly, the analysis was done using performance optimised polymer 4 (POP 4), 1x Buffer with EDTA, and 5 seconds injection time at 15 KV. 1μl of PCR product was added to 15μl of highly deionised formamide with ROX 500bp marker, heat denatured for 3 minutes at 95°C, cooled to 37°C and run at 15 KV, 60°C for 30 minutes.

2.8   Sequencing

All samples positive for the *MLL* PTD mutation were sequenced in the forward and reverse orientation to confirm the sequence. All sequencing was performed using the ABI Big Dye 3.1 Sequencing Kit according to the manufacturer's instructions. The PCR products were run out on an agarose gel to verify the product and the number of bands.

Single band products were purified using the Qiagen PCR Purification Kit according to manufacturer's instructions. Five volumes of buffer PB was added to the PCR sample and applied to the QIAquick column and centrifuged at ~17,900xg for 45 seconds. After discarding the flow-through, the column was washed with 0.75 ml PE buffer and centrifuged for 45 seconds. The eluate was discarded and the column spun for an additional minute. To elute the purified DNA 50μl water was added to the centre of the membrane and centrifuged for 1 minute. If the amplification yielded multiple bands, the individual bands were cut out from the gel and purified using the Qiagen Gel purification Kit following the manufacturer's protocol. Briefly, the DNA fragment was cut from the agarose gel using a clean scalpel. The sample was weighed and three volumes of Buffer QG to one volume of gel were added. The sample was incubated for 10 minutes at 50°C, and vortexed every three minutes to help dissolve the gel. One gel volume of isopropanol was added and the tube was mixed. The sample was applied to the QIAquick spin column and centrifuged at ~17,900xg for one minute. The eluate was discarded, 0.5ml of Buffer QG was added to remove all traces of agarose, and the spin was repeated. The DNA was washed with the addition of 0.75ml of Buffer PE and a one minute spin. Eluate was discarded and the spin repeated to remove all traces of ethanol. The DNA was eluted with the addition of 50μl and a 1 minute spin. The purified products were analysed on the spectrophotometer to determine the concentration

(Section 2.4.3) using the formula of 50μg/ml per 1 OD for DNA. The amount of DNA used was relative to the size in base pairs of the DNA to be sequenced.

200-500bp	3-10ng
500-1000bp	5-20ng

The appropriate amount of template was added to 2μl of Ready Reaction mix, 1μl of BigDye Sequencing Buffer, 1.6pmol *MLL* 6.1 and water added to a final volume of 20μl. The programme used for cycle sequencing was: 96°C for 30 seconds, 50°C for 15 seconds, 60°C for 4 minutes repeated for 25 cycles.

The Qiagen Dye Ex 2.0 Spin Kit was used to remove the unincorporated dNTPs from the sequencing reaction. The manufacturer's instructions were observed as follows. The spin column was vortexed and then spun for 3 minutes at 2800 rpm on a Heraeus Biofuge to pack the resin. All 20μl of the sequencing reaction was dripped on to the centre of the resin bed and spun again for 3 minutes. The dNTPs remain in the column while the purified DNA elutes during centrifugation. The purified sequencing reaction was then taken to CBS to be electrophoresed on a 3130xl Genetic Analyzer. The results were analysed to verify the type of mutation.

2.9 Collection of Clinical Data

Excel spreadsheets were designed and updated for over 386 patient samples. The clinical details were collected and collated for all patients after selecting for samples with satisfactory PCR controls. The non-trial information was collected manually from the Hospital Telepath System and trial data was provided by Dr. Rob Hills, Cardiff University. Information collected included the date of birth (DOB), date of collection, diagnosis, age at diagnosis, trial number if applicable, WBC and BM blast percentage at diagnosis, AML RT-

PCR screen status, sample type, cytogenetic data, cytogenetic risk factor, sex, FAB, molecular

marker data for *FLT3* TKD, *NPM1* and *WT-1*, as well as post diagnostic data for trial patients.

2.10 Data Obtained From Other Sources

A portion of the data analysed in this study was produced by other groups. This

included *NPM1* and *FLT3* TKD results provided by the Cardiff University Department of

Haematology and the *WT-1* expression level results kindly provided by Professor John Yin,

Manchester Royal Infirmary. The Department of Haematology at Cardiff University has

collected GEP data that can be analysed for the determination of gene expression in AML.

Around 300 AML patient GEPs are available, derived from two Affymetrix GeneChip

platforms including the human genome HG 133A and the HG-U133 plus 2. These procedures

are described in brief below and the protocols are provided (Appendix B).

The analysis of the variable mutation sites in *NPM1* and *FLT3* TKD are performed

using the PyroMark ID (Biotage, Uppsala, Sweden) which provides quantitative mutation

analysis by using pyrosequencing to analyse short sections of DNA directly from PCR

templates. cDNA from the reverse transcription as described above (Section 2.5) was

amplified using the same amplification conditions but with primers specific to the mutation

being analysed; (94°C 10 min x1; 50x 94°C 30 seconds, 65°C 1 minute, 72°C 1 min; 72°C 10

min). The success of the amplification is verified before the PCR products are analysed on a

PyroMark ID programme individualised for the gene of interest. The region where mutations

occur was sequenced and the mutation type was determined by comparison with a set of

known mutation profiles.

The diagnostic expression level of *WT-1* in 133 AML 15 trial patients was determined

by quantitative PCR. The assay used primers and fluorescent hybridisation probes located in

exon 7 and 8 of the gene to determine if the gene was overexpressed at presentation. *WT-1* copy number was normalised using the *ABL* gene (Liu-Yin *et al.* 2008)

2.11 Gene expression profiling

Mononuclear cells were isolated by Ficoll separation and then pelleted for RNA isolation. RNA isolation for the Affymetrix analysis was performed with either Trizol (HG-U133A) as described above (Section 2.4) or using the Qiagen RNeasy protocol (HG-U133 Plus2) (Appendix B). Comparison of probe set expression levels between samples with RNA isolated by different methods indicated that biological variation due to acute leukaemia overrode any differences that arose from RNA isolation methodology (Campo Dell'Orto 2007). The other slight modification in procedure between the chips was the *in vitro* transcription time. The RNA was quantified using a spectrophotometer and assessed for quality using the Agilent Bioanalyser which visualises and quantitates the RNA peaks (Agilent Technologies UK Ltd, Stockport, Cheshire, UK). RNA samples with insufficient concentration were to be efficiently transcribed were concentrated by re-precipitation. The first reverse transcription was performed with a poly-T oligomer to create the first strand cDNA (Figure 2.1). This was followed by the synthesis of the second strand so that a double stranded cDNA solution was produced. The solution was purified with a spin column and then an *in vitro* transcription (IVT) reaction was carried out to create biotin labelled cRNA. The cRNA was purified with a spin column and then fragmented for use in the hybridisation cocktail. Fragmentation was verified by visualisation using the Agilent Bioanalyser. The hybridisation cocktail was heat denatured before adding to the pre-hybridised Affymetrix GeneChip and incubated. The GeneChip was washed and stained on the fluidics station

before being scanned and fluorescent data collected. The normalisation process used for the

.cel files is described in Chapter 5, Section 5.2.2.
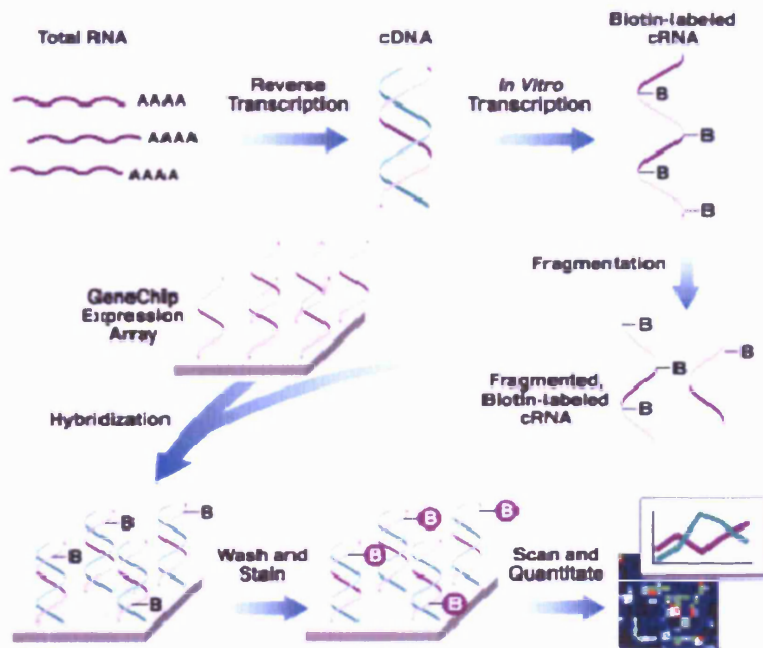
Figure 2.1 Diagram of Affymetrix GeneChip Protocol
Brief description of the protocol for the standard eukaryotic gene expression assay.
Affymetrix ™.

3 Prognostic Significance of *MLL* PTD in AML Patients

3.1 Introduction and Objectives

Molecular markers are important in the accurate prognosis of CN-AML patients who are currently classified by their lack of a prognostic marker. Diagnostic karyotypes are used to stratify patients into three prognostic groups favourable, intermediate and adverse. However, the technique of karyotyping only identifies large chromosomal aberrations, which means that many underlying mutations go undetected. In fact, approximately 45% of all AML cases are described as cytogenetically normal and are then categorised as intermediate prognosis due to the absence of a karyotypic marker (Byrd *et al.* 2002; Grimwade *et al.* 1998; Grimwade *et al.* 2010). The clinical outcome for cytogenetically normal AML patients is variable, with the 5 year overall survival between 35% and 45% (Byrd *et al.* 2002; Grimwade *et al.* 1998; Slovak *et al.* 2000). New molecular markers would add further stratification to the intermediate group, so patients could be treated in a manner appropriate to their level of risk (Grimwade & Hills 2009).

In recent years, molecular genetic techniques have been used to identify several new sub-microscopic mutations associated with normal karyotype AML. These include mutations in the *FLT3* and *NPM1* genes (Boissel *et al.* 2005; Dohner *et al.* 2005; Falini *et al.* 2005; Nakao *et al.* 1996; Schnittger *et al.* 2005b). The *FLT3* ITD was one of the first markers to be identified and is found in approximately 25% of newly diagnosed AML and 30% to 35% of all CN-AML patients (Section 1.2.4.b) (Boissel *et al.* 2005; Chillon *et al.* 2004; Kottaridis *et al.* 2001; Whitman *et al.* 2001). The variable length duplication creates an in-frame transcript whose protein is constitutively activated and promotes survival and proliferation of leukaemic blasts (Gilliland & Griffin 2002). This marker is associated with leukocytosis, a high BM blast count, an increased risk of relapse and reduced survival (Kottaridis, Gale, & Linch 2003; Small 2006; Stirewalt & Radich 2003). *NPM1* mutations are found in approximately 30% of

all AML patients (Suzuki *et al.* 2005a; Verhaak *et al.* 2005a) and 50% of CN-AML patients (Section 1.2.4.c) (Falini *et al.* 2005; Schnittger *et al.* 2005a; Suzuki *et al.* 2005b; Verhaak *et al.* 2005b). The function of the *NPM1* gene is not as well understood as the function of the *FLT3* ITD. However, the *NPM1* mutation is known to disrupt protein localisation and export, causing patients with this mutation to exhibit cytoplasmic rather than nucleic nucleophosmin. This altered expression pattern may disrupt the normal functioning of the protein leading to cancer pathogenesis (Falini *et al.* 2006). Patients with mutated *NPM1* are reported to have a favourable prognosis in terms of achievement of complete remission (CR) relapse free survival (RFS) and overall survival (OS) (Dohner *et al.* 2005; Falini *et al.* 2005; Gale *et al.* 2007; Schnittger *et al.* 2005b). The prognostic significance of these additional mutations is illustrated by the inclusion of the *NPM1* mutation as a provisional entity in the 4th edition of the WHO classifications of AML with recurrent genetic abnormalities (Swerdlow *et al.* 2008). The *MLL* PTD is another sub-microscopic mutation that may provide further prognostic information, particularly to patients with CN-AML (Section 1.2.4.a). The duplication was originally identified as a recurrent abnormality in AML patients with trisomy 11 identified as the sole abnormality (Bernard *et al.* 1995; Caligiuri *et al.* 1994; Caligiuri *et al.* 1996; Schichman *et al.* 1994a). However, it was soon recognised that the duplication was also found in AML patients lacking trisomy 11 and that the majority occurred in CN-AML, with reported frequencies of 6%-21% (Caligiuri *et al.* 1998; Schnittger *et al.* 2000; Yu *et al.* 1996). The preliminary papers on *MLL* PTD indicated a reduced duration of complete remission and reduced overall survival for patients with the *MLL* PTD (Caligiuri *et al.* 1998; Schnittger *et al.* 2000; Yu *et al.* 1996). The majority of these early studies assessed small numbers of AML patients, less than 100, from limited subtypes including cytogenetically normal AML and trisomy 11.

This chapter addresses the hypothesis that the *MLL* PTD is a poor prognostic indicator in diagnostic and CN-AML. In this study, the frequency and clinical characteristics of the *MLL* PTD were analysed in 386 (342 unselected diagnostic AML) patients with haematological disorders.

The specific objectives of the study were:

1) to determine the frequency of the *MLL* PTD in unselected and CN-AML

2) to assess the impact of the *MLL* PTD PCR assay on the proportion of patients without established prognostic markers (for cytogenetics and RT-PCR)

3) to determine the association of the duplication with cytogenetic abnormality, FAB classification, *FLT3* mutations, *NPM1* mutations and other characteristics of clinical presentation

4) to assess the prognostic impact of the *MLL* PTD on AML patients

3.2   Materials and Methods

3.2.1   Patient Samples

A total of 386 patients and 20 healthy controls were tested for the presence of *MLL* PTD. The healthy control samples were obtained from donor samples collected for bone marrow transplant. The patient samples were derived from two groups, Medical Research Council AML Trial-United Kingdom (MRC-UK) patients and non-trial patients assayed by the Haematology department's RT PCR AML screen assay. Of the 386 patients assayed, there were 342 AML, 18 AML follow up, 3 ALL and 23 that were not diagnosed as AML. The majority of the patients in the latter group were diagnosed with myelodysplastic syndrome (MDS) or myeloproliferative neoplasms (MPN) and were excluded from the primary analyses. The 342 diagnostic AML patients were composed of 302 *de novo* AML and 40 secondary AML (sAML) cases, these samples were analysed together unless otherwise stated. The clinical and molecular characteristics of these patients are shown in Table 3.1 and Table 3.2. The treatment protocols of the patients were described previously (Section 2.3.2 Materials and Methods). The 342 diagnostic AML patients studied fell into two groups, 117 were non-trial patients and 225 were enrolled into one of the MRC AML trials, 130 of the trial samples were randomly selected from the stored material maintained by the Cardiff University trials unit and the remaining 95 were processed directly by the Haematology department. There was no significant difference observed between the results for the trial data alone and trial and non trial data combined in the majority of tests performed, therefore all results shown in this study are a combination of trial and non trial patients unless otherwise stated. Results requiring post-diagnostic information were performed on the trial samples only due to the lack of outcome data available on the non-trial patients.

Characteristic			
Sex	*n=342*	**n**	**%**
Male		187	54.7
Female		155	45.3
Age Group	*n=342*	**n**	**%**
<35 years		52	15.2
35-60 years		131	38.3
≥60 years		158	46.2
Age	*n=342*	**years**	
Median		58.0	
Range		1-89	
WBC	*n=323*	**x10^9/L**	
Median		15.7	
Range		0.6-311.0	
Bone marrow blasts count	*n=295*	**%**	
Median		65.0	
Range		1-100	
Type of AML	*n=342*	**n**	**%**
de novo AML		306	89.5
sAML		36	10.5
FAB type	*n=342*	**n**	**%**
M0		29	8.5
M1		56	16.4
M2		86	25.1
M3		19	5.6
M4		46	13.5
M5		33	9.6
M6		6	1.8
M7		1	0.3
Bilineage		1	0.3
RAEB/RAEB-t		6	1.8
Not determined		59	17.3

Table 3.1 Clinical Characterisation of Diagnostic Patient Cohort
WBC: white blood count; L: litre; sAML: secondary AML (sAML included both AML that developed after a myelodysplastic syndrome and AML developed after therapy); FAB: French-American-British. Complete information was not available for some patients .RAEB: refractory anaemia with excess blasts; RAEB-t: refractory anaemia with excess blasts in transformation.

Characteristic	Patients (n)	Patients (%)
CytogeneticAbnormalities *n=342*		
t(15;17)	17	5.0
t(8;21)	15	4.4
inv(16)/t(16;16)	20	5.8
+8	10	2.9
+11	8	2.3
-5	1	0.3
-7	11	3.2
3q	2	0.6
t(6;9)	1	0.3
t(9;22)	2	0.6
t(11q23)	8	2.3
Complex karyotype (>5 chromosomal abnormalities)	29	8.5
Other chromosomal abnormalities	29	8.5
Normal karyotype	154	45.0
Not determined	35	10.2
Cytogenetic Prognostic Classification *n=342*		
Favourable	52	15.2
Intermediate	201	58.8
Adverse	44	12.9
Not determined	45	13.2
Molecular Abnormalities		
MLL PTD - no./total no.	17/342	5.0
FLT3 ITD - no./total no.	76/336	22.6
FLT3 TKD - no./total no.	25/200	12.5
NPM1 - no./total no.	52/181	28.7
WT1 over-expression - no./total no.	108/133	81.2

Table 3.2 Molecular Characterisation of the Diagnostic Patient Cohort
PTD: partial tandem duplication; n.: number; ITD: internal tandem duplication; TKD: tyrosine kinase domain.

3.2.2 Molecular Biology Techniques

PCR conditions were optimised to prevent amplification of *MLL* PTDs that were not relevant to AML (as described in Introduction 3.1). When the *MLL* PTD PCR was performed as near to the conditions described by Schnittger *et al.* (1998) as possible (the $MgCl_2$ concentration and buffer conditions were not stated) multiple PCR products were produced (Schnittger *et al.* 1998). The conditions for amplification were 50 ng/μl cDNA, 0.2 mM primers *MLL* 6.1 and *MLL* E3AS, 0.2 mM dNTPs, 2 mM $MgCl_2$, 1 mM Tris-HCl, pH8.3, 5 mM KCl and 0.025 units/μl Taq polymerase. The cycling conditions were 4 minutes at 94°C, followed by 35 cycles (1 minute at 94°C, 1 minute at 63°C, 1 minute at 72°C) then a final extension of 10 minutes at 72°C. Reducing the $MgCl_2$ concentration to 1mM and replacing Taq polymerase with TaqGold polymerase achieved more stringent results and prevented the amplification of products in *MLL* PTD negative patients (Figure 3.1). The final optimised conditions were described previously (Materials and Methods Section 2.6.1.e).

A Standard RT-PCR conditions:

M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M

480 bp
350 bp

- + - - - + - - - - - - - - - - - -
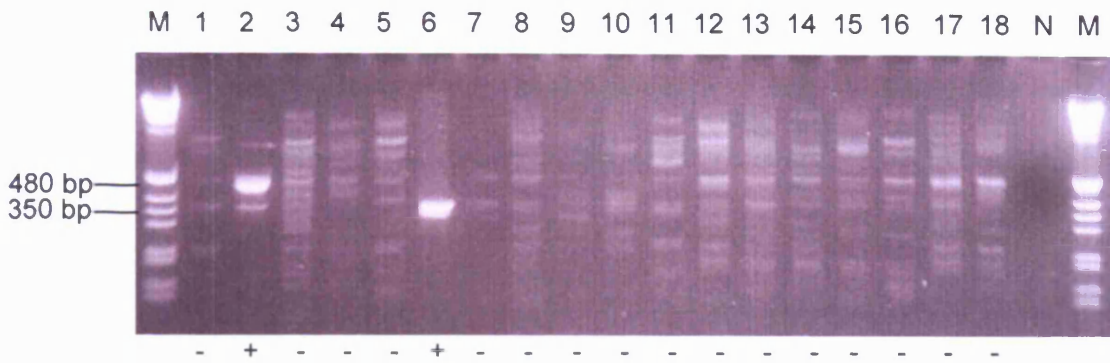
B Optimised RT-PCR conditions:

M 1 2 3 4 5 6 7 8 N M

480 bp
350 bp

e11/e3
e10/e3

- - + + - - - -
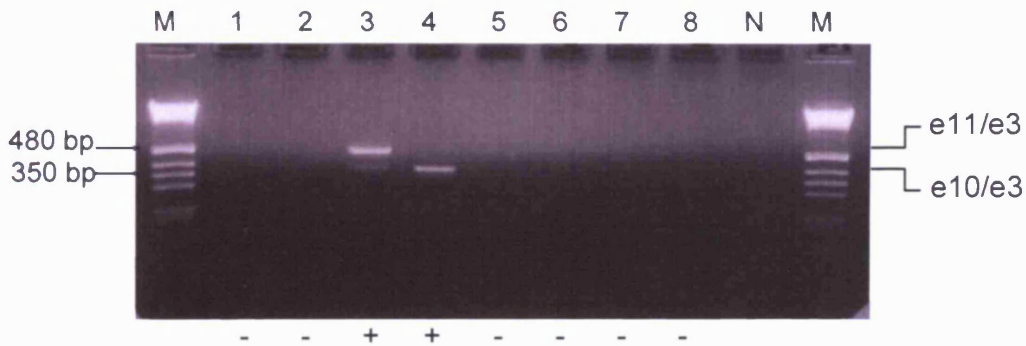
Figure 3.1 Optimisation of *MLL* PTD PCR
PCR products separated by size on an ethidium bromide stained agarose gel. A: Amplification of multiple *MLL* PTD transcripts from AML patients prior to PCR optimisation. Lanes 2 and 6 were positive for *MLL* PTD, all other numbered lanes are negative. B: Amplification of samples after PCR optimisation. Lanes 3 and 4 are positive for the duplication (lanes 2 and 6 in gel A), all other numbered lanes are negative. *MLL* PTD transcript types and sizes are indicated. M: Molecular weight standard; N: negative control containing water.

3.2.3 Endpoints and Statistical Analysis

Clinical outcome data were analysed for trial patients with post diagnostic information available. Complete remission (CR) was defined as less than 5% blasts in a normocellular BM with no requirement for peripheral blood count recovery. The two classifications of remission failure were induction death (ID) (treatment related death and/or hyperplasia occurring within 30 days of diagnosis), or resistant disease (RD) (failure of treatment to eliminate the disease including partial remission (5% to 20% blasts)). Overall survival is defined as the time from trial entry to death, relapse free survival is the time from trial entry to first event either relapse or death and the risk of relapse is the cumulative risk of relapse discounting death in CR.

Statistical analysis was performed using the SAS System 9.1 statistical software package (SAS Institute, Cary, NC, USA). Differences in the clinical and demographic data for *MLL* PTD positives were assessed using the Wilcoxon rank sum test (for continuous data), Mantel-Haenszel test for trend (for ordinal data), χ^2 test and Fisher's exact test. Odds ratios (OR) and hazard ratios (HR) with 95% confidence intervals were calculated. An OR or HR >1 indicated a worse outcome for the condition considered. Two tailed P-values below 0.05 were accepted as a marker of significance throughout the study. The probability of RR, RFS and OS were determined using the Kaplan-Meier method and survival curves were compared using a double-sided log-rank test for univariate analyses and Cox regression for multivariate. If p<0.05, models were adjusted for two sets of additional variables using logistic regression. The first set were the basic clinical variables (age, sex, secondary disease, cytogenetic prognostic group and WHO performance status) followed by molecular markers (*FLT3* ITD, *FLT3* TKD and *NPM1*).

3.3 Results

3.3.1 *MLL* PTD Qualitative PCR Optimisation

To identify *MLL* PTDs for use as positive controls, diagnostic AML samples were first screened with the PCR conditions described by Schnittger *et al.* (1998). The amount of cDNA used was increased from 25 to 50ng to ensure that amplification occurred, as some of the amplification conditions were not fully described (Section 3.2.2). These conditions amplified multiple bands in the majority of samples tested (Figure 3.1A); such bands represented *MLL* PTD transcripts that are produced at high levels of amplification in *MLL* PTD negative AML patients and healthy controls. The majority of the transcripts have a different sequence to those identified in AML and are not considered to be related to the disease (Marcucci *et al.* 1998; Schnittger *et al.* 1998). Further explanation of these *MLL* PTD transcripts is covered in the Discussion (Section 3.5). Of the 18 samples tested, two of the samples showed greater product amplification at the expected sizes of the *MLL* PTD transcripts associated with AML. When the PCR was repeated using conditions optimised to increase the stringency of the amplification, only the recognised *MLL* PTD transcripts (e10/e3 and e11/e3) were amplified. Three modifications were made to increase the stringency of the PCR; the template concentration was reduced to 25 ng, the concentration of $MgCl_2$ was decreased from 2mM to 1mM and Taq polymerase was replaced with TaqGold polymerase. TaqGold polymerase increases the specificity of PCR reactions by allowing a "hot start" technique to be employed. TaqGold was activated by a ten minute 94°C incubation (hot start) thus preventing non-specific amplification caused by low temperature amplification before the start of the cycling protocol (Chou *et al.* 1992). The specificity of the amplification was confirmed by sequencing each of the different transcripts and comparing with the *MLL* sequence. All three of the transcripts amplified matched the sequence of *MLL* and confirmed

that the PCR identified the partial duplication by fusion of either *MLL* exon 9, 10 or 11 with exon 3, creating an in-frame protein for translation (Figure 3.2, Figure 3.3). These samples were used as positive controls for future assays and subsequent positives were verified by basepair size.
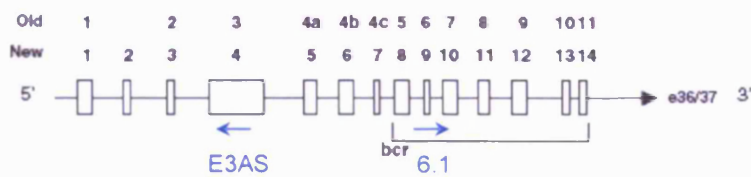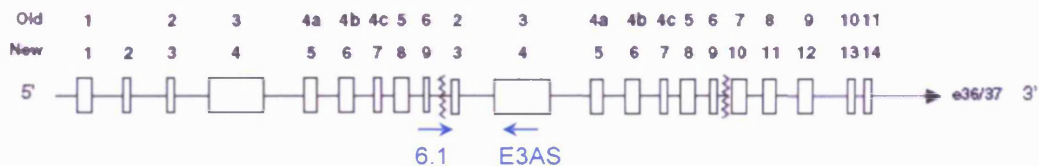
Figure 3.2 Schematic Stucture of the *MLL* PTD
Schematic structure of the 5' region of the *MLL* gene and the exon 9/exon 3 *MLL* PTD. The primers used to amplify the duplications are indicated in blue and are named according to the "Old" nomenclature. The "New" exon nomenclature ((Nilson *et al.* 1996) is used to designate the PTDs in this study. *BCR*: breakpoint cluster region.
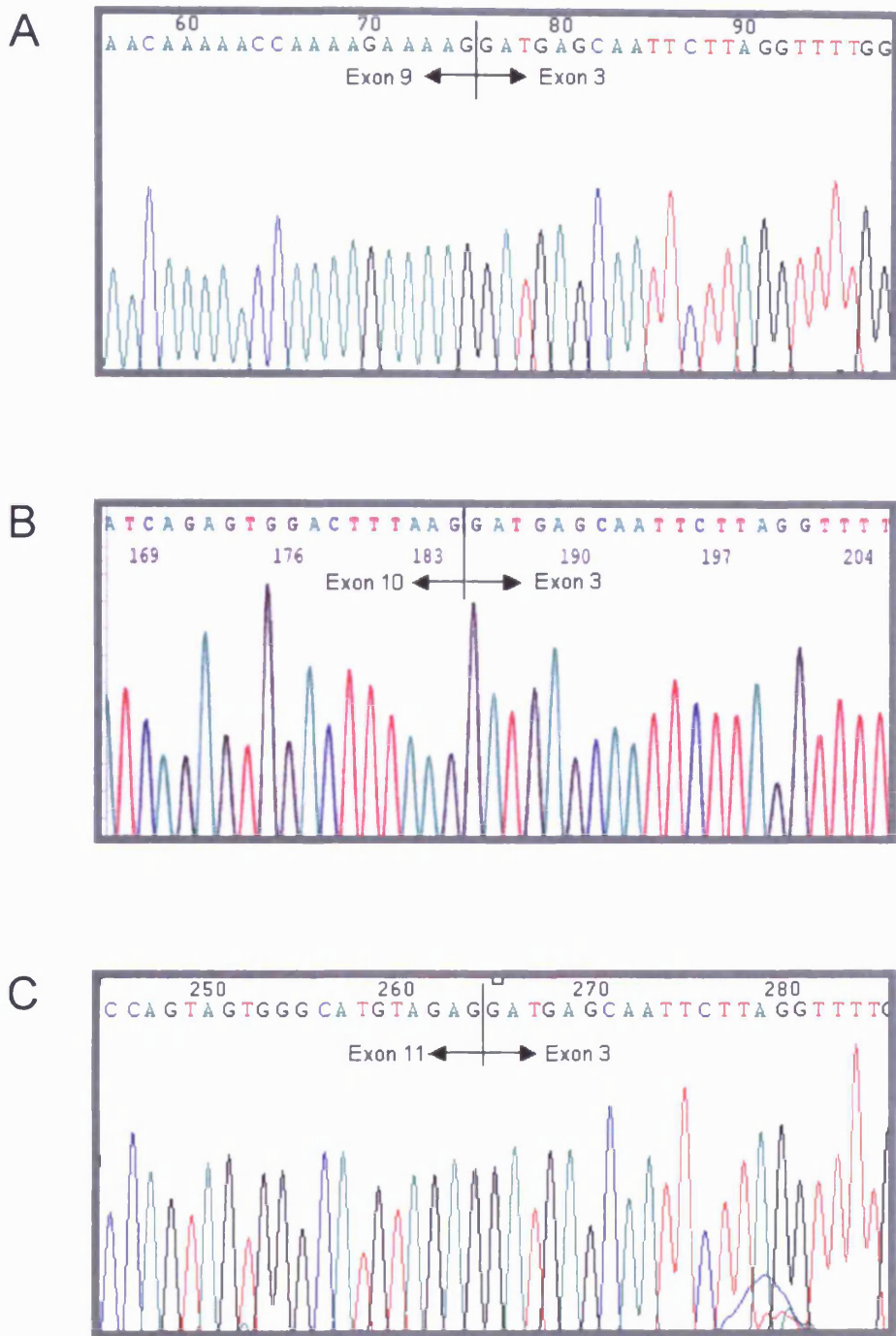
Figure 3.3 Sequence Analysis of the *MLL* PTD Transcripts
Characterisation of the DNA sequence of the three partial tandem duplications identified in
MLL-PTD.   A: Exon 9/exon3; B: Exon 10/exon3; C: Exon 11/exon3.

3.3.2   Incidence of *MLL* PTD

3.3.2.a   MLL PTD Incidence for all Diagnostic AML samples

In total, 386 patient samples and 20 healthy controls were analysed for the presence of the *MLL* PTD. All patients were either enrolled in MRC AML trials or were determined to need an AML PCR screen by their clinical details. The 20 healthy control samples were all negative for the duplication. There were 19 *MLL* PTD positives (4.9%) in the cohort of 386 patients; however, after completion of all clinical tests, the final number of diagnostic AML samples was determined to be 342 and the final number of *MLL* PTD positives was 17 (two were from relapse samples). To gain a broad picture of the frequency of *MLL* PTD in AML, the assay was not limited to a particular subgroup of patients; therefore *de novo* AML, secondary AML and all FAB subtypes were included in the overall analysis. *MLL* PTD products were detected in 5.0% (17/342) of patients overall,the frequency was 5.2% (16/306) in *de novo* AML and 2.8% (1/36) in sAML (Section 3.3.2.i) (Table 3.3). The clinical details of the patients positive for *MLL* PTD are shown in Table 3.4. Three different transcripts were identified by sequence analysis, the majority of which were the fusion of exon 9 to exon 3 (e9/e3 transcript). The e9/e3 transcript represented 70.6% (12/17) of the total positives, the e10+e11/e3 transcripts 23.5% (4/17) and the e10/e3 transcript 5.9% (1/17). In the 154 patients who were cytogenetically normal, 6.5% (10/154) were positive for the duplication. Thus, the incidence of the duplication was increased in the cytogenetically normal subgroup compared to the overall group but was not significant (p=0.3). The duplication also occurred in samples with cytogenetic abnormalities, albeit at a lower frequency. The incidence was 3.9% (6/153) with all aberrant karyotypes included, but dropped to 2.7% (4/150) when the 3 samples with trisomy 11 as the sole karyotypic abnormality were excluded. In summary, 5.0% (17/342) of the diagnostic AML patients had an *MLL* PTD. Six of the 17 cases had

chromosomal abnormalities including trisomy 11 (2), monosomy 7 (2), complex karyotype with multiple abnormalities (1) and unusually, an t(8;21) (1). Ten of the remaining patients were normal karyotype and one was not determined (Table 3.4). These results indicated the duplication was not limited to patients with a normal karyotype or to the intermediate prognostic group.

AML Cohort	total no.	MLL PTD +ve n	Frequency %
Overall	368	19.0	4.9
Diagnostic AML	342	17.0	5.4
de novo AML	306	16	5.2
sAML	36	1	2.8
CN-AML	154	10	6.5
Aberrant karyotype	153	6	3.9
Aberrant karyotype (minus +11)	150	4	2.7

Table 3.3 Prevalence of *MLL PTD* in Different AML Cohorts
Prevalence of the *MLL PTD* in different AML cohorts. PTD: partial tandem duplication; sAML: secondary AML; CN-AML; cytogenetically normal AML.

UPN	Sex	Age (years)	WBC (x10⁹/L)	Blasts (%)	MLL PTD Type	FLT3 ITD	FLT3 TKD	NPM1	FAB Type	Cytogenetic Abnormality	Remission Status	Clinical Details	AML Type
9	F	66	59.8	66	9/3	M	N/D	N/D	N/D	Normal	N/D	?ET to AML	s AML
14	F	47	4.4	83	10/3	WT	N/D	N/D	M1	-7	N/D	?AML	*de novo* AML
17	F	62	4.8	N/D	10/3+11/3	WT	N/D	N/D	M2	Complex	N/D	?Relapse	FU
41	M	64	60.0	63	10/3+11/3	WT	WT	WT	M2	Normal	CR	AML 14	*de novo* AML
49	F	61	3.0	56	9/3	WT	N/D	N/D	M2	Normal	N/D	?AML	*de novo* AML
68	M	38	8.0	73	10/3+11/3	WT	N/D	N/D	M2	-7(q)	N/D	?Relapse	FU
77	M	61	0.8	60	10/3+11/3	WT	WT	WT	M0	Normal	CR	AML 15	*de novo* AML
126	M	82	4.5	80	10/3+11/3	WT	N/D	N/D	M1	+11	N/D	?AML	*de novo* AML
150	F	35	3.6	10	9/3	WT	M	WT	N/D	N/D	CR	AML 15	*de novo* AML
173	F	43	1.0	56	9/3	WT	N/D	N/D	M2	Normal	CR	AML 15	*de novo* AML
188	F	72	1.9	50	9/3	WT	N/D	N/D	M6	Normal	N/D	?AML	*de novo* AML
262	F	72	125.0	90	9/3	M	WT	WT	M5	Normal	CR	AML 11	*de novo* AML
263	M	62	118.0	N/D	9/3	M	WT	WT	M1	Normal	CR	AML 11	*de novo* AML
282	F	65	2.6	63	10/3+11/3	WT	WT	WT	M2	-7	CR	AML 14	*de novo* AML
296	M	76	46.2	84	9/3	M	WT	WT	M1	Normal	RES DIS	AML 14	*de novo* AML
301	M	71	50.5	92	9/3	WT	M	WT	M1	+11	CR	AML 14	*de novo* AML
320	F	41	79.9	71	9/3	WT	WT	WT	M2	t(8;21)	CR	AML 15	*de novo* AML
370	M	60	1.0	87	9/3	WT	WT	WT	M1	Complex	RES DIS	AML 15	*de novo* AML
384	F	75	1.6	N/D	9/3	WT	WT	WT	M2	Normal	N	?AML	*de novo* AML

Table 3.4 Clinical Characteristics of *MLL* PTD Patients
Clinical characteristics of patients with the *MLL PTD*. The two relapse patients UPN 17 and UPN 68 were not used in further analyses. UPN: unique patient number; M:male; F: female; N/D: not determined; M: mutant; WT: wild type; CR complete remission; RES DIS: resistant disease; and sAML: secondary AML.

Karyotype data was available on 299 *MLL* WT AML patients, of these 2.7% (8/299) of patients had a trisomy 11; three of these as the sole abnormality and five in combination with other abnormalities. Trisomy 11 was present in 13.3% (2/15) *MLL* PTD patients and 2.1% (6/284) *MLL* negative patients (p=0.06) (Table 3.5). Furthermore, two of the three cases of trisomy 11 as the sole abnormality were *MLL* PTD positive (p=0.02) indicating *MLL* PTD is significantly over represented in this group of patients.

A	Trisomy 11 (sole and other abnormalities)	*MLL* WT n=284	%	*MLL* PTD n=15	%	p-value
	-	278	97.9	13	86.7	
	+	6	2.1	2	13.3	0.06

B	Trisomy 11 (sole abnormality)	*MLL* WT n=284	%	*MLL* PTD n=15	%	p-value
	-	283	99.6	13	86.7	
	+	1	0.4	2	13.3	0.02

Table 3.5 Frequency of *MLL* PTD in Trisomy 11 Patients
Co-occurrence of *MLL* PTD in patients with trisomy 11. A. *MLL* PTD association with all trisomy 11 patients. B. *MLL* PTD association with patients with trisomy 11 as the sole abnormality. WT: wild type; PTD: partial tandem duplication.

3.3.2.b Stability of MLL PTD in Relapse Samples

The stability of the *MLL* PTD as an MRD marker was assessed by the number of follow up samples that retained the marker upon relapse. Retention would increase the possibility of successfully using the duplication as an MRD marker. Eight of the 18 post-treatment samples were relapse samples (BM blast counts greater than 5%) and two were positive for the *MLL* PTD. Unfortunately, only one of these patients had a diagnostic sample available for testing. In this patient the same duplication, e10/e3 + e11/e3, was identified in both the diagnostic and relapse sample, indicating that the mutation was retained (Figure 3.4). The *MLL* PTD screening identified 8 patients for whom follow up material could be obtained. These patients were followed through treatment to determine if the duplication was an effective marker for MRD. The results of these studies can be found in Chapter 4.
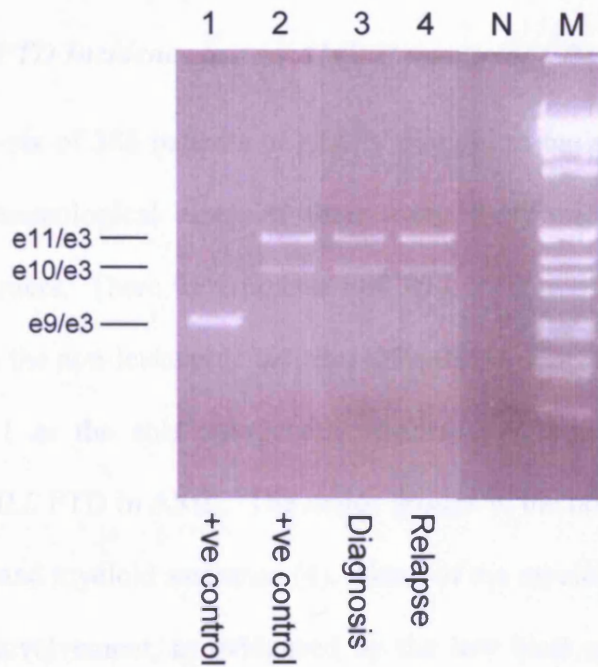
Figure 3.4 *MLL* PTD PCR of Patient at Diagnosis and Relapse
Lane 1 and 2 are positive controls, lane 3 is UPN 68 at diagnosis and lane 4 is UPN 68 at relapse. Transcript type and size are indicated on the left. M: molecular weight marker; N: negative control using water. This is a composite gel.

3.3.2.c MLL PTD Incidence in non AML Haematology Patients

Twenty-six of 386 patients originally thought to have AML were later determined to have other haematological diseases; three were diagnosed with ALL and 23 with non-leukaemic disorders. There were no cases of *MLL* PTD in diagnostic ALL patients (0%; 0/3) and no cases in the non-leukaemic patients (0%; 0/23). Interestingly, one of the ALL patients had trisomy 11 as the sole cytogenetic abnormality, which is associated with a higher frequency of *MLL* PTD in AML. The major groups in the non-leukaemic patients were MDS (11), MPN (5) and myeloid sarcomas (4). None of the myeloid sarcomas had any evidence of bone marrow involvement, as evidenced by the low blast counts observed; therefore AML abnormalities in the bone marrow would not be expected. Although the numbers tested were small, there was no evidence of the *MLL* PTD detected in any of the non-AML samples tested.

3.3.2.d MLL PTD Co-occurrence with Other Molecular Aberrations

The possible co-occurrence of *MLL* PTD with other molecular aberrations was examined in *FLT3* ITD, *FLT3* TKD, *NPM1* and *WT-1*. The *FLT3* TKD and *NPM1* data was provided by Professor A. K. Burnett, Cardiff University Trials Unit and the *WT-1* data on the Cardiff University MRC AML 15 samples was kindly provided by Professor John Yin, Department of Haematology, Manchester Royal Infirmary. The co-occurrence of the *MLL* PTD with the *FLT3* ITD and TKD mutations was analysed in the 342 diagnostic patients due to the co-occurrence of *MLL* PTD with either a *FLT3* ITD or TKD in 50% (6/12) of the *MLL* PTD samples. The *FLT3* ITD mutation was detected in 22.6% (76/336) of the patients assayed; 23.5% (4/17) of *MLL* PTD positive patients were *FLT3* ITD mutants compared with 22.6% (72/319) of *MLL* negative patients (p=1.0) (Table 3.6). *FLT3* TKD data was available for 200 patients and detected in 12.5% (25/200); 18.2% (2/11) of the duplication positive

patients were positive for the TKD mutation versus 12.2% (23/189) of the *MLL* negative patients (p=0.6). The number of patients available for the overall *FLT3* mutation status, dictated by the presence of either mutation, was 224 with 44.6% (100/224) mutants detected. Six patients positive for *MLL* PTD tested positive for a *FLT3* mutation (6/12; 50%) and 44.3% (94/212) of *MLL* negative patients had a *FLT3* mutation (p=0.8).

Despite the high proportion of *MLL* PTD patients harbouring *FLT3* mutations there was no significance associated with any of the aberrations. There was no significant correlation when the same analysis was performed in the CN-AML subgroup where there is a higher incidence of all three mutations. In summary, despite the high percentage of co-occurrence between *MLL* PTD and *FLT3* mutations the frequency was no higher than expected.

	MLL WT n	%	*MLL* PTD n	%	p value
FLT3 ITD	n=319		n=17		
-	247	77.4	13	76.5	
+	72	22.6	4	23.5	1.0
FLT3 TKD	n=189		n=11		
-	166	87.8	9	81.8	
+	23	12.2	2	18.2	0.6
FLT3 overall	n=212		n=12		
-	118	55.7	6	50.0	
+	94	44.3	6	50.0	0.8
NPM1	n=170		n=11		
-	118	69.4	11	100.0	
+	52	30.6	0	0.0	0.04
WT1 Over-expression	n=128		n=5		
-	20	15.6	0	0.0	
+	108	84.4	5	100.0	1.0

Table 3.6 Co-occurrence With Other Molecular Markers
WT: wild type; PTD: partial tandem duplication; no.:number; ITD: internal tandem duplication; TKD: tyrosine kinase domain; -: negative for mutation; +: positive for mutation.

The co-occurrence of the *NPM1* mutation and *MLL* PTD was analysed in the diagnostic AML cohort. The *NPM1* mutation assay was performed on 181 patients and 28.7% (52/181) were positive. Within the 11 *MLL* PTD patients there were no *NPM1* mutations detected (0/11; 0%) compared to 30.6% (52/170) *NPM1* mutations in the *MLL* negative group. There was a significant negative association between the two mutations (p= 0.04). Subsequent analysis of the CN-AML subgroup (n=93) reproduced this result (p=0.06). Taken as a whole, these results suggest that the mutations *NPM1* and *MLL* PTD have a negative association, but a larger cohort of patients would be required to confirm this finding. *WT-1* is generally over-expressed in AML and has been linked with poor prognosis (Liu-Yin *et al.* 2008; Paschka *et al.* 2007; Weisser *et al.* 2005a)(Garg *et al.* 2003b). The correlation of *WT-1* expression level with the presence of the *MLL* PTD was analysed in 133 diagnostic trial patients. High expression levels of *WT-1* were detected in all five *MLL* PTD positive patients (5/5; 100%) and in 108 *MLL* negative patients (108/128; 81.2%). The association of *MLL* PTD with over expression of *WT-1* was not statistically significant (p=1.0).

3.3.2.e *MLL PTD Incidence in Cytogenetic Prognostic Groups*

The frequency of the *MLL* PTD was examined in relation to the cytogenetic prognostic groups in *de novo* AML patients, to determine if the duplication was associated with a particular prognostic classification. In the 297 patients analysed, the duplication positive patients had 6.7% (1/15) favourable, 73.3% (11/15) intermediate and 20.0% (3/15) adverse classification types. This compared to the *MLL* negative patients who showed 18.1% (51/282) favourable, 67.4% (190/282) intermediate and 14.5% (41/282) adverse types (Table 3.7A). Despite the slightly higher proportion of intermediate classifications in the *MLL* PTD patients, there was no significant difference between the distribution of the *MLL* PTD and *MLL* WT patients in the cytogenetic prognostic groups (p=0.3). To determine if there was an

association between *MLL* PTD and cytogenetically normal patients, the intermediate group was further stratified into cytogenetically normal and those with a cytogenetic abnormality. Although over half of the duplication positive patients (9/15; 60.0%) were in the cytogenetically normal group, there was no significant difference between the duplication positive and negative patients (p=0.6).

A.

	MLL WT *n=282*	%	*MLL* PTD *n=15*	%	p value
Favourable	51	18.1	1	6.7	
Intermediate	190	67.4	11	73.3	0.3
Adverse	41	14.5	3	20.0	

B.

	MLL WT *n=282*	%	*MLL* PTD *n=15*	%	p value
Favourable	51	18.1	1	6.7	
Intermediate (normal)	143	50.7	9	60.0	0.6
Intermediate (abnormal)	47	16.7	2	13.3	
Adverse	41	14.5	3	20.0	

Table 3.7 Frequency of *MLL* PTD in Cytogenetic Prognostic Classifications
A: Prevalence of the *MLL* PTD in patients classified by cytogenetic prognostic classifications. B: Prevalence of *MLL* PTD when the intermediate prognosis group is sub-divided into those with normal karyotype ('normal') or with cytogenetic abnormalities ('abnormal'). WT: wild type; PTD: partial tandem duplication.

These data show that the proportion of patients classified with an intermediate prognosis due to the lack of a marker would be reduced by the inclusion of the *MLL* PTD assay. There were 201 patients classified as intermediate prognosis by using cytogenetic classification of which 75.1% (151/201) were cytogenetically normal. The *MLL* PTD assay increased the number of patients with a significant prognostic marker (Section 3.3.3) through detection of nine duplications (9/151; 6.0%) in the cytogenetically normal cohort. Therefore, the *MLL* PTD assay increased the number of patients with a prognostic marker when compared to cytogenetic testing alone. Addition of the *FLT3* ITD to the 146 patients tested for both mutations added a prognostic marker to 39 more patients 32.8% (48/146).

3.3.2.f   *MLL PTD Incidence in Relation to RT-PCR of Fusion Genes*

The incidence of *MLL* PTD in the diagnostic AML screen cohort was analysed to determine how the inclusion of the *MLL* PTD assay would affect the number of non-informative results in the RT PCR AML screen assay. There were 231 diagnostic AML patients screened by RT PCR and only 18.2% (42/231) had a detectable mutation. Fourteen of the mutations were *CBFβ-MYH11* (14/231; 6.1%), 12 *RUNX1-RUNX1T1* (12/231; 5.2%), 14 *PML-RAR* (14/231; 6.1%) and 2 *BCR-ABL* (2/231; 0.9%). This left 82% (189/231) of the patients without a marker detectable by these RT PCR tests. The *MLL* PTD was detected in 10 (10/231; 4.3%) of the samples, with only one co-occurring with a detectable RT PCR translocation, a *RUNX1-RUNX1T1*. Therefore, the *MLL* PTD assay increased the number of informative RT PCR results by 21.4% (9/42). Inclusion of the *FLT3* ITD assay, available on 186 of the 189 negative samples, detected 35 positives and increased the positive results by 68.6% (35/51) when combined with the results of the *MLL* PTD assay. The final proportion of informative markers detected when both tests were performed was 37.7% (86/228). The addition of the *MLL* PTD test whether independently or in conjunction with the *FLT3* ITD

assay increased the number of informative results obtained from RT PCR testing of *de novo* AML patients.

3.3.2.g *Association of Clinical Features with MLL PTD*

The demographic data on age, sex, FAB classification, presentation WBC and presentation blast percentage for the 342 diagnostic AML patients was analysed to determine if there were associations with *MLL* PTD.

In the only instance of a difference between trial and non-trial patients, elderly trial patients had a significantly higher incidence of the mutation (p=0.03). The median age of the *MLL* PTD negative patients was 53.5 years (range 15-87) while the median for the positive patients was 62 years (range 35-76). The addition of non-trial patient's data raised the number of patients from 225 to 342 and increased the median of the *MLL* WT to 57 (range 1-89) and the *MLL* PTD patients to 64 (range 35-82) (Table 3.8). These data still demonstrated a higher frequency of *MLL* PTD in elderly patients, but were not significant (p=0.07).

Age (years)	MLL WT n=325	%	MLL PTD n=17	%	p value
0-14	11	3.4	0	0.0	
15-29	29	8.9	0	0.0	
30-39	24	7.4	1	5.9	
40-49	37	11.4	3	17.6	
50-59	80	24.6	0	0.0	
60+	144	44.3	13	76.5	
Median	57 (1-89)		64 (35-82)		0.07

Table 3.8 Distribution of *MLL* PTD by Age
Distribution of the *MLL* PTD across age groups in the combined trial and non-trial cohort. WT: wild type; PTD: partial tandem duplication.

The greater proportion of *MLL* PTD were female 58.8% (10/17), and the occurrence of the mutation in females (10/155; 6.5%) was higher than in males (7/187; 3.7%), however, these differences were not significant (p=0.3).

The distribution of the FAB classifications for the 325 *MLL* WT and 17 *MLL* PTD patients are displayed in Table 3.9. Although the majority of the duplications occurred in M1 and M2, the distribution broadly mirrored that of the *MLL* WT and no statistical differences were seen (p=0.4).

FAB	*MLL* WT *n=325*	%	*MLL* PTD *n=17*	%	p value
	n		n		
M0	28	8.6	1	5.9	
M1	50	15.4	6	35.3	
M2	80	24.6	6	35.3	
M3	19	5.8	0	0.0	
M4	46	14.2	0	0.0	0.4
M5	32	9.8	1	5.9	
M6	5	1.5	1	5.9	
M7	1	0.3	0	0.0	
Bilineage	1	0.3	0	0.0	
RAEB/RAEB-T	6	1.8	0	0.0	
N/D	57	17.5	2	11.8	

Table 3.9 Incidence of *MLL* PTD in FAB Classifications
Incidence of FAB subtypes in patients with and without the *MLL* PTD. RAEB: refractory anaemia with excess blasts; RAEB-t: refractory anaemia with excess blasts in transformation; N/D: not determined.

To determine if there was an association between the diagnostic WBC and *MLL* PTD, the data from 323 patients was analysed. The median WBC for *MLL* negative patients was 16.2×10^9/L (range $0.6\text{-}311.0 \times 10^9$/L) while the median for the duplication positive patients was 4.4×10^9/L (range $0.8\text{-}125.0 \times 10^9$/L) (Table 3.10). There was no significant difference between the distribution of WBC for positive or negative *MLL* PTD patients (p=0.17). Additionally, the *FLT3* ITD is associated with a significant increase in the WBC of *MLL* PTD patients (p=0.02).

WBC ($\times 10^9$/L)	*MLL* WT $n=306$	%	*MLL* PTD $n=17$	%	p value
0-9.9	125	40.8	10	58.8	
10-49.9	95	31.0	1	5.9	
50-99.9	42	13.7	4	23.5	
100+	44	14.4	2	11.8	
Median (range)	16.2 (0.6-316.0)		4.4 (0.8-125.0)		0.17

Table 3.10 Association of *MLL* PTD with Presenting WBC
Comparison of the distribution of *MLL* WT and *MLL PTD* in presenting WBC. L: litre; WT: wild type; PTD: partial tandem duplication; WBC: white blood count.

The proportion of blasts in the diagnostic analysis from 299 samples were generally similar between wild type and positive patients with the largest number of patients found in the >50% subset for both, 56.6% (184/325) for wild type and 82.4% (14/17) for the duplication (Table 3.11). The median and range were 65% (1-100%) in *MLL* WT versus 66% (range 10-92%) in the *MLL* PTD. There was no significant difference between the distributions of the two groups (p=0.7).

Blast Percentage	*MLL* WT n=325	%	*MLL* PTD n=17	%	p value
<20%	21	6.5	2	11.8	
20-29%	26	8.0	0	0.0	
30-49%	52	16.0	0	0.0	
50%+	184	56.6	14	82.4	
Median (range)	65 (1-100)		66 (10-92)		0.7
Unknown	42	12.9	1	5.9	

Table 3.11 Association of *MLL* PTD with Diagnostic Blast Percentage
WT: wild type; PTD: partial tandem duplication

In summary, the *MLL* PTD patient cohort was not significantly different to the *MLL* WT cohort in terms of the baseline clinical characteristics of sex, FAB classification, WBC or blast percentage. However, evidence from this study suggested a higher incidence of *MLL* PTD in the elderly.

3.3.2.h   *Correlation of MLL PTD with WHO Performance Status*

Patients with the *MLL* PTD were studied to determine if there was an association between the duplication and the WHO performance status. The WHO performance status is a measure of the patients well being and is used as a factor in determining patient treatment (Oken *et al.* 1982). The status levels range from 0-5, with 0 being asymptomatic and 5 death. The distribution of patients in status level 0 to status level 4 (bed-bound or completely disabled) was compared in 225 trial patients with and without the duplication (Table 3.12). Of the *MLL* negative patients, 0.9% (2/214) were at status level 4, 2.3% (5/214) at level 3, 2.3% (5/214) at level 2, 27.6% (59/214) at level 1 and 66.8% (143/214) at level 0. The *MLL* PTD distribution showed none (0/11) at level 4, 9.1% (1/11) at level 3, none (0/11) at level 2, 54.5% (6/11) at level 1 and 36.4% (4/11) at level 0 (Table 3.12). Thus, the numbers of *MLL* negative patients increased as patient well-being increased from level 4 to level 0, whereas the greatest number of *MLL* PTD patients were in performance level 1. However, these data were not significant (p=0.09).

WHO Performance Status	*MLL* WT n=214	%	*MLL* PTD n=11	%	p value
0	143	66.8	4	36.4	
1	59	27.6	6	54.5	
2	5	2.3	0	0.0	0.09
3	5	2.3	1	9.1	
4	2	0.9	0	0.0	

Table 3.12 Association of *MLL* PTD with WHO Performance Status
WT: wild type; PTD: partial tandem duplication; WHO: World Health Organisation; PTD: partial tandem duplication.

3.3.2.i   Association of MLL PTD With sAML

Within the total cohort of 342, the patient data were studied for an association between *MLL* PTD and sAML. The *MLL* negative patients exhibited a higher level of sAML than the positive cases, with 11% (35/325) compared to 6% (1/17) with the duplication. This difference was not significant (p=1.0).

3.3.3   *MLL* PTD Patient Outcome

3.3.3.a   Association of MLL PTD with Complete Remission

The MRC trial data from 220 intensively treated patients was used to determine if there was an association between the *MLL* PTD and one of three outcomes: induction death, resistant disease or complete remission. A total of 31 patients negative for *MLL* PTD died during induction (31/209; 14.1%) compared to no deaths for *MLL* PTD patients (0/11; 0%). Although fewer *MLL* PTD patients died during the induction stage of treatment the difference in death rate was not significant (Odds Ratio (OR), 0.30 [0.05-1.68]; p=0.17) (Table 3.13). The incidence of RD did not vary significantly between the positive and negative patients, 10.5% (22/209) of negative patients had RD compared with 18.1% (2/11) of *MLL* PTD patients (OR, 2.19 [0.31-15.2]; p=0.4). The higher proportion of duplication positive patients reaching CR, 81.8% (9/11) versus 75.0% (156/209) *MLL* WT, was not significant (OR, 0.68 [0.17-2.76]; p=0.6). Additionally, the treatment outcome of 103 normal karyotype patients demonstrated no difference in these three remission indicators for *MLL* PTD patients compared with *MLL* WT. In summary, there was no association between remission indicators and the presence of the *MLL* PTD.

	MLL WT n=209	%	MLL PTD n=11	%	OR	95% CI	p value
ID							
-	178	85.2	11	100.0	0.3	0.05-1.68	0.17
+	31	14.8	0	0.0			
RD							
-	187	89.5	9	81.8	2.19	0.31-15.2	0.4
+	22	10.5	2	18.2			
CR							
-	53	25.4	2	18.2	0.68	0.17-2.76	0.6
+	156	74.6	9	81.8			

Table 3.13 Comparison of Complete Remission Outcome
Complete remission outcome compared between patients with and without *MLL* PTD. WT: wild type; PTD: partial tandem duplication; OR: odds ratio; CI: confidence interval; ID: induction death RD: resistant disease; CR: complete remission; -: negative for mutation; and +: positive for mutation.

3.3.3.b Outcome of MLL PTD Patients

The outcome indicators, relapse free survival (RFS), risk of relapse (RR), and overall survival (OS), were compared between patients positive and negative for the duplication using the post diagnostic MRC trial data of AML patients. The presence of the duplication had a significant negative influence on the five year RFS rate in the 165 patients with available data, the Kaplan Meier (KM) survival estimates were 38.4% for the *MLL* PTD negative patients versus 0% for the *MLL* PTD positive (Hazard Ratio (HR), 18.78 [5.11-69.09]; p<0.0001). In fact, only one patient with the duplication passed the one-year mark before relapsing (Figure 3.5). The results remained significant (HR, 2.66 [1.19-5.91]; p=0.01) after RFS was adjusted for the influence of the baseline covariates (age, sex, secondary disease, cytogenetic prognosis status and WHO performance status). However, when adjusted for the molecular markers *FLT3* ITD, TKD and *NPM1* the influence of *MLL* PTD on RFS was no longer significant (HR, 1.72 [0.67-4.41]; p=0.3).

A   MRC AML Trials: Relapse-free Survival

	No. Patients	No. Events
MLL–PTD WT	156	96
MLL–PTD Mutant	9	9

2P < 0·0001

MLL–PTD WT
MLL–PTD Mutant

37%

Adjusted for baseline covariates: HR 2.66 (1.19-5.91) p=0.01.
Adjusted for baseline, molecular: HR1.72 (0.67-4.41) p=0.3

B   MRC AML Trials: Relapse Risk

	No. Patients	No. Events
MLL–PTD WT	156	73
MLL–PTD Mutant	9	8

2P < 0·00001

MLL–PTD WT
MLL–PTD Mutant

100%

54%

Adjusted for baseline: HR 3.95 (1.72-9.10) p=0.0005
Adjusted for baseline and molecular: HR 2.50 (0.93-6.72) p=0.06

C   MRC AML Trials: Overall Survival

	No. Patients	No. Events
MLL–PTD WT	209	138
MLL–PTD Mutant	11	10

2P = 0·04

MLL–PTD WT
MLL–PTD Mutant

Adjusted for baseline covariates: HR 1.23 (0.58-2.59) p=0.6
Adjusted for baseline and molecular: HR 0.99 (0.42-2.30) p=1.0

33%

9%

Figure 3.5 *MLL* PTD influence on Clinical Outcome in diagnostic AML.
A. Relapse free survival, B. risk of relapse, and C. overall survival of *MLL* PTD WT compared to *MLL* PTD patients. HR: hazard ratio.

Patients positive for the duplication demonstrated a significantly higher RR than negative patients (n=156). The RR in *MLL* WT patients at five years post-diagnosis was 46.7% compared to 100% of the eight *MLL* PTD patients (HR, 75.19 [15.12-373.9]; p<0.00001). One patient reached CR, yet died without relapsing; therefore this patient was not censored. The significance of the relationship between *MLL* PTD and a higher RR was maintained after the adjustment was made for baseline co-variates (p=0.005, HR=3.95 (17.2-9.10)) and remained highly suggestive after the influence of the molecular markers was included (p=0.06, HR=2.5 (0.93-6.72)). The presence of *MLL* PTD was correlated with the OS at 5 years post diagnosis in 220 patients. Patients with the duplication had a significantly lower rate of OS than *MLL* negative patients, with survival estimates of 9.1% and 33.9% respectively (p=0.04, HR=2.5 (1.05-5.97)). The *MLL* PTD patient that survived to five years post diagnosis had a favourable t(8;21) marker in addition to the duplication, which may have been dominant over the negative influence of the PTD. The negative relationship between *MLL* PTD and OS was not retained after baseline co-variate adjustments were made (p=0.6, HR=1.23 (0.58-2.59)), or with the additional adjustment that accounted for the impact of molecular markers (p=1.0, HR=0.99 (0.42-2.3)). In summary, the *MLL* PTD does not appear to add any prognostic information relating to the OS beyond the scope of the standard baseline prognostic variables (age, sex, secondary disease, cytogenetic prognosis classification and WHO performance status). In contrast, patients positive for the duplication fared significantly worse in both RFS and RR even after the baseline variables were included, thus demonstrating the negative impact of the *MLL* PTD on these outcomes.

The clinical outcome of CN-AML patients with and without the *MLL* PTD was investigated due to the high proportion of duplications that occur in this cytogenetic subgroup. Patients with an *MLL* PTD detected had a significantly lower level of RFS and a significantly higher RR (p=0.002, HR=15.4 (2.79-84.4) and p<0.0001, HR=145 (16.9-1242) respectively).

There were 82 patients with relapse data, 36.4% of the *MLL* negative patients achieved RFS at five years compared with 0% of the *MLL* PTD patients (Figure 3.6). When comparing RR analyses, 48.1% of *MLL* negative patients relapsed compared to all of the patients with a detected PTD. Patients with a detectable *MLL* PTD exhibited a lower level of OS at 5 years, 0% versus 34.7% of the *MLL* negative, but this result was not statistically significant (p=0.1, HR= 2.76 (0.79-9.73)). In summary, the presence of the *MLL* PTD within the cytogenetically normal AML subgroup had a significant negative affect on the relapse outcomes, relapse free survival and risk of relapse, but no significant effect on overall survival.
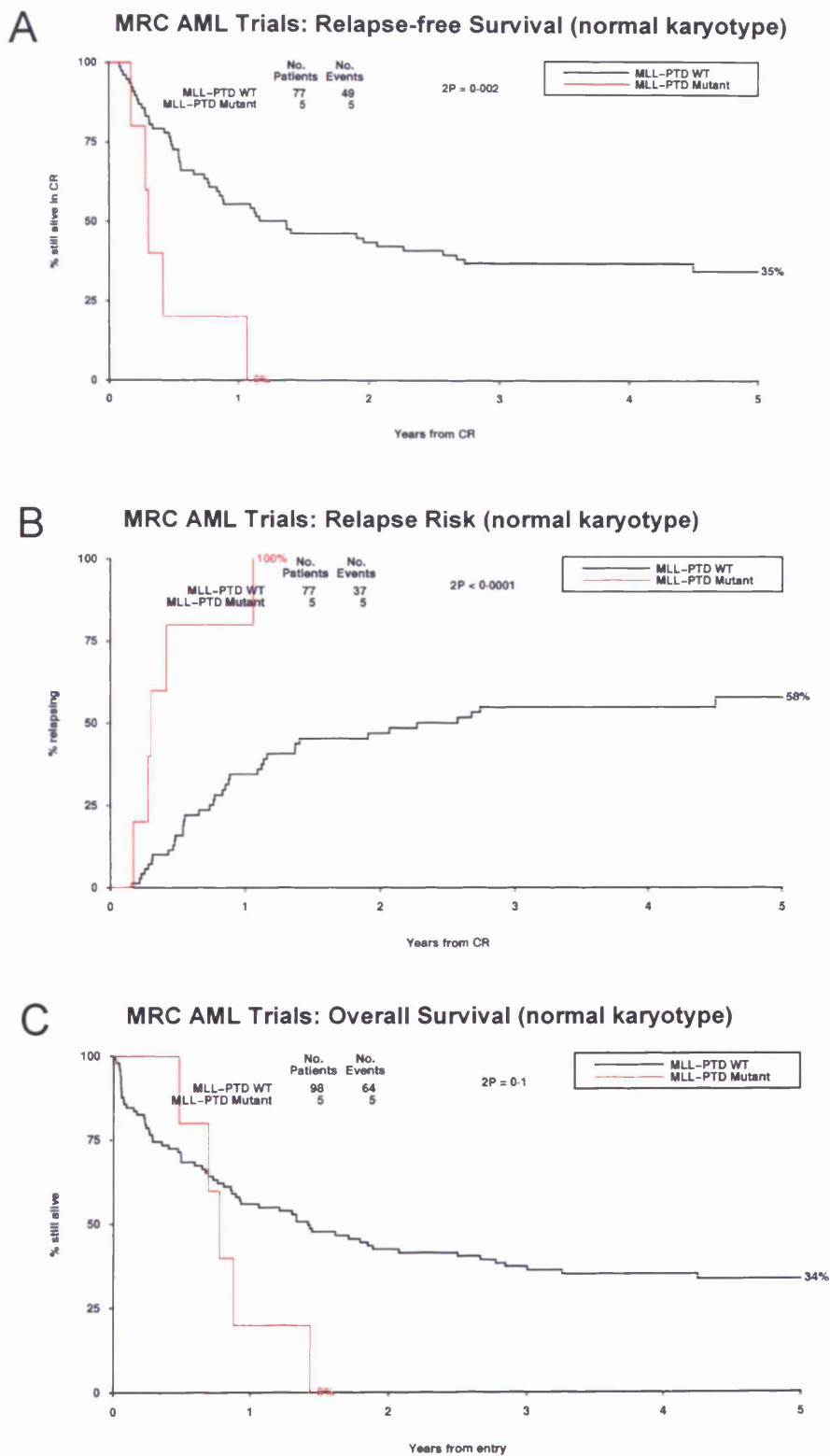
Figure 3.6 *MLL* PTD influence on Clinical Outcome of CN-AML Patients
A. Relapse free survival, B. risk of relapse, and C. overall survival of *MLL* PTD WT compared to *MLL* PTD patients. HR: hazard ratio.

As a large proportion of *MLL* PTD mutations co-occur with *FLT3* mutations, the prognostic impact of the combination was compared by outcome analysis. The patients were divided into four groups, *MLL* PTD-/*FLT3*-, *MLL* PTD-/*FLT3*+, *MLL* PTD+/*FLT3*- and *MLL* PTD+/*FLT3*+ and the three outcome indicators RFS, RR and OS were estimated. The *FLT3* mutations were analysed in three categories, *FLT3* ITD, *FLT3* TKD and *FLT3* overall, where *FLT3* was considered mutant if either ITD or TKD was mutant and WT when both were WT. The low numbers in the co-incident groups were insufficient for statistical analysis of the associations, therefore the study can be considered primarily as a preliminary analysis of the trends which may encourage further investigations using greater numbers. The numbers of patients with mutation status and clinical outcome data are given in the figure legends.

The RFS and RR for the three *FLT3* categories in conjunction with *MLL* PTD described similar scenarios, where *MLL* PTD+ patients fared worse than the others. In RFS, none of the *MLL* PTD+/*FLT3*+ and *MLL* PTD+/*FLT3*- patients were alive at five years, with the double positive patients surviving around half as long as the *MLL* PTD+/*FLT3*- patients (Figure 3.7). The KM plots for RR showed that all *MLL* PTD patients relapsed, with the time to relapse shorter for the double positive group (Figure 3.8). Interestingly, in both analyses, the *MLL* PTD-/*FLT3* TKD+ was beneficial to the patient outcome in comparison to all other categories, while the *MLL* PTD+/*FLT3* TKD+ was markedly detrimental.
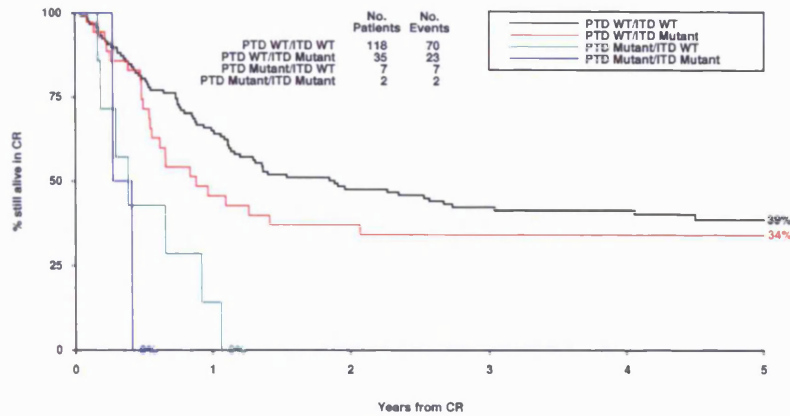
Figure 3.7 Influence of *MLL* PTD and *FLT3* mutations on RFS
Relapse free survival in patients positive or negative for *MLL* PTD with A. *FLT3* ITD mutations, B. *FLT3* TKD mutations or C. either *FLT3* mutation. MRC: Medical Research Council; PTD: partial tandem duplication; ITD: internal tandem duplication, TKD: tyrosine kinase domain.

A MRC AML Trials: Relapse Risk in FLT3 ITD Mutants

	No. Patients	No. Events
PTD WT/ITD WT	118	54
PTD WT/ITD Mutant	35	18
PTD Mutant/ITD WT	7	6
PTD Mutant/ITD Mutant	2	2

Legend:
- PTD WT/ITD WT
- PTD WT/ITD Mutant
- PTD Mutant/ITD WT
- PTD Mutant/ITD Mutant

B MRC AML Trials: Relapse Risk in FLT3 TKD Mutants

	No. Patients	No. Events
PTD WT/TKD WT	111	56
PTD WT/TKD Mutant	16	6
PTD Mutant/TKD WT	6	6
PTD Mutant/TKD Mutant	2	1

Legend:
- PTD WT/TKD WT
- PTD WT/TKD Mutant
- PTD Mutant/TKD WT
- PTD Mutant/TKD Mutant

C MRC AML Trials: Relapse Risk in all FLT3 Mutants

	No. Patients	No. Events
PTD WT/FLT3 WT	80	39
PTD WT/FLT3 Mutant	51	23
PTD Mutant/FLT3 WT	4	4
PTD Mutant/FLT3 Mutant	4	3

Legend:
- PTD WT/FLT3 WT
- PTD WT/FLT3 Mutant
- PTD Mutant/FLT3 WT
- PTD Mutant/FLT3 Mutant

Figure 3.8 Influence of *MLL* PTD and *FLT3* mutations on RR
Relapse risk in patients positive or negative for *MLL* PTD with A. *FLT3* ITD mutations, B. *FLT3* TKD mutations or C. either *FLT3* mutation. MRC: Medical Research Council; PTD: partial tandem duplication; ITD: internal tandem duplication, TKD: tyrosine kinase domain.

The *MLL* PTD appeared to decrease the overall survival more than either of the *FLT3* mutations alone. In fact, the KM graphs for the double negative, double positive and the *MLL* PTD+/*FLT3*- in all three graphs were similar, with the double positive faring the worst (0% survival), *MLL* PTD+/*FLT3*- (ITD, TKD or overall) at 13% or 20% and the double negative ranging from 30%- 37% (Figure 3.9). As shown by data for RFS and RR, the *MLL* PTD-/*FLT3* TKD+ was beneficial to the OS with a KM estimate of 50% in comparison with 24% for *MLL* PTD-/*FLT3* ITD+ and 30% for *MLL* PTD-/*FLT3* + in *FLT3* overall status. Although the numbers are low due to the low frequency in multiple mutation categories, the outcome results revealed a negative influence of *MLL* PTD in all three of the outcome indicators suggestive of a negative influence in combination with *FLT3* mutations.

A   MRC AML Trials: Overall Survival in All FLT3 ITD Mutants

	No. Patients	No. Events
PTD WT/ITD WT	157	98
PTD WT/ITD Mutant	49	37
PTD Mutant/ITD WT	8	7
PTD Mutant/ITD Mutant	3	3

Legend: PTD WT/ITD WT; PTD WT/ITD Mutant; PTD Mutant/ITD WT; PTD Mutant/ITD Mutant

% still alive vs Years from entry; curves end at 37%, 24%, 13%, 0%

B   MRC AML Trials: Overall Survival in FLT3 TKD Mutants

	No. Patients	No. Events
PTD WT/TKD WT	151	104
PTD WT/TKD Mutant	18	9
PTD Mutant/TKD WT	8	7
PTD Mutant/TKD Mutant	2	2

Legend: PTD WT/TKD WT; PTD WT/TKD Mutant; PTD Mutant/TKD WT; PTD Mutant/TKD Mutant

% still alive vs Years from entry; curves end at 50%, 30%, 13%, 0%

C   MRC AML Trials: Overall Survival in All FLT3 Mutants

	No. Patients	No. Events
PTD WT/FLT3 WT	108	69
PTD WT/FLT3 Mutant	67	47
PTD Mutant/FLT3 WT	5	4
PTD Mutant/FLT3 Mutant	5	5

Legend: PTD WT/FLT3 WT; PTD WT/FLT3 Mutant; PTD Mutant/FLT3 WT; PTD Mutant/FLT3 Mutant

% still alive vs Years from entry; curves end at 35%, 30%, 20%, 0%

Figure 3.9 Influence of *MLL* PTD and *FLT3* mutations on OS
Overall survival in patients positive or negative for *MLL* PTD with A. *FLT3* ITD mutations, B. *FLT3* TKD mutations or C. either *FLT3* mutation. MRC: Medical Research Council; PTD: partial tandem duplication; ITD: internal tandem duplication, TKD: tyrosine kinase domain.

3.4 Discussion

The importance of molecular mutations in determining the prognosis and the response to treatment in AML patients was originally demonstrated in relation to cytogenetics (Grimwade *et al.* 1998). More recently, the importance of sub-microscopic mutations has been evaluated in AML, with particular emphasis on the sub-group of CN-AML. In this study, the frequency of the *MLL* PTD was investigated in 342 patients who were either in an MRC AML trial or being tested for AML markers using a standard AML RT PCR screen. The frequency of the *MLL* PTD in AML patients was 5.0% overall, 5.2% in *de novo* AML and 2.8% in sAML. The frequency within CN-AML rose to 6.5% while the 3.9% incidence observed in aberrant karyotypes was reduced to 2.7% when patients with +11 as the sole abnormality were removed, as there is a known association between *MLL* PTD and +11. Interestingly, the association of +11 patients with old age and few long term disease free survivals (Heinonen *et al.* 1998) may be due in part to *MLL* PTD. The frequencies of *de novo*, CN-AML and aberrant karyotypes registered in this study are in line with those reported in other large number studies assessed over the same time period (Steudel *et al.* 2003), (Dohner *et al.* 2002a), (Schnittger *et al.* 2000) (Bacher *et al.* 2005). Higher frequencies (10-21%) have also been reported (Munoz *et al.* 2003; Shiah *et al.* 2002; Yu *et al.* 1996), however most of these are in studies of less than 100 patients and the 21% was for CN-AML patients only. There is also the likelihood that variation in PCR sensitivity could affect the detection frequency, with amplified *MLL* PTD transcripts clearly detected in both healthy normal samples and AML patients not considered positive for *MLL* PTD when the PCR sensitivity is increased (Schnittger *et al.* 1998), (Marcucci *et al.* 1998). The first step in this study was to optimise the PCR protocol to prevent the amplification of the multiple fusion transcripts detectable in all AML patients clearly demonstrated in Figure 3.1. The effectiveness of the

optimisation strategy was confirmed by the negative results of all 20 healthy control samples assayed in this study.

The incidence of the duplication in patients with sAML in this study, 2.8%, is lower than that reported elsewhere. The few early studies in which sAML and tAML were examined reported frequencies of 15.4% (4/26) and 3.8% (1/26) respectively (Christiansen & Pedersen-Bjergaard 2001; Schnittger *et al.* 2000). Apart from possible differences arising from the difference in sample size, the sAML frequency in this study may be reduced as a result of tAML patients being categorised as sAML in the trial studies. The higher frequency of *MLL* PTD in AML compared to MDS (0/11) observed in this study may indicate the duplication is an early mutation event that promotes the pathogenesis of AML over that of MDS. Furthermore, the reported increase in frequency as disease evolves from MDS 2.7% (9/238) to sAML 7.4% (28/378) suggests the *MLL* PTD may be an acquired event promoting the transformation of MDS to AML (Bacher *et al.* 2007a). A recent publication supports this idea with the same incidence of *MLL* PTD observed in +11 MDS as in +11 AML, resulting in a high rate of transformation to AML and poor prognosis (Wang *et al.* 2010).

The *MLL* PTD was detected in 11.1% (2/18) of the relapse patient samples. The e10+e11/e3 transcript was detected in both the diagnosis and relapse samples of the one patient where samples at both timepoints were available. This indicated that the marker was stable over the time course of the disease. The gain of cytogenetic aberrations not present at diagnosis, in addition to the marker stability, suggests the *MLL* PTD may be an early event leading to leukaemogenesis. Stability of a mutation at relapse would suggest potential as an MRD marker, but there are difficulties due to the levels of transcripts present in healthy volunteers and non *MLL* PTD AML patients (Schnittger *et al.* 1998). Further data and discussion of the stability of the duplication, MRD applicability and clonal evolution are presented in Chapter 4.

Analysis of RT PCR AML screens and cytogenetic prognostic classification cohorts revealed the usefulness of the *MLL* PTD as a diagnostic and prognostic marker. Performance of the test alongside the traditional RT PCR AML screen, including *RUNX1-RUNX1T1*, *PML-RAR* and *CBFB-MYH11* increased the number of informative results by 21.4%, as 6% of CN-AML patients gained a marker. The proportion of informative results was raised further when combined with the *FLT3* ITD assay. The use of the *MLL* PTD assay would provide an increased level of prognostic information for AML patients, information that may be used to alter treatment strategies and improve clinical outcome.

Investigation of the prevalence of *MLL* PTD and other molecular mutations revealed a significant difference between the occurrence of *NPM1* in *MLL* WT and patients with the duplication. There were no *NPM1* mutations present in *MLL* PTD positive cases versus 31% of *MLL* WT. This negative association fits with the grouping of mutations into complementary classes. It has been proposed that molecular mutations are divided into two groups in AML: class I mutations, which increase proliferation and survival through the activation of signal transduction pathways, and class II mutations that impair differentiation by disruption of transcription (Gilliland 2002). *FLT3* mutations are class I mutations while *MLL* PTD and *NPM1* are generally considered to be class II, and overlap is generally seen between mutations of different classes (Section 1.2.4.c) (Ishikawa *et al.* 2009; Schlenk *et al.* 2008). The prevalence of the *FLT3* ITD in duplication positive samples (23.4%) was not significant in the diagnostic cohort as a whole or in the CN-AML cohort. Previous results for this observation are varied with some finding significant association and others no significance at all (Olesen *et al.* 2005b; Steudel *et al.* 2003). The *FLT3* TKD was found to co-exist with the duplication in 18.2% of *MLL* PTD positives and high expression of *WT-1* was found in 100% of duplication positives, however neither of these associations were significant. The presence of *MLL* PTD with other aberrations suggests that the duplication

may require the proliferative and survival traits of complementary mutations before leukaemia develops.

Additional comparisons between *MLL* PTD and *FLT3* ITD patients revealed further differences in the clinical characteristics. There was no significant association of *MLL* PTD with the WBC or FAB type. Although there was a lower WBC in patients positive for the duplication, this finding was not significant. These results were in agreement with previous studies, with the exception of a study that found the WBC significantly lower in duplication positive patients (Steudel *et al.* 2003). The duplication has been detected in all FAB types except M3 and M7, which indicates that it is not restricted to lineage or level of differentiation (Caligiuri *et al.* 1998; Dohner *et al.* 2002b; Schnittger *et al.* 2000; Steudel *et al.* 2003). Taken together, the lower WBC, difference in FAB type distribution and association with different cytogenetic abnormalities demonstrated different clinical characteristics for the *MLL* PTD and *FLT3* ITD. In this study a significantly higher WBC in co-duplicated patients was observed compared to *MLL* PTD alone. A similar result has been reported and may indicate a dominant effect of *FLT3* ITD on the proliferative capacity of the cells (Steudel *et al.* 2003). These differences suggest that the two mutations identify different AML subtypes with similarly poor prognostic outcomes. Despite the suggestion that the breakpoint regions of the two genes may be susceptible to a similar action of genotoxic stress (Libura *et al.* 2003), the association of each mutation with a specific cytogenetic abnormality would indicate a common mechanistic cause was unlikely (Steudel *et al.* 2003)).

These data show that neither sex nor blast percentage could be used as an indicator for *MLL* PTD, in accordance with previous studies (Schnittger *et al.* 2000), (Caligiuri *et al.* 1998; Dohner *et al.* 2002b). *MLL* duplication positive patients tended to be older than the negative patients (p=0.07) and the WHO performance status was poorer than negative patients, with a greater proportion of patients in status level one (p=0.09). Taken as prognostic indicators,

both of these factors would suggest a poorer outcome. A significant association with higher age was also observed in previous studies, this would suggest a different mechanism of pathogenesis to that of balanced translocations, which were evenly distributed over all age groups (Bacher *et al.* 2005; Schnittger *et al.* 2000; Shih *et al.* 2006a)).

The *MLL* PTD cohort achieved a slightly higher level of complete remission in accord with other reports, but no significant association between *MLL* PTD and CR was demonstrated (Dohner *et al.* 2002b; Whitman *et al.* 2007). In addition, there was no significant difference between the incidence of the ID or RD rates between positive and negative patients. This would suggest that the poor outcome of these patients is not caused by a failure to reach remission, however, morphological determination of remission may not be sufficient to detect small populations of malignant clones capable of causing patient relapse by repopulating the bone marrow.

The clinical outcome for patients with the *MLL* PTD was worse in the three clinical outcomes studied. The 100% RR and RFS in duplication positive patients were clearly significant (p=0.001 and 0.006 respectively) as well as the 9% OS (p=0.04). Previous investigations of the clinical outcome of duplication positive patients have shown varied results, though the consensus is an association with a generally worse overall outcome. Reduced complete remission duration without reduced overall survival, reduced disease free survival without reduced overall survival and reduced overall survival alone have all been reported in cohorts of unselected AML (Dohner *et al.* 2002b; Schnittger *et al.* 2000; Steudel *et al.* 2003). The differences in the reported outcome may be due to the low frequency of the mutation, although the sample sizes in the studies were all more than 387 patients. Consideration of other prognostic variables within the cohorts, in addition to *MLL* PTD, may be another reason for the discrepancies. For this reason, the clinical outcomes in the present study were adjusted to account for the impact of baseline clinical prognostic factors and

molecular markers. Multivariable regression analysis adjusts for the influence of known prognostic variables on the population of patients being studied, thereby preventing over interpretation of the impact of the condition being studied. Adjustment for the basic factors age, sex, secondary disease, cytogenetic prognostic group and WHO performance status was done first, followed by adjustment for *FLT3* ITD, *FLT3* TKD and *NPM1*. After adjustment for other prognostic variables, the difference in OS between patients positive and negative for the duplication was not significant, which implied that the original difference was not due to *MLL* PTD. In contrast, while the numbers of positives are small, the analysis does suggest that the negative outcomes observed in RR and RFS reflect the influence of the duplication after the basic clinical prognostic factors are considered. Furthermore, the negative impact of the *MLL* PTD is maintained in the RR after adjusting for molecular markers as well as basic clinical prognostic factors, indicating it is an independent prognostic marker and provides a rationale for *MLL* PTD testing at diagnosis.

One of the objectives of this study was to determine the effect of the *MLL* PTD on the CN-AML cohort of patients. Although the number of positive patients is decreased, the results are similar to those seen in the group overall, with a significant negative impact on RR and RFS and no effect in OS. Adjustments were not made to these analyses due to the smaller starting numbers. In summary, the *MLL* PTD proved to be a useful independent prognostic marker that identified a subset of patients with a higher RR and lower RFS in CN-AML and AML with chromosomal aberrations.

In experimental models, the influence of *MLL* PTD in leukaemogenesis is unclear. *In vivo* studies using *MLL* PTD/WT mice show an increased number of haematopoietic progenitor cells and epigenetic alterations in the promoter region of *Hoxa* genes, however the mice never develop leukaemia (Dorrance *et al.* 2006). The duplication was detected in CD34[+] progenitor cells, as well as myeloid and lymphoid subfractions of healthy cord blood

samples which indicated it was present throughout life and occurred at an early stage of haematopoiesis (Basecke *et al.* 2006b). The high incidence of the mutation in a subfraction of cells prone to leukaemic transformation together with the low frequency of AML patients with the mutation suggests that it is not an initiating event. These findings are similar to the reports of *BCR-ABL* (CML) and *RUNX1-RUNXT1* (AML) transcripts being identified in healthy individuals, however both were found at lower frequencies than *MLL* PTD (Basecke *et al.* 2002; Bose *et al.* 1998). The *MLL* PTD transcripts detected in AML patients are restricted to three transcript types and are 3-4 magnitudes higher than in healthy donors, which may be relevant to the development of leukaemia (Chapter 4 data)(Basecke *et al.* 2006b; Weisser *et al.* 2005b). Taken with the *in vivo* study, this may indicate that further alterations in addition to transcriptional up-regulation are necessary before leukaemogenesis occurs. In addition, the argument for the leukaemogenic potential of *MLL* PTD due to its appearance as a singular aberration in early studies should be revised with the identification of more co-occurring mutations (Basecke *et al.* 2006c). The current evidence would suggest the duplication may be an early event in leukaemic transformation that requires the co-operation of an additional mutation(s). Evidence that the co-expression of *FLT3* ITD is necessary for the fusion gene *MLL-SEPT6* to induce myeloid leukaemia in mice supports the hypothesis that multiple mutations may be needed to unlock the leukaemogenic potential of *MLL* mutations (Ono *et al.* 2005).

The clinical outcome studies in *MLL* PTD patients indicated the high proportion of patients achieving complete remission was offset by the increased relapse free survival and risk of relapse resulting in a lower rate of overall survival. The multitude of theories regarding leukaemogenesis provides a number of possible reasons for a relapsing population of cells. The short duration of remission coupled with the relatively insensitive definition of complete remission as less than 5% blasts in a normocellular BM may indicate that a

substantial sub-population of *MLL* PTD positive cells survived treatment and rapidly repopulated the marrow to cause patient relapse. A subpopulation of quiescent *MLL* PTD cells located in the microenvironment of a BM niche would be unaffected by chemotherapy aimed at killing proliferating cell populations. A change in the niche environment promoting the re-expansion of these cells would result in patient relapse. This population could be intrinsically leukaemic or may be a "seeding" population that possessed the property of clonal expansion but required a co-operating mutation to become leukaemic. Furthermore, relapsing *MLL* PTD patients have a higher frequency of adverse molecular markers than *MLL* PTD patients that achieved long term disease free survival (Whitman *et al.* 2007). Lastly, a subpopulation of the leukaemic cells may have obtained a mutation that conferred drug resistance. Determining the cause of early relapse would be instrumental in the pursuit of treatment targeted for these poor outcome patients. Regardless of the cause, the stability of the *MLL* PTD upon relapse implies it could be used as a more sensitive method of MRD detection. Identification of patients harbouring positive cells despite reaching morphological remission may identify a subset of patients who would benefit from different treatment (e.g. bone marrow transplant). The application of *MLL* PTD RQ-PCR for detection of MRD is studied in Chapter 4. In a recent clinical trial, CN AML *MLL* PTD positive patients exhibited an improved rate of disease free survival rate, 41% (9/24). This improvement may be attributable to a treatment regime that included an autologous PBSCT transplant in CR1 for 82% (18/22) of the duplication positive patients (Whitman *et al.* 2007).

The *MLL* duplication had a negative effect that was observed in all three clinical outcomes with all three *FLT3* conditions. The numbers preclude statistical analysis, but the co-duplicated and the *MLL* PTD+/*FLT3*- patients exhibited a more negative outcome than the *MLL* PTD-/*FLT3*+ and *MLL* PTD-/*FLT3*- in each of the three outcomes studied. The results of previous studies regarding the co-duplication of *FLT3* ITD and *MLL* PTD reported no

significant difference in OS, however, one group reported a lower OS for co-duplicated patients and *MLL* PTD+/*FLT3*- that was not significant and another showed significant difference between the co-duplication and all other combinations in DFS for patients over 60 years of age. All of these studies were limited by low numbers of *MLL* PTD and co-duplicated samples (Olesen *et al.* 2005b; Steudel *et al.* 2003)]. The prognostic impact of *FLT3* TKD remains uncertain, with previous studies reporting significantly improved survival, poorer outcome or no impact at all (Bacher *et al.* 2007b; Mead *et al.* 2007; Yanada *et al.* 2005). A likely cause of these different conclusions may be due to the selection of the cohorts analysed. Recently, the prognostic impact of the *FLT3* TKD was shown to be dependent upon its combination with other molecular markers (Bacher *et al.* 2007a). In regard to *MLL* PTD, the study reported a significant negative effect of the co-occurrence of the two mutations when compared to the double negative and the *MLL* PTD-/*FLT3* TKD+ subset but not when compared to the *MLL* PTD+/*FLT3* TKD- patients. This effect was amplified when *FLT3* ITD patients were excluded. A recent report indicated that *FLT3* ITD and TKD should be considered as two independent mutations occurring in the same gene due to the biological and prognostic differences (Bacher *et al.* 2007a). However, the outcome of patients harbouring co-occurring *FLT3* and *MLL* PTD mutations in this study appears to be similarly poor. Clinical outcome is associated with the combined mutational status (e.g. *NPM1, FLT3, CEBPA* and *MLL*) of CN-AML patients (Schlenk *et al.* 2008). Several groups have proposed schema for mutational testing of AML and CN-AML to refine prognosis and assign risk-adapted post-remission therapies (Mrozek *et al.* 2007; Renneville *et al.* 2008).

The low frequency of the *MLL* PTD in AML patients makes the limited numbers of available positives an inherent problem in this and other studies of the duplication. The trial samples used in this study were AML 15 samples processed at Cardiff University or by the hospital Department of Haematology and considerable difficulties have been encountered

previously even when trying to obtain sample numbers smaller than those needed by this study.   Given the low incidence of the mutation, ~2000 samples would be needed substantially increase the significance of the results.   Therefore, the results presented in this study regarding the prognostic significance of the duplication should be considered as a proof of principle.

3.5   Conclusion

The data support the hypothesis that the *MLL* PTD is an indicator of poor prognosis. Consideration of additional prognostic variables, including baseline variables and molecular markers, provided evidence that the duplication was an independent negative prognostic indicator for RR. The duplication gave additional information on relapse free survival beyond that obtainable with the baseline variables. Evidence of a higher rate of CR in the *MLL* PTD coupled with a high RR and low RFS would be consistent with a scenario in which the majority of leukaemic cells were killed yet a subpopulation of *MLL* PTD cells survived, possibly through quiescence or drug resistance, to repopulate the bone marrow causing relapse and in most cases, death.

The introduction of this assay would provide a forewarning of poor clinical outcome in around 5% of CN-AML patients who are presently without RT PCR or cytogenetic information. Ideally, other molecular markers would be included, not only for the prognostic information they would provide in isolation, but also because of possible changes in prognosis due to mutations that occur in combination. The most striking possibility would be the change from a favourable/intermediate prognosis in the case of a patient with *FLT3* TKD to an that of adverse prognosis if an *MLL* PTD was also identified.

Difficulties in identifying large numbers of *MLL* PTD positives due to their low frequency dictates that these results should be considered a working hypothesis. Further work should be completed as an integrated molecular study in conjunction with a large scale clinical trial, thus eliminating difficulties arising from differences in molecular techniques, patient selection and patient treatment which are encountered when comparing results from different studies. A large scale clinical trial that included the *MLL* PTD assay could

determine whether changes in the treatment strategy would improve the clinical outcome for

this cohort of patients.

4 Assessment of *MLL* PTD RQ-PCR as a Marker for Minimal Residual Disease

4.1 Introduction

Beyond the initial prognostic categories established by cytogenetic analysis at diagnosis, there are additional criteria affecting the outcome of the AML patient which subdivide the prognoses still farther and one of these is the level of minimal residual disease (MRD) detected as treatment progresses. MRD detected by PCR has been shown to be clinically relevant using gene fusions RUNX1-*RUNX1T1*, PML/RARA and CBFB/*MYH11* to monitor the patient's response to treatment and to predict early relapse (Grimwade *et al.* 2009; Perea *et al.* 2006; Scholl *et al.* 2005). Qualitative PCR can be an effective method for the detection of MRD and is still widely used, however the advent of RQ-PCR allows for a much more detailed picture of the response than simply a positive or negative result; this is especially important since a positive result need not indicate impending relapse (Buonamici *et al.* 2002; Jurlander *et al.* 1996; Lo *et al.* 1999c). The ability to measure precise levels of expression using RQ-PCR has led to the identification of many new prognostic parameters. A number of studies have shown the prognostic significance of decreased expression levels and the attainment of log reduction thresholds (e.g. 2 log reduction) in fusion genes (Buonamici *et al.* 2002) (Stentoft *et al.* 2006) (Leroy *et al.* 2005). RQ-PCR provides the means to study the kinetics of the changes in the levels of disease in patients, which has led to the development of prognostic scoring systems (Schnittger *et al.* 2003). This scoring system found that a decrease in the transcript of the fusion genes, RUNX1/*RUNX1T1*, PML/RARA and CBFB/*MYH11*, between diagnosis and post consolidation could be used to identify a cohort of patients with favourable prognosis (100% EFS) from those with a significantly worse outcome. Conversely, consecutive substantial increases in the level of transcript invariably predicts for patient relapse. Relapsing patients with core binding factor leukaemia, RUNX1/*RUNX1T1* and CBFB/*MYH11*, were shown to have increasing levels of MRD

(Krauter *et al.* 2003; Stentoft *et al.* 2006), often this increase was seen before the relapse was identified by morphology and cytogenetic analysis (Schnittger *et al.* 2003).

It is important to establish new markers of MRD because the most common gene rearrangements used to measure MRD in AML (RUNX1-*RUNX1T1*, PML/RARA and CBFB/*MYH11*) only occur in 21-37% of AML patients (van Dongen *et al.* 1999). Up to 50% of AML patients appear cytogenetically normal, ruling out the use of the "traditional" gene rearrangement markers or cytogenetic analysis for MRD. However there are a number of genes with mutations or changes in gene expression level which are specifically associated with CN AMLs; these have been studied for use as MRD markers and include *FLT3* ITD, *NPM1* and *WT-1* (Barragan *et al.* 2008; Ostergaard *et al.* 2004; Schnittger *et al.* 2004). The *MLL* PTD also occurs in a sizeable proportion of CN patients, shown to be 5% in Chapter 3 and reported to be 5%-10% in other publications, hence it may be a useful MRD marker (Caligiuri *et al.* 1998; Schnittger *et al.* 2000). Furthermore, while many of the previously mentioned fusion gene markers are categorised as favourable prognosis, the *MLL* PTD has been shown to indicate poor prognosis with a higher probability of relapse and a shorter period of complete remission (Chapter 3)(Dohner *et al.* 2002b). Early identification of unresponsive and relapsing patients using molecular MRD status would give the clinician time to make pro-active changes in the treatment strategy; this approach has been used successfully in acute promyelocytic leukaemia (Grimwade *et al.* 2009) .

However, the use of the *MLL* PTD for MRD detection is not straightforward, due to the presence of low background levels of *MLL* PTD transcripts detected in healthy donor PB and BM as well as progenitor subfractions (Basecke *et al.* 2006a; Marcucci *et al.* 1998; Schnittger *et al.* 1998). An effective assay requires the ability to identify the transcripts relevant to the leukaemia at as high a level of sensitivity as possible while avoiding the amplification of transcripts in normal haematopoietic cells; alternatively there should be as

much separation between the two sets of transcript levels as possible. Optimisation of conditions to preferentially amplify the relevant transcripts is critical if clinically relevant results are to be produced while preventing false positives.

The aim of these studies was to test the hypothesis that the expression level of *MLL* PTD determined by RQ-PCR can be used as a marker for MRD, and is able to indicate remission and predict relapse.

The objectives of this study were:

1) to develop and optimise an *MLL* PTD RQ-PCR assay

2) to develop an analysis methodology

3) to determine the stability of the *MLL* PTD mutation

4) to evaluate patient data to determine efficacy of the *MLL* PTD RQ-PCR as an MRD marker.

4.2   Results

4.2.1   *MLL* PTD RQ-PCR Optimisation

An ideal clinical RQ-PCR assay for MRD requires efficient amplification of the target template and clear differentiation of the positive and negative samples. The RQ-PCR assay protocol was optimised to meet these two criteria. The primers used for the quantitative *MLL* PTD RQ-PCR were the same as those used for the qualitative assay described in Chapter 2 and were used at the same concentration (Chapter 2 Material and Methods Section 2.6.2.a). The RQ-PCR assay was performed using the LightCycler Faststart DNA Master SYBER Green I kit from Roche. Syber Green fluoresces in the presence of double stranded DNA (dsDNA) and the signal is proportional to the amount of dsDNA present. The reagents used for the qualitative and quantitative assays are not the same, which necessitated re-optimisation of the conditions for the RQ-PCR assay. The three variables optimised were $MgCl_2$, template concentration and primer annealing temperature. Three patient samples were used in this optimisation process; one sample positive for the e9/e3 *MLL* PTD, one positive for the e10/e3+e11/e3 *MLL* PTD and one AML sample negative for *MLL* PTD. The AML *MLL* WT sample was used in place of a donor sample because of the scarcity and high demand for the donor RNA. The amount of donor RNA used was minimised by delaying the determination of the level of *MLL* PTD expression in donor samples until after the RQ-PCR was fully optimised. According to the manufacturer's recommendations $MgCl_2$ concentration was optimised first, followed by template concentration and then primer annealing temperature. The efficiency of the assay was judged by the crossing point value (Cp), steepness of the curve, fluorescence intensity and melting temperature (Tm) as shown in the $MgCl_2$ concentrations evaluated in Figure 4.1. On the LightCycler the software calculated the Cp using the "second derivative maximum" method. The Cp was defined as the maximum of the

second derivative of the amplification curve, the point of the maximum rate of increase in DNA amplification. The "second derivative maximum" method was chosen for comparisons between runs because it allows the Cp to be calculated independent from user input and initial fluorescence value. The amount of starting template influences Cp, such that the value of the Cp is inversely correlated to the template concentration. The slope of the curve is a measure of the efficiency of the amplification, with steeper curves showing higher efficiency. Fluorescence intensity is an arbitrary measurement of the Syber Green dye fluorescence during the amplification. However, because the reaction of the Syber Green dye is non-specific the amount of fluorescence detected is a result of all dsDNA produced in the reaction. The specificity of the amplification must be determined using the DNA melting temperature (Tm) to ensure that values arising from non-specific amplification such as primer dimers are not evaluated. The Tm of the DNA product was calculated from the negative derivative of fluorescence to determine the temperature at which the amplicon denatures. The temperature was gradually increased to a final temperature of 95°C and the Tm was calculated from the decrease in fluorescence that occurred when the DNA denatured. Every PCR product will have a characteristic Tm, in the same way as it has a characteristic size in base pairs. The Tm is determined by the GC content and the length of the PCR product.

A.
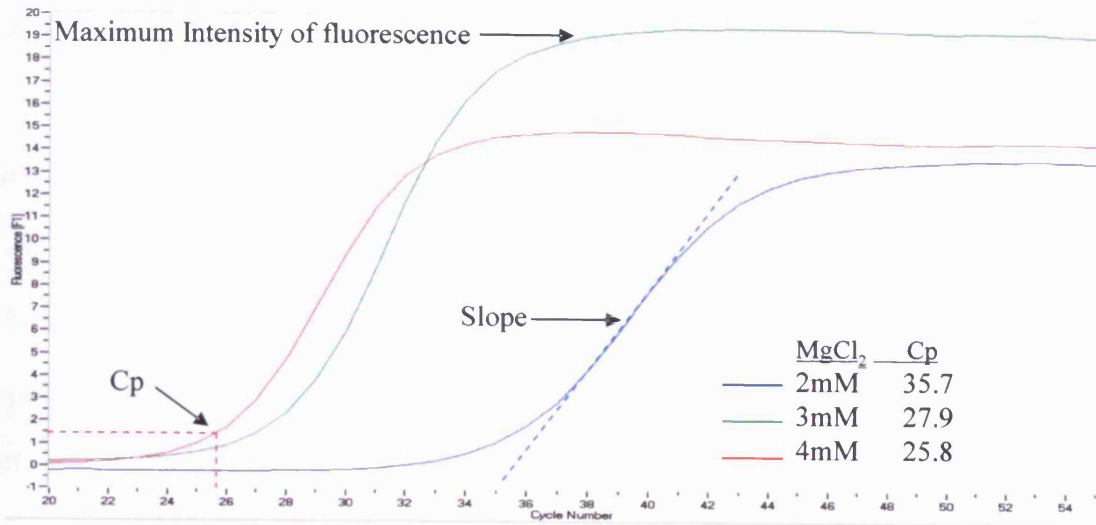
B.
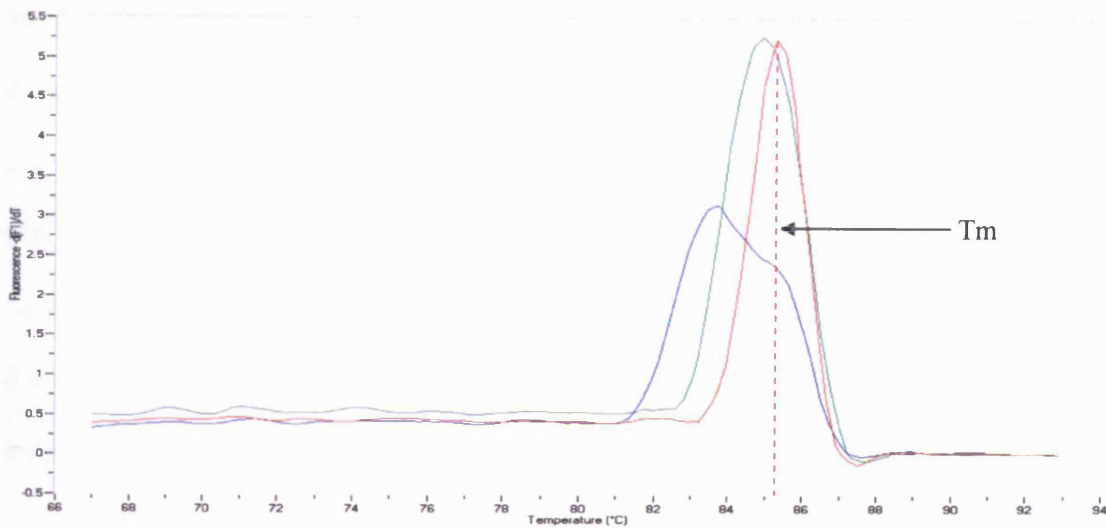
Figure 4.1 Light Cycler Traces of *MLL* PTD e9/e3 Sample
A. Real time Syber Green fluorescent history plotted against the cycle number of *MLL* PTD e9/e3 amplified with varying concentrations of MgCl₂. The three parameters used to assess RQ-PCR quality are shown (Cp, slope and intensity of fluorescence). B. Fluorescence history versus temperature. The fluorescence values were transformed and plotted against temperature to calculate the characteristic melting temperature of the PCR products. Cp: crossing point; Tm: melting temperature; mM: millimolar.

4.2.2   Optimisation of MgCl$_2$ Concentration

The MgCl$_2$ concentration was optimised first because it is a critical factor for primer binding and also affects the activity level of Taq polymerase. Higher concentrations of MgCl$_2$ are known to have an inverse affect on the stringency of primer template binding and as Mg$_2^{2+}$ is a cofactor for Taq polymerase, low concentrations limit the efficiency of the enzyme (Figure 4.1) (Kramer & Coen 2001). The range of MgCl$_2$ concentrations tested for optimisation was 1-5mM. Table 4.1 shows the results of altering MgCl$_2$ concentration on RQ-PCR efficiency, with cDNA template volume and primer annealing temperature held constant at 1μl and 55°C respectively. Amplification was not observed in any of the samples using 1mM, the lowest MgCl$_2$ concentration. The highest MgCl$_2$ concentration, 5mM, showed no improvement in any of the measured variables when compared to 4mM MgCl$_2$. Low Cp values indicated 3-4 mM MgCl$_2$ were improved amplification efficiency compared with 2mM. The specificity of the amplification was confirmed by using agarose gel electrophoresis to calculate the size of the products in base pairs (described in chapter 2). The size of the RQ-PCR products was identical to that amplified using the qualitative assay, which confirmed that the correct products were being produced. The Tm values generated for the two positive samples were within one degree Celsius of each other. Each positive sample had one major Tm indicating a specific amplification with the larger e10/e3+e11/e3 mutation having a slightly higher final temperature. Multiple Tm peaks were calculated for the negative sample indicating that non-specific binding produced more than one product. The intermediate concentrations showed the most efficient amplification.

UPN	MLL PTD Status	MgCl₂ (mM)	Cp	Fluor. Intensity (Arbitrary Units)	Slope	Tm (°C)
49	e9/e3	1	n/a	n/a	n/a	n/a
		2	>41	n/a	++	85.6
		3	25.9	12.0	+++	85.0
		4	25.3	9.0	++	85.3
		5	25.1	8.0	++	85.3
77	e10/e3+e11/e3	1	n/a	n/a	n/a	n/a
		2	29.8	15.0	+++	85.7
		3	31.0	14.0	++	85.7
		4	27.5	9.0	++	86.0
		5	27.6	9.0	++	86.0
250	Negative	1	n/a	n/a	n/a	n/a
		2	35.7	14.0	+++	81-89
		3	35.4	12.0	+++	81-89
		4	33.4	10.0	++	81-89
		5	33.8	8.0	++	81-89
Water		3	n/a	n/a	n/a	n/a

Table 4.1 Optimisation of RQ-PCR Using a Range of MgCl₂ Concentrations.
The *MLL* PTDs e9/e3 and e10+e11/3, and an AML patient negative for the *MLL* PTD were amplified with cDNA template and annealing temperature held constant at 1µl and 55°C respectively. UPN: unique patient number; Cp: crossing point; Tm: melting temperature; Fluor. Intensity: fluorescence intensity. Cp, Tm, fluorescence intensity and slope are measures of amplification efficiency while Tm indicates specificity.

As the highest levels of amplification were apparent at the intermediate concentrations, these assays were repeated to demonstrate reproducibility and to determine one optimum concentration for each mutation (Table 4.2). The high amplification was represented by the low Cp values determined in all (Table 4.2A). Consistent Tm values for the *MLL* PTD positive samples were produced using 3 and 4mM $MgCl_2$ but the Tm value produced in 2mM $MgCl_2$ was 1.5°C lower indicating poor amplification specificity at this lower concentration. Tm values generated for the positive samples were similar, with the larger e10/e3+e11/e3 mutation having a final temperature approximately 1°C higher than the e9/e3 mutation. The *MLL* PTD negative sample showed a similar pattern to the positive samples, with decreasing Cp values as $MgCl_2$ concentrations increased. However, the number of different PCR products increased as $MgCl_2$ concentration increased as indicated by multiple Tm values. The trend of decreasing Cp as $MgCl_2$ concentration increased was reproducible between Table 4.1 and Table 4.2 and can be seen in all samples despite the variation in individual Cps. The results of all three of the parameters of efficiency (Cp, fluorescent intensity and the steepness of the curve) showed 3mM to be the optimum concentration for the e9/e3 *MLL* PTD and 4mM the most efficient for e10/e3+e11/e3 *MLL* PTD. The $MgCl_2$ concentration of 3.5mM was chosen because it provided the lowest Cp and consistent Tm for both positive *MLL* PTD samples (Table 4.2A) and generated the largest difference in Cp values between the positive and the negative samples (Table 4.2B).

UPN	MLL PTD Status	MgCl$_2$ (mM)	Cp	Fluor. Intensity (Arbitrary Units)	Slope	Tm (°C)
A.						
49	e9/e3	2	35.7	14.0	++	83
		3	27.9	19.0	+++	84.9
		4	25.8	15.0	++	85.5
77	e10/e3+e11/e3	2	42.6	17.0	++	84.5
		3	32.5	17.0	+++	86
		4	28.2	16.0	+++	86.5
250	Negative	2	>51	n/a	+++	78-90
		3	36.3	13.0	+++	78-90
		4	34.8	10.0	++	78-90
Water		3	n/a	n/a	n/a	n/a
B.						
49	e9/e3	3.5	24.2	15.0	+++	85.4
77	e10/e3+e11/e3	3.5	26.8	14.5	+++	86.5
250	Negative	3.5	33.7	13.0	+++	78-90

Table 4.2 Reproduciblity of MgCl$_2$ Optimisation
The *MLL* PTDs e9/e3 and e10/e3+e11/e3, and an AML patient negative for the *MLL* PTD were amplified with cDNA template and annealing temperature held constant at 1μl and 55°C respectively. Cp, Tm, fluorescence intensity and slope are measures of amplification efficiency while Tm indicates specificity. A. Repeat of intermediate MgCl$_2$ concentrations to show reproducibility of Table 4.1. B. Optimised assay using 3.5mM MgCl$_2$. Cp: crossing point; Tm: melting temperature; Fluor.:fluorescence intensity.

4.2.3 Optimisation of Template Concentration

The optimisation of template concentration was important for establishing a successful RQ-PCR because high levels of the template cDNA, can inhibit amplification and low levels can reveal evidence of inhibitors in the sample (when a decrease in the amount of template shows an unexpected increase in amplification it can be an indication of a PCR inhibitor being diluted out). An additional reason for testing the template concentration for *MLL* PTDs is the fact that the transcripts were found in normal samples if the sensitivity was too high (Chapter 3). The volumes of template used were 0.01 μl, 1 μl and 2 μls. The lowest volume of 0.01 μl was used to test for evidence of inhibitors and to establish if it would create a greater difference between *MLL* PTD and *MLL* PTD WT Cp values. All three conditions worked in the assay, amplifying *MLL* PTD product; these results are summarised in Table 4.3. The greatest discrimination between the positive and negative Cps, calculated by subtracting the Cp of the negative sample from the Cp of the positive, was seen using 1 μl template, with a resulting difference in Cps of 9.4 for e9/e3 *MLL* PTD and 6.9 for e10/e3+e11/e3 *MLL* PTD. The 0.01 μl sample showed comparable discrimination between the positive and negative samples, and had a greater difference than with the 2 μl sample for the e9/e3 and the e10/e3+e11/e3 *MLL* PTDs. However, the 1 μl sample is more useful because its higher starting concentration of cDNA, as measured by a lower Cp, would enable detection of lower amounts of transcript before the lowest detection level (sensitivity) of the assay was reached. Therefore the optimal template volume was set at 1 μl.

UPN	MLL PTD Status	cDNA Temp. (µl)	Cp	Fluor. Intensity (Arbitrary Units)	Slope	Tm (°C)
49	e9/e3	2	23.6	15	+++	85.4
		1	24.2	15	+++	85.4
		1 (1:10)	29	15	+++	85.4
77	e10/e3+e11/e3	2	26.1	15.5	+++	86.5
		1	26.8	14.5	+++	86.5
		1 (1:10)	30.9	13.5	+++	86.5
250	Negative	2	30.7	10.5	+++	78-90
		1	33.7	13	+++	78-90
		1 (1:10)	37.7	14	+++	78-90
Water		1	46.9	9	++	82

Table 4.3 cDNA Template Volume Optimisation
The *MLL* PTDs e9/e3 and e10/e3+e11/e3, and an AML patient negative for the *MLL* PTD were amplified with MgCl$_2$ concentration and annealing temperature held constant at 3.5mM and 55°C respectively. Cp, Tm, fluorescence intensity and slope are measures of amplification efficiency while Tm indicates specificity. Cp: crossing point; Tm: melting temperature; Fluor: fluorescence intensity.

4.2.4 Optimisation of Primer Annealing Temperature

Whereas the $MgCl_2$ and the template concentration were the first to be optimised because they affect multiple aspects of the PCR, the primer annealing temperature, which only affects the stringency of primer binding, was optimised last and used strictly to maximise the difference in Cp value between the positive and the negative *MLL* PTD samples. A second AML patient sample negative for the *MLL* PTD was used in this optimisation, to take into account for possible variations in the Cp values of negative samples. Raising the annealing temperatures from a starting temperature of 55°C to 69°C gradually increased the difference between the Cp values for the positive and negative *MLL* PTD samples. Table 4.4 shows how increasing the annealing temperatures from 65°C to 69°C increased the difference in Cp between the positive and negative *MLL* PTD samples, while the other measures of efficiency remained largely unchanged. At 69°C, the average difference in Cp was >9 for both the e9/e3 and the e10/e3+e11/e3 mutations (Table 4.5). However, the previously moderate rate of increase in the Cp value observed as the temperature was raised, increased to a >1°C change in Cp as the temperature rose from 67°C to 69°C for the e10/e3+e11/e3 *MLL* PTD. This large increase indicated the beginnings of loss of efficiency for the RQ-PCR. Therefore, the temperature of 69°C was selected as the optimum primer annealing temperature because the separation between the positive and negative samples was nearly equivalent to a 3-log difference and the Cp of the positives was starting to increase indicating that a further increase in temperature would be detrimental.

UPN	MLL PTD Status	Annealing Temp. (°C)	Cp	Fluor. Intensity (Arbitrary Units)	Slope	Tm (°C)
49	e9/e3		23.1	19.5	+++	85.7
77	e10/e3+e11/e3		25	19.5	+++	86.9
250	Negative	65	32.3	17.5	+++	85.7+86.9+90.5
125	Negative		33.4	17.0	++	84.9+87.75+90.5
Water			n/a	n/a	n/a	n/a
49	e9/e3		23.3	17.0	+++	85.7
77	e10/e3+e11/e3		25.6	17.5	+++	86.9
250	Negative	67	34.3	16.0	+++	87.7+88.9+91.0
125	Negative		33.3	14.5	+++	86.4+89.5
Water			n/a	n/a	n/a	n/a
49	e9/e3		24	17.0	+++	85.8
77	e10/e3+e11/e3		26.7	19.5	+++	87.0
250	Negative	69	36.9	14.5	+++	86+89.5
125	Negative		34.8	15.0	+++	86.3+90.5
Water			n/a	n/a	n/a	n/a

Table 4.4 Optimisation of Primer Annealling Temperatures
The *MLL* PTDs e9/e3 and e10/e3+e11/e3, and two AML patients negative for the *MLL* PTD were amplified with MgCl$_2$ concentration held constant at 3.5mM. Cp, Tm, fluorescence intensity and slope are measures of amplification efficiency while Tm indicates specificity. Cp: crossing point; Tm: melting temperature; Fluor.: fluorescence intensity.

		Annealing Temperature					
		65°C	67°C	69°C	65°C	67°C	69°C
UPN		250	250	250	125	125	125
	MLLPTD Status	Negative			Negative		
49	e9/e3	9.7	11	12.9	10.7	10	10.8
77	e10/e3+e11/e3	7.8	8.7	10.2	8.9	7.7	8.1

Table 4.5 Cp Difference Between *MLL* PTD Positive and Negative Samples
Discrimination of RQ-PCR assay between positive and negative samples determined by Cp value at 3 primer annealing temperatures. The Cp difference=Cp(*MLL* Negative)-Cp(*MLL* PTD). Cp: crossing point.

4.2.5 Final Optimisation of the Assay

The results of the final optimisation conditions can be seen in Figure 4.2. The superior reproducibility of the *MLL* PTD Cp and Tm compared to *MLL* WT is shown by the proximity of the duplicate traces. Evidence of the separation between the starting levels of template is shown by the positive samples having lower Cp values (shifted to the left) than the MLL WT samples. The ability to discriminate positive samples at this level, without encountering competing non-specific amplification from WT cDNA would enable testing for MRD to be clinically useful. The level of 3-log reduction from diagnosis was the original gold standard for CML patients being treated with Imatinib, after which there is a 95% chance of disease free progression for the following year (Hughes *et al.* 2003). Detection of MRD to that level would make this assay more sensitive than morphology, cytogenetics and many flow cytometry leukaemia associated immunophenotypes (LAIPs).
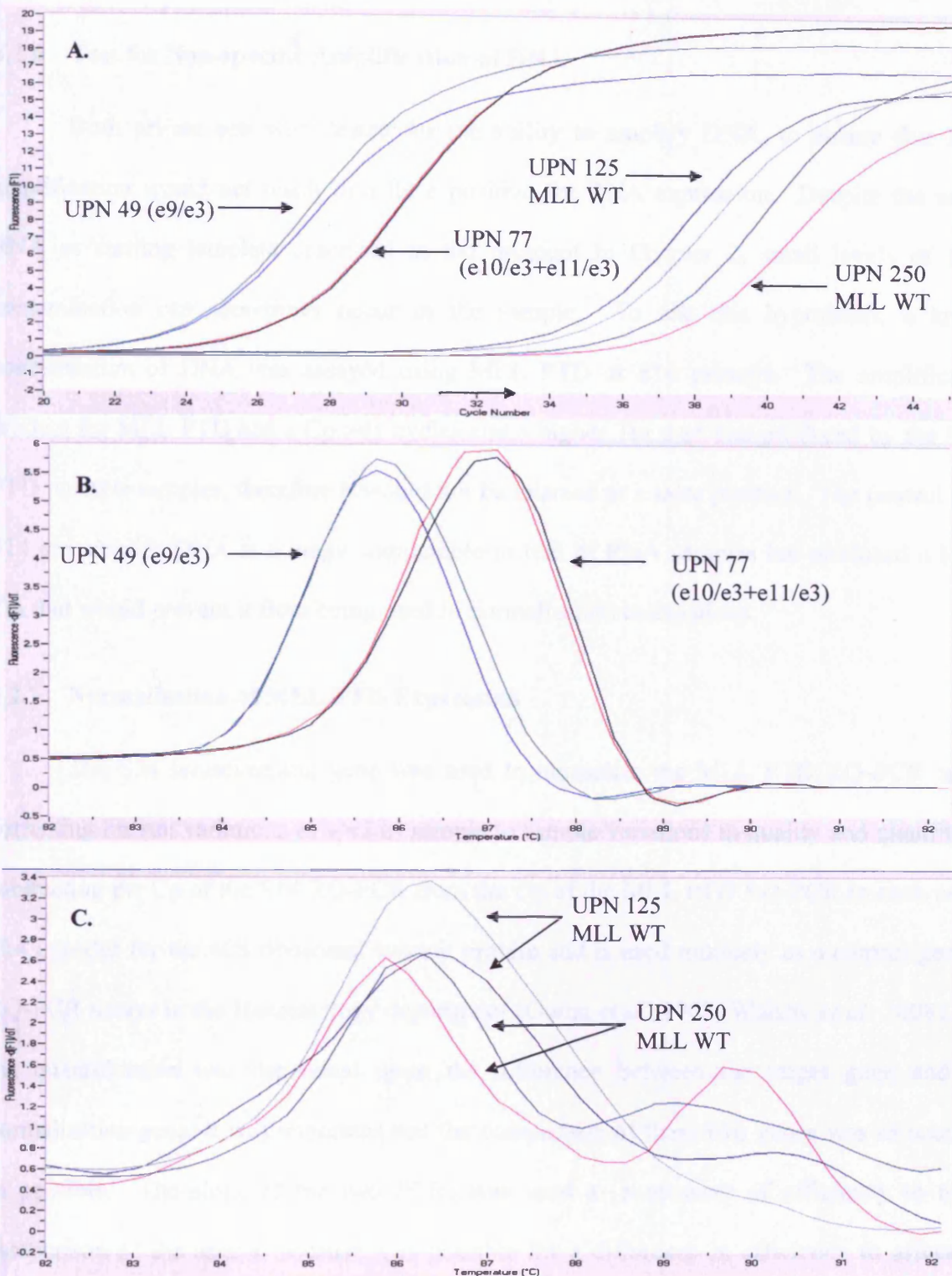
Figure 4.2 RQ-PCR Assays Using Optimised Conditions
A. Real time Syber Green fluorescent history plotted against the cycle number of the MLL
PTDs e9/e3 and e10/e3+e11/e3, and two AML patients negative for the MLL PTD. B. and C.
Fluorescent history versus temperature. The negative derivative of fluorescence over time
was plotted against temperature to calculate the melting temperature of the PCR products. B.
MLL PTD positives. C. MLL PTD negatives.

4.2.6 Test for Non-specific Amplification of DNA

Both primer sets were tested for the ability to amplify DNA, to ensure that DNA amplification would not result in a false positive for RNA expression. Despite the use of RNA as starting template described in the protocol in Chapter 2, small levels of DNA contamination can sometimes occur in the sample. To test this hypothesis, a known concentration of DNA was assayed using MLL PTD or S14 primers. The amplification product for MLL PTD had a Cp >41 cycles and a higher Tm than that produced by the MLL PTD positive samples, therefore it would not be misread as a false positive. The control gene S14 did amplify DNA at a range comparable to that of RNA samples but produced a lower Tm that would prevent it from being used in normalisation calculations.

4.2.7 Normalisation of MLL PTD Expression

The S14 housekeeping gene was used to normalise the MLL PTD RQ-PCR assay, correcting for run variations as well as sample to sample variations in quality and quantity by subtracting the Cp of the S14 RQ-PCR from the Cp of the MLL PTD RQ-PCR in each assay. S14 encodes for the 40S ribosomal subunit protein and is used routinely as a control gene in RQ-PCR assays in the Haematology department (Guinn *et al.* 2005; Walsby *et al.* 2008). As the quantification was dependent upon the difference between the target gene and the normalisation gene, it was important that the comparison of these two genes was as accurate as possible. The slope of the two PCRs was used as a measure of efficiency to allow comparison of the assays, because it is possible for a difference in efficiency to affect the calculation of expression level. PCR efficiency was calculated using the following the formula from Roche Diagnostic technical notes (overpage):

$$E=10^{(-1/slope)} \qquad [4.1]$$

Where E is the efficiency and the slope is calculated from a relative standard curve of
the log concentration of the standard dilutions vs. cycle number.

In an ideal PCR reaction, a 10-fold dilution series would have a slope of –3.3 and an
efficiency of 2, meaning that the amount of PCR product would double after the completion
of every PCR cycle. This means that every 10-fold increase in starting template concentration
would decrease the Cp by a value of 3.3. For the purpose of quantification the slope of a PCR
that fell within the range of -3.0 and –3.9 and had a coefficient of determination (i.e. a
measure of the association between the dilution points and the calculated trend line) of >0.95
should be acceptable (Van, V *et al.* 2003b). The slope of the MLL PTD RQ-PCRs was
calculated by plotting the Cp value against the log of a relative standard curve of the two
MLL PTD samples used in the optimisation assays. cDNA from both positives was serially
diluted 1:10 from 100% to 0.1%. Tm values indicated successful amplification down to 1%,
and the corresponding Cp values were plotted on a scatter graph to calculate the slope (Figure
4.3). The e9/e3 mutation had a slope of –3.4±0.4 (n=3) and a coefficient of determination of
0.99. The slope of the e10/e3+e11/e3 MLL PTD was slightly lower, -3.8±0.5 (n=3), and a
coefficient of determination of 0.99. S14 was retrospectively analysed and showed a
reproducible but lower slope for both samples with a slope of -4.7±0.1 (n=4) and a coefficient
of determination of 0.99. Due to the difference between the slopes of the target gene and the
control gene, the slope of the target gene was used because, by its nature as the unknown
variable, its expression level will display a wider range of Cp values (Beillard *et al.* 2003).
Therefore, despite the different assay efficiencies, the narrow range of Cp values displayed for
S14 as a control gene should have minimal impact on the calculation.

A. **MLL PTD e9/e3 versus S14**

$y = -3.385x + 36.273$
$R^2 = 0.9803$

$y = -4.86x + 31.04$
$R^2 = 0.9922$

Legend:
- ◆ 9/3
- ■ S14
- Linear (S14)
- Linear (9/3)

y-axis: Cp
x-axis: Log of dilution
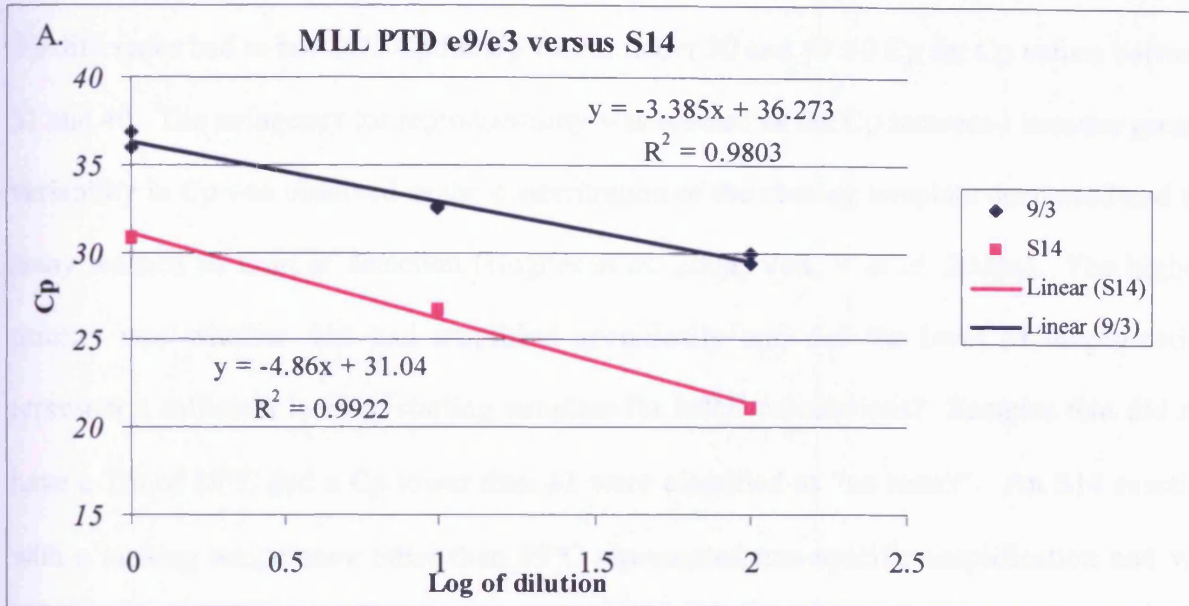
Figure 4.3 RQ-PCR Slopes for Serial Dilutions of MLL PTD Compared to S14
The slope of the expression levels of a serial dilution series from 100-1% for RQ-PCR MLL
PTD transcripts was compared to the slope of a dilution series 100-1% for the control gene
S14. The log value of the dilution was plotted against the Cp value. A. Slope of e9/e3 (blue)
dilutions compared to S14 dilutions (pink). B. Slope of e10/e3+e11/e3 dilutions (blue)
compared to S14 dilutions (pink). Cp: crossing point; R^2: Coefficient of determination.

4.2.8 Analysis Methodology

Steps were taken to verify the specificity of the RQ-PCR amplifications. A protocol

for the analysis of MLL PTD RQ-PCR assays was developed to prevent reporting false

positive due to non-specific amplification of S14 or MLL PTD (Figure 4.4). Due to the non-

specific nature of Syber Green fluorescence, any amplified product, including primer dimer

formation or non-specific amplification (caused by low template), created fluorescence and

generated a Cp value. The reproducibility of Tm values between the duplicates was assessed,

to verify that a specific product was amplified. Reproducibility of the Tm values was checked

visually using the LightCycler histogram that plotted the first negative derivative of the

fluorescence *vs.* temperature. Reproducibility of Cp between duplicates was defined on a

sliding scale with the strictest definition being applied to the lower Cp values; to be accepted, the difference had to be <0.75 Cp for Cp values under 30 and <1.50 Cp for Cp values between 31 and 40. The stringency for reproducibility was relaxed as the Cp increased because greater variability in Cp was observed as the concentration of the starting template decreased and the assay reached its limit of detection (Hughes *et al.* 2006; Van, V *et al.* 2003a). The highest priority was whether S14 had amplified specifically and did the level of amplification represent a sufficient level of starting template for MRD calculations? Samples that did not have a Tm of 89°C and a Cp lower than 41 were classified as "no result". An S14 reaction with a melting temperature other than 89°C represented non-specific amplification and was not used for further analysis. If the S14 amplification was satisfactory the same analysis was repeated with the MLL PTD product data. The MLL PTD amplification required a reproducible Tm equal to 85 or 86°C and a Cp value less than 41 in both duplicates to continue analysis; samples that failed these criteria were classified as negative for the MLL PTD and assigned a ΔCp of 0.
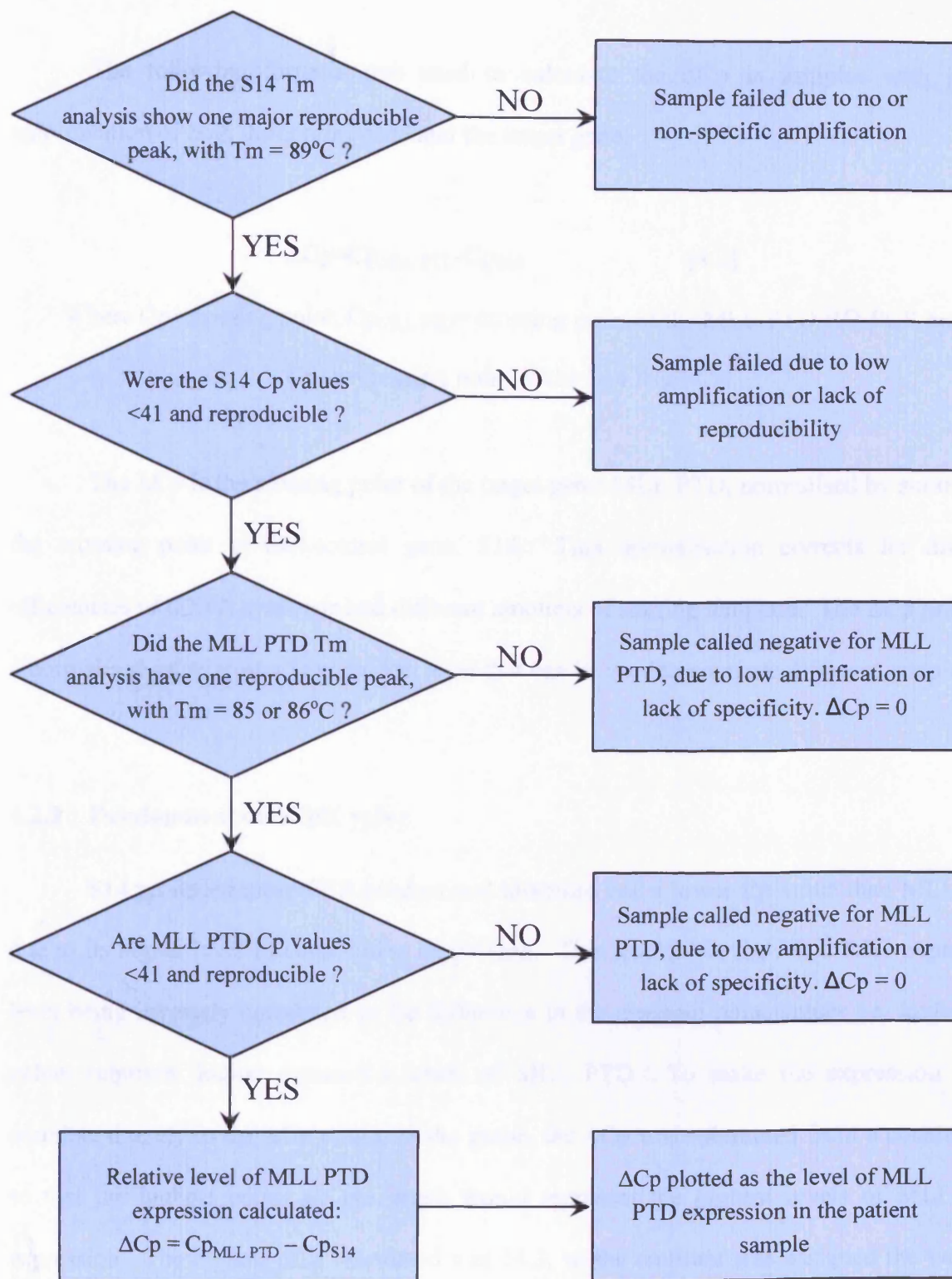
Figure 4.4 Analysis Methodology
Steps taken to verify the specificity of RQ-PCR amplifications. The Tm and Cp values were used to validate the amplification before calculating and plotting the ΔCp. Tm: melting temperature; Cp: crossing point and ΔCp: change in Cp.

The following formula was used to calculate the ΔCp in samples with proven amplification of both the control gene and the target gene:

$$\Delta Cp = Cp_{MLL\ PTD} - Cp_{S14} \qquad [4.2]$$

Where Cp=crossing point, $Cp_{MLL\ PTD}$= crossing point of the MLL PTD RQ-PCR and

Cp_{S14}=crossing point of the S14 RQ-PCR.

The ΔCp is the crossing point of the target gene, MLL PTD, normalised by subtracting the crossing point of the control gene, S14. This normalisation corrects for different efficiencies of cDNA synthesis and different amounts of starting template. The ΔCp produces a normalised calculation of expression level that can be used to compare different samples.

4.2.9 Development of ΔCpK value

S14 produced more PCR product and therefore had a lower Cp value than MLL PTD due to its higher level of constitutive expression. This resulted in the MLL PTD expression level being inversely correlated to the difference in the crossing point values *i.e.* lower ΔCp values represent higher expression levels of MLL PTD. To make the expression levels correlate directly to the ΔCp points on the graph, the ΔCp was subtracted from a constant, K, so that the highest points on the graph would represent the highest levels of MLL PTD expression. The highest ΔCp calculated was 14.3, so the constant was assigned the value of 15 to make all ΔCp values greater than 0. The new variable was designated ΔCp_K and was determined using the following formula (overpage):

$$\Delta Cp_K = K - (Cp_{MLL\,PTD} - Cp_{S14}) \qquad [4.3]$$

Where Cp=crossing point, $Cp_{MLL\,PTD}$= crossing point of the MLL PTD RQ-PCR,

Cp_{S14}=crossing point of the S14 RQ-PCR and the constant K=15.

All results were analysed using this methodology. The change of expression level for minimal residual disease was plotted on a log scale. The ΔCp_K value equivalent to 1 log was calculated. The average slope for the two mutations was –3.6; therefore every 10-fold increase in starting template concentration would decrease the Cp by a value of 3.6. To facilitate this understanding the y-axis of all graphs plotting ΔCp_K were marked in increments of 3.6, with each increment representing a 1 log difference in the expression level of MLL PTD.

The fluorescent history of Syber Green as a function of the cycle number for the MLL PTD samples shown in Figure 4.2 were analysed using the methodology flow chart (Figure 4.4), and the calculated ΔCp_K values were plotted for each sample (Figure 4.5). The results for the positive MLL PTD samples had acceptable Cp and Tm values for both genes so the ΔCp_K was calculated and plotted. The large ΔCp_K values (UPN 77,49) are indicative of high levels of MLL PTD expression. The negative sample, UPN 250 passed all S14 criteria and was categorised as ΔCp_K=0 due to the lack of reproducibility in Tm values produced from the MLL PTD RQ-PCR. However, negative sample UPN 125 passed the criteria for both assays with a calculated ΔCp_K of 4.8; this value is 7.2 units below UPN 77 and 9.6 units below UPN 49, making it at least of 2 logs lower than the positive samples. This indicated that there is clear separation between the expression levels seen in MLL PTD positive samples and the levels seen in MLL WT AML samples.

	49	77	250	125	NC
■ Cp	14.4	12.0	0.0	4.8	0.00

UPN
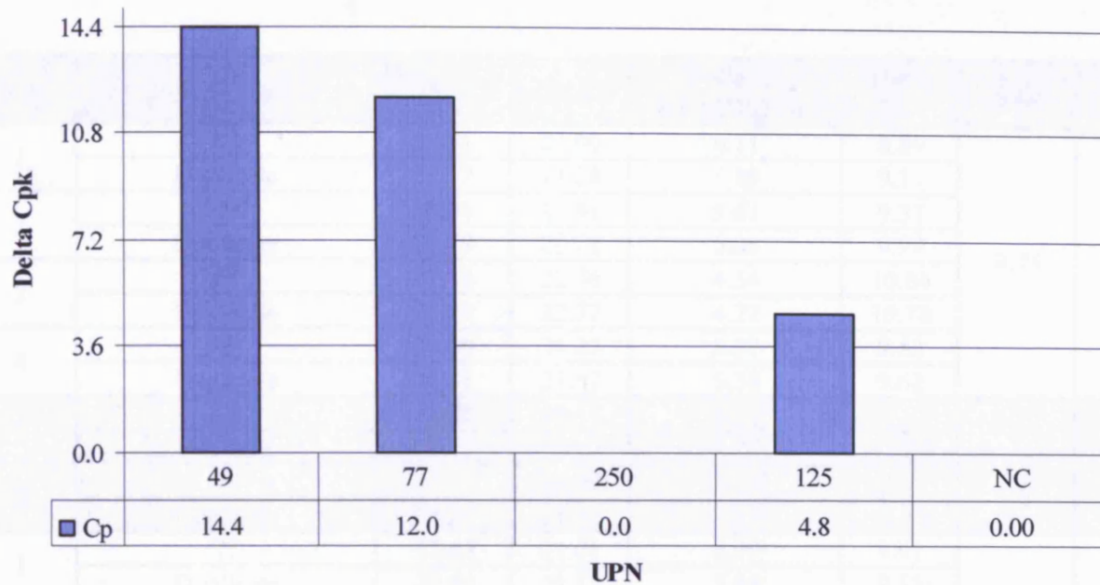
Figure 4.5 CpK Values of Optimised RQ-PCR
ΔCpK normalised expression values of MLL PTD of the samples shown in Figure 4.2. UPN 49 and 77 are MLL PTD positive, UPN 250 and 125 are MLL WT AML samples.

4.2.10 Inter Run Reproducibility of the Assay

The reproducibility of a RQ-PCR assay designed for the detection of MRD is essential if it is to be used as the determining factor for changing patient therapy. The inter and intra run reproducibility was assessed for both MLL PTD and AML MLL WT samples and was found to be consistent. Three separate positive samples were tested to determine the inter-run reproducibility of the assay. The samples were run on different days and used separately prepared reagent mixes to provide the best measure of independence. The results can be seen in Table 4.6. UPN 77 and 17 were run twice and showed a 1.3% and a 4.7% difference from the mean ΔCp_K respectively. Patient sample 49 was run four separate times with a mean ΔCp_K of 9.8±0.6 (n=4). Repeat runs using the AML MLL WT optimisation samples showed inconsistency, this was expected of samples that have template concentrations near the limits of detection of the assay. The samples produced positive and negative results on subsequent runs; all of the positive results had a low ΔCp_K. The donor samples were not tested because no positives were observed.

4-180

Run	UPN No.	MLL PTD Cp	S14 Cp	Delta Cp (MLLPTD-S14)	Delta CpK	Mean	SD
1	49	27.11	21.00	6.11	8.89		
	Duplicate	27.17	21.28	5.89	9.11		
2	49	28.04	22.41	5.63	9.37		
	Duplicate	27.79	22.71	5.08	9.92	9.75	0.64
3	49	27.10	22.76	4.34	10.66		
	Duplicate	26.99	22.77	4.22	10.78		
4	49	26.60	21.23	5.37	9.63		
	Duplicate	26.81	21.43	5.38	9.62		
1	17	23.85	20.13	3.72	11.28		
	Duplicate	23.82	20.12	3.70	11.30	11.56	0.27
2	17	24.17	20.94	3.23	11.77		
	Duplicate	24.23	21.12	3.11	11.89		
1	77	27.48	21.01	6.09	8.91		
	Duplicate	27.00	21.11	5.89	9.11	8.86	0.22
2	77	27.13	21.04	6.09	8.91		
	Duplicate	27.52	21.03	6.49	8.51		

Table 4.6 Inter-run Reproducibility of MLL PTD RQ-PCR
Inter run reproducibility shown in three ML PTD positive samples. All samples were run in duplicate. UPN: unique patient number; No.: number; Cp: crossing point; SD: standard deviation.

4.2.11 Intra Run Reproducibility of the Assay

Assessment of the intra-run reproducibility showed a very high correlation between the duplicates for both assays. To measure the level of intra run variation for the MLL PTD RQ-PCR, the mean Cp of the duplicates was plotted against the minimum and maximum values (Figure 4.6). For the MLL PTD data, all samples that passed the methodology criteria for S14 and had reproducible Tm for MLL PTD were plotted, resulting in 40 samples. The Cp values ranged from 23.2-37.7 due to the differences in the expression level of MLL PTD, yet the mean difference between the minimum and maximum value remained low at 0.3 ± 0.3 (n=40). A gradual loss of reproducibility occurred in parallel with the increase of mean Cp values as the assay approached the maximum level of sensitivity. Still larger variations were observed at higher Cp values, but these samples were not plotted because they failed to pass the Tm reproducibility checkpoint (non-specific amplification) flowchart.
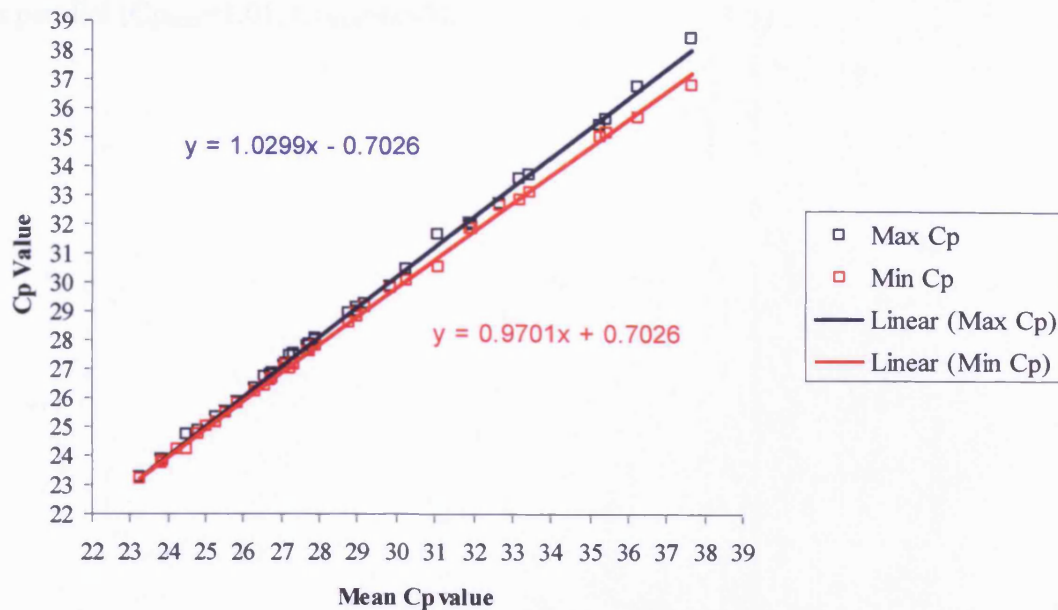
Figure 4.6 MLL PTD Intra Run Reproducibility
The minimum and maximum Cp values of the duplicates of all MLL PTD samples that met analysis criteria, n=40. Cp: crossing point; Max: maximum; Min: minimum.

A gradual increase in the difference of the duplicates in the S14 assay was observed as Cp value increased and was used in combination with the analysis methodology to establish a low template (high Cp) cut-off value for poor quality samples. In Figure 4.7A, the mean Cp value was plotted against the minimum and maximum duplicate values of all S14 samples if they met the reproducible Tm criteria for the S14 methodology, resulting in 73 samples. A slight difference in slope (Cp_{max}=1.03, Cp_{min}=0.97) is observed with a greater separation between duplicates after the Cp of 25. When the criterion of reproducible Cp was applied to this set of samples, 3 of 11 (27.3%) samples with Cp>25, failed compared to 1 of 62 (1.6%) samples with Cp<25, indicating a higher percentage of poor quality samples in the Cp>25 group. Many of the samples that failed to meet the reproducible Tm criteria for S14 also exhibited high variation in the Cp duplicates as the result of non-specific amplification. After the reproducible Cp criteria was applied the remaining samples were plotted as shown in

Figure 4.7B and the linear trend lines graphed for the minimum and maximum Cp values are more parallel ($Cp_{max}=1.01$, $Cp_{min}=0.99$).
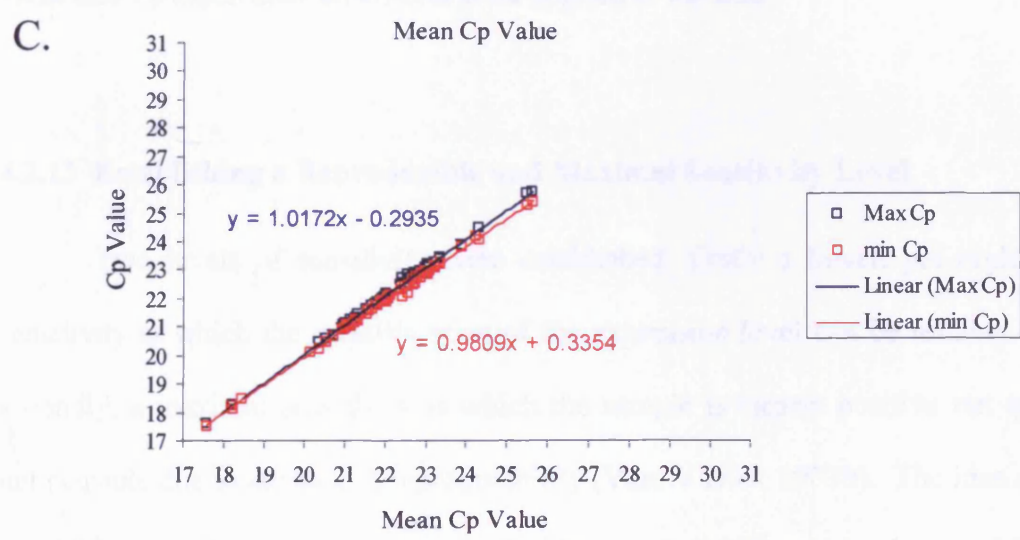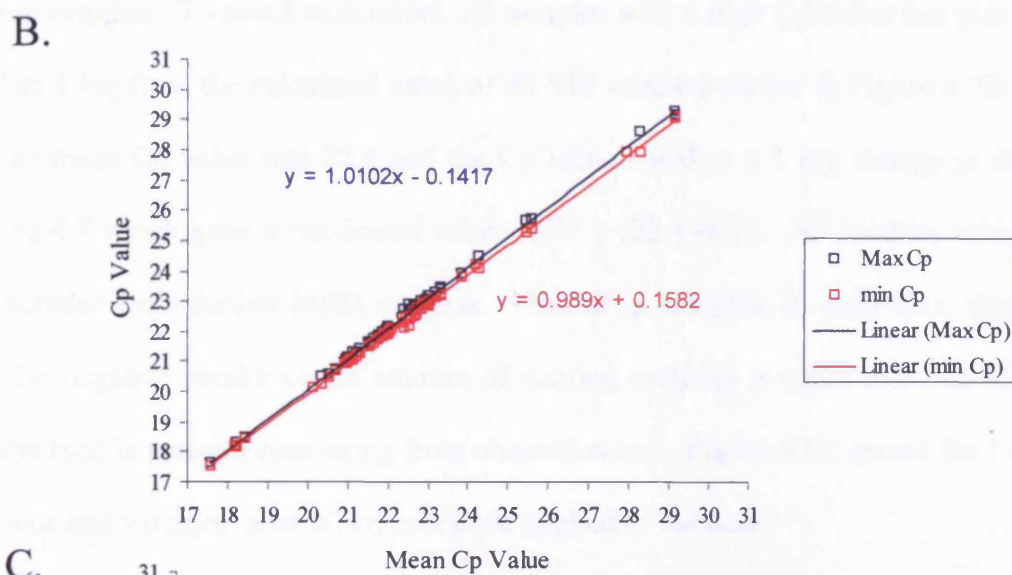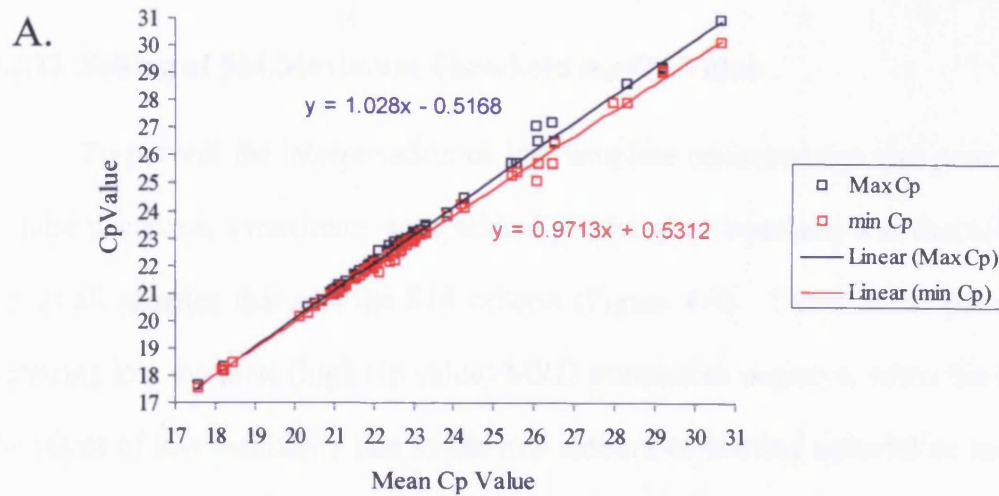
Figure 4.7 S14 Intra Run Reproducibility
The minimum (blue) and maximum (red) Cp values of the duplicates of S14 samples. A. All samples with a reproducible melting temperature (Tm), n=73. B. Samples with reproducible Tm and Cp values, n=67. C. Samples with Cp≤27 and reproducible Tm and Cp. Cp: crossing point; Max: maximum; Min: minimum.

4-184

4.2.12 Setting of S14 Maximum Threshold for Cp Value

To prevent the interpretation of low template concentration and poor quality samples as false negatives, a maximum acceptable Cp value (low template) was derived from the mean Cp of all samples that met the S14 criteria (Figure 4.4). There is the potential for falsely declaring low template (high Cp value) MRD samples as negative, when the high Cp may be the result of low sensitivity due to the low amount of starting material or inefficient reverse transcription. To avoid such errors, all samples with a high Cp value that placed them farther than 1 log from the calculated mean of all S14 samples shown in Figure 4.7B were excluded. The mean Cp value was 22.4 and the Cp value equal to a 1 log change in starting template was 4.7 which gave a maximum value of 27.1 (22.4+4.7). All samples with a Cp>27 were excluded from further MRD analysis. Follow up samples, in particular, would be prone to false negative results as the amount of starting material is often low due to the low WBC observed in patients recovering from chemotherapy. Figure 4.7C shows the final range of Cp value and variation after all criteria were applied to the data.

4.2.13 Establishing a Reproducible and Maximal Sensitivity Level

Two levels of sensitivity were established; firstly a lower, yet highly reproducible sensitivity at which the quantification of the expression level can be reliably calculated and, secondly, a maximal sensitivity at which the sample is clearly positive but quantification is not possible due to the lack of reproducibility (Van, V *et al.* 2003b). The idea of two levels of sensitivity was first suggested by van Velden *et al.* (2003) and is also used in the proposed *BCR-ABL* guidelines for reporting disease levels for CML patients. In the *BCR ABL* guidelines, the level of "confidence" in the result is reduced as the copy number of the control gene decreases, until eventually the lowest level positive is reported as a positive of

undetermined quantity. The reproducible sensitivity was calculated as 10^{-2} using the cDNA

dilution experiment (Section 4.2.15). The lack of reproducibility observed in low-level MLL

PTD positives was reflected by variation in the duplicate values of the Cp and, more

frequently, in the visualisation of the Tm, where samples failed because they had more than

one Tm peak. These double peak positive (DPP) MLL PTD samples were identified by the

appearance of two Tm peaks instead of one; with one major reproducible peak at a recognised

MLL PTD Tm and a minor peak at a slightly higher Tm. The major peak was defined as the

taller of the two peaks present. A sample designated $\Delta Cp_K=0$ because of a high or varied Cp

value was re-classified if it met the definition of a DPP. DPPs were easily distinguished from

the donor samples, which showed either no amplification, no reproducibility in Tm or had

more than two peaks with one below the expected Tm of MLL PTD. The use of maximal

sensitivity to detect DPPs potentially increased the detection of a MLL PTD positive sample

by an order of 10, to 10^{-3} .

The use of the amended analysis methodology, including the maximum Cp value for

S14, resulted in fewer samples being analysed overall and a higher number of positives. All

of the incorporated amendments to the protocol can be seen in Figure 4.8. The higher

stringency for S14, to prevent analysis of poor quality samples, excluded 5 samples from

analysis. All 5 samples were $\Delta Cp_K=0$. Six samples previously classified as $\Delta Cp_K=0$ were re-

classified as low positives (DPPs) using the guidelines for maximal sensitivity. Three of the

samples were from MLL WT AML patients and 3 were from MLL PTD positive patients. No

donor samples changed from $\Delta Cp_K=0$ to DPP. As expected, the amended guidelines excluded

poor samples, which may have led to false negative reporting, left samples with high ΔCp_K

values unchanged and reclassified some low positive samples previously excluded due to low

reproducibility to DPP.
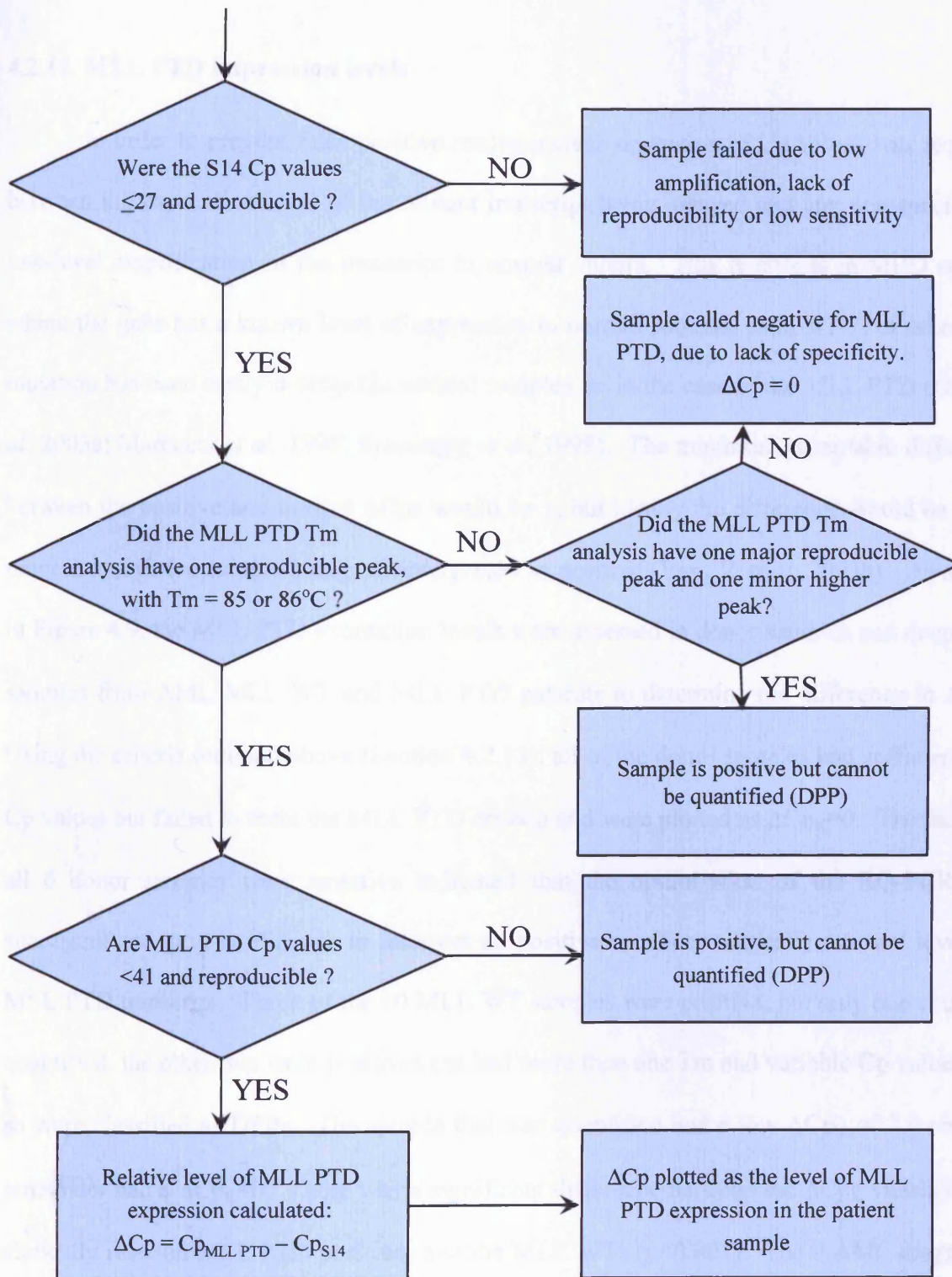
Figure 4.8 Amended Analysis Methodology
Analysis protocol with S14 Cp threshold and maximal sensitivity amendments incorporated. The Tm and Cp values were used to validate the amplification before calculating and plotting the ΔCp. Tm: melting temperature; Cp: crossing point and ΔCp: the change in Cp; DPP: double peak positive.

4.2.14 MLL PTD Expression levels

In order to prevent false positive results a clear separation of Cp values was required between the expression level of the mutant transcript being assayed and any non-specific or low-level amplification of the transcript in normal donors. This is critical in MRD studies where the gene has a known level of expression in normal patients, as in *WT-1*, or where the mutation has been easily detected in normal samples, as in the case of the MLL PTD (Garg *et al.* 2003a; Marcucci *et al.* 1998; Schnittger *et al.* 1998). The minimum acceptable difference between the positive and normal ΔCps would be 1, but ideally the difference would be ≥ 3 to prevent a negative sample being misinterpreted as positive (Van, V *et al.* 2003b). As shown in Figure 4.9, the MLL PTD expression levels were assessed in donor samples and diagnostic samples from AML MLL WT and MLL PTD patients to determine the difference in ΔCp_K. Using the criteria outlined above (Section 4.2.13), all of the donor samples had sufficient S14 Cp values but failed to meet the MLL PTD criteria and were plotted as $\Delta Cp_K=0$. The fact that all 6 donor samples were negative indicated that the optimisation of the RQ-PCR was successful and gave confidence to interpret all positive results as clinically relevant levels of MLL PTD transcript. Three of the 10 MLL WT samples were positive, but only one could be quantified, the other two were positives but had more than one Tm and variable Cp values and so were classified as DPPs. The sample that was quantified had a low ΔCp_K of 3.0 and the remainder had a $\Delta Cp_K=0$. There was a significant difference between the ΔCp_K values of the clinically relevant MLL PTD positives and the MLL WTs (p<0.001). The 9 AML diagnostic samples with MLL PTD all tested positive with a median ΔCp_K of 9.8 (range, 8.0-12.4). Using a ΔCp_K of 3.6 as an equivalent to 1 log, the range of the positives was 1.2 logs and with a median value of 9.8 the assay should provide nearly 3 log of sensitivity for MRD follow up samples.
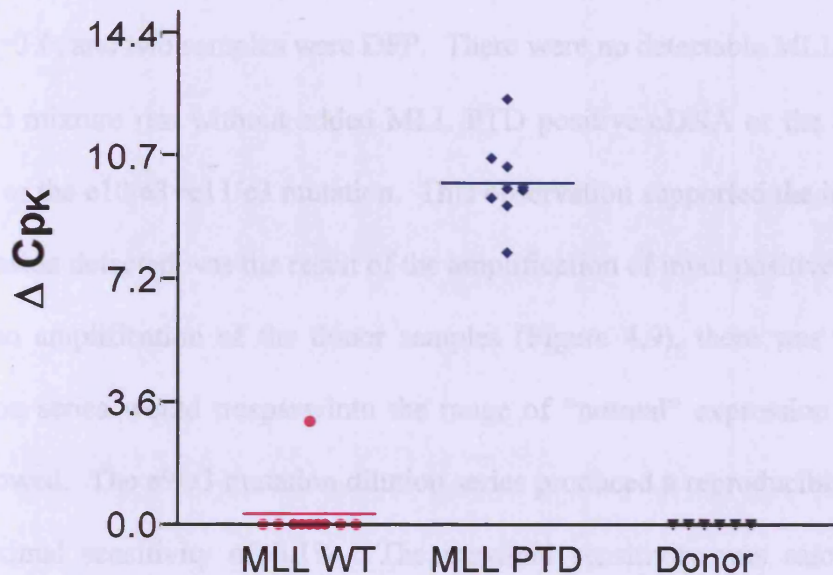
Figure 4.9 MLL PTD Expression Levels in Different Populations
Scatter graph of MLL PTD expression levels using delta Cpk values. MLL WT AML patients, MLL PTD positive patients and healthy donors. Cp: crossing point; Δ Cp: change in Cp.

4.2.15 Sensitivity of the RQ-PCR Assay

The sensitivity of an MRD RQ-PCR is important for determining the clinical relevance of the assay. The type of RQ-PCR assay has a bearing on the achievable sensitivity; a level of 1:10,000 to 1:1,000,000 is a desirable level to aim for in most cases (Grimwade & Hills 2009). However, in the case of the MLL PTD assay a much lower level of sensitivity would be clinically relevant for two reasons: 1) the majority of MLL PTD AML cases are cytogenetically normal and so have no straightforward MRD marker other than morphology, and 2) an assay that is made too sensitive will begin to amplify MLL PTD transcripts that are not clinically relevant. The sensitivity of the MLL PTD RQ-PCR assay in follow up samples was estimated by diluting a positive diagnostic sample representative of

each of the MLL PTD mutations into a pool of AML WT cDNA (Figure 4.10). The pooled dilution mixture of cDNA was composed of the samples used to calculate the expression level of MLL PTD in MLL WTs, where one sample displayed a low level of MLL PTD expression, $\Delta Cp_K=3.0$, and two samples were DPP. There were no detectable MLL PTD transcripts in the pooled mixture run without added MLL PTD positive cDNA or the 0.01% dilutions of the e9/e3 or the e10/e3+e11/e3 mutation. This observation supported the idea that any MLL PTD expression detected was the result of the amplification of input positive cDNA. Because there was no amplification of the donor samples (Figure 4.9), there was no point at which the dilution series would trespass into the range of "normal" expression levels and have to be disallowed. The e9/e3 mutation dilution series produced a reproducible sensitivity of 1% and a maximal sensitivity of 0.1%. The maximal sensitivity was calculated using the DPP criteria. Despite the lower initial ΔCp_K value of the 100% sample, the e10/e3+e11/e3 mutant obtained the same sensitivity levels as the e9/e3 mutation. Taking into consideration that these were patient samples that did not have the MLL PTD in every cell, both samples had a blast count of around 60%, the maximum sensitivity would be expected to be greater.
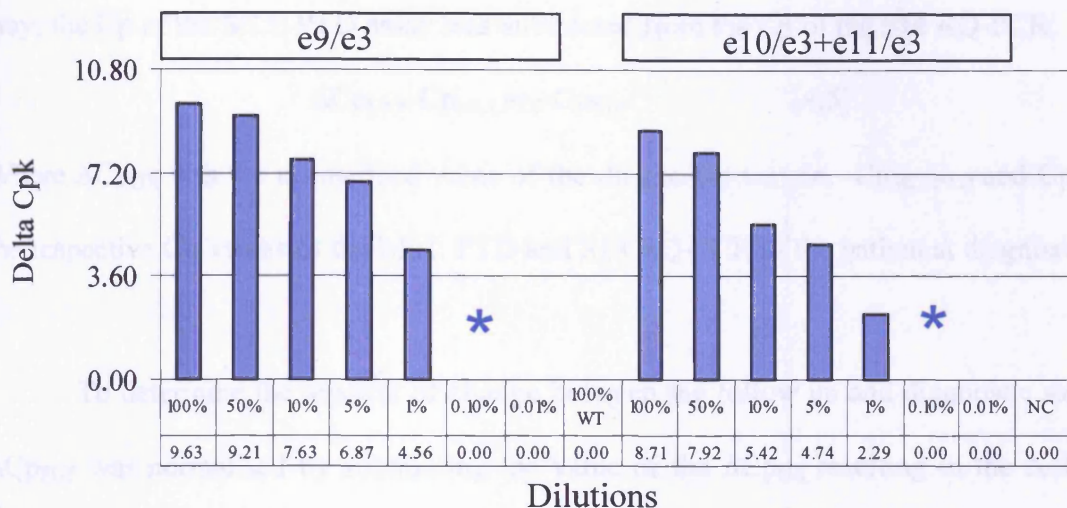
Figure 4.10 Sensitivity of MLL PTD RQ-PCR
Series of 9/3 and 10/3+11/3 positive patient samples diluted into MLL WT AML cDNA. Negative control used water as template. "*" indicates a sample that was positive by maximal sensitivity (DPP). Cp: crossing point; WT: MLL wild type; NC: negative control.

4.2.16 Calculation of MRD Using ΔΔCp

The data from the sensitivity test was used as a mock MRD case to determine if The ΔΔCp formula developed by the Europe Against Cancer network for the calculation of MRD could be applied to the MLL PTD RQ-PCR assay (Gabert *et al.* 2003). The ΔΔCp method was developed to allow the level of fusion genes transcripts detected in a patient during the course of cytoreductive treatment to be presented graphically in an intuitive format. Using this method, the follow up Cp of the target gene, MLL PTD, was first normalised by subtracting the Cp of the control gene, S14, from the fusion gene to give the ΔCp_{FUP}.

$$\Delta Cp_{FUP} = Cp_{MLLPTD} - Cp_{S14} \qquad [4.4]$$

Where ΔCp_{FUP} was the normalised value of the follow up sample, Cp_{MLLPTD} was the Cp value of the follow up MLL PTD RQ-PCR and the Cp_{S14} was the follow up S14 Cp value of the follow up sample.

The normalised value for the diagnostic sample, ΔCp_{DX}, was calculated in the same way; the Cp of the MLL PTD assay was subtracted from the Cp of the S14 RQ-PCR.

$$\Delta Cp_{DX}= Cp_{MLLPTD}-Cp_{S14} \qquad [4.5]$$

Where ΔCp_{DX} was the normalised value of the diagnostic sample. Cp_{MLLPTD} and Cp_{S14} were the respective Cp values of the MLL PTD and S14 RQ-PCR of the patient at diagnosis.

To determine the amount of change between the follow up and diagnostic sample the ΔCp_{FUP} was normalised by subtracting the value of the ΔCp_{DX} resulting in the $\Delta\Delta Cp$. The MRD value was then calculated using the formula:

$$MRDv=10^{((\Delta CpFUP-\Delta CpDX)/(-S))} \qquad [4.6]$$

Where MRDv is the minimal residual disease value, ΔCp_{FUP} and ΔCp_{DX} are the normalised values for the follow up and diagnostic samples respectively, and S is the slope of the RQ-PCR assays.

As outlined above (Section 4.2.7), when there is a difference between the slopes of the fusion gene and the control gene, the slope of the fusion gene should be used because the Cp values of the control gene occur in a narrower range and therefore would have only a minimal impact on the result. The smaller range of S14 Cp values were due to the equal amount of RNA as template for the cDNA reaction and the exclusion of high Cp values representative of poor quality samples. The two mutation types created by the MLL PTD have slight differences in their relative slopes, so to increase the accuracy at low levels of MRD the individual slope values of each mutation were used in the calculation. The results plotted in Figure 4.11 show concordance between the dilutions made and the level of MRD calculated using the $\Delta\Delta Cp$ method despite the difference in slopes between the MLL PTD and S14

assays and between the MLL PTD assays themselves. The lowest reproducible sensitivity for both mutations was between 2% and 3%, and the maximal sensitivity was detected down to the level of 0.1%. No product was detected in the 0.01% dilution for either mutation. There was a slight increase in the level of S14 Cp values (decrease in template concentration) as the concentration of MLL PTD cDNA decreased; this may cause the MRD value to be higher than the dilution value. The assay sensitivity achieved was 2 logs with the reproducible sensitivity and 3 logs overall using the maximal sensitivity.

	100%	50%	10%	5%	1%	0.10%
e10/e3 +e11/e3	100.0	61.8	13.6	9.0	2.0	0.0
e9/e3	100.0	75.2	25.9	15.4	3.2	0.0

Sample Concentration
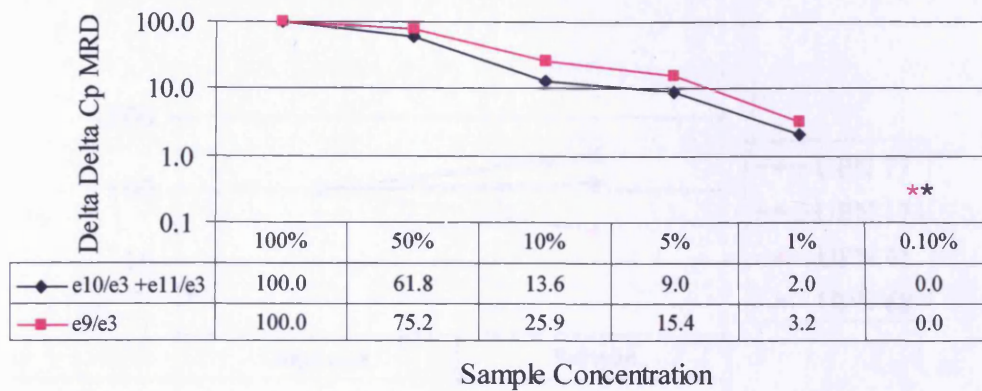
Figure 4.11 MRD Plot of MLL PTD Transcript Sensitivity Dilutions
Dilution series of e9/e3 and e10/e3+e11/e3 positive patient samples diluted into MLL WT AML cDNA plotted using the delta delta Cp method. "*" indicates a sample that was positive by maximal sensitivity (DPP). MRD: minimal residual disease; Cp: crossing point.

4.2.17 Stability of the MLL PTD Marker

To determine if the MLL PTD was a stable molecular marker, 4 paired samples were tested at diagnosis and relapse. All four of the markers were positive at both time points and all relapsed with the same transcript type seen at diagnosis, which suggested the marker was stable and that it was a potential marker for MRD. The diagnostic average using ΔCp_K was 9.2$\pm$1.2 (n=4) and the average relapse using $\Delta\Delta Cp$ was 186$\pm$133.2 (n=4). The ΔCp_K values indicate variation in the diagnostic samples, these were normalised using the $\Delta\Delta Cp$ method so differences in levels at diagnosis between individual patients are not apparent in the figure. All of the relapse samples showed a substantial level of transcript; none were more than one log from the diagnostic level, as shown in Figure 4.12. Three of the four patients relapsed with expression levels higher than that seen at diagnosis and the fourth was lower, at 27% of the diagnosis value. Expression levels of this magnitude may indicate a more rapid rate of relapse, which would require more frequent monitoring to detect the relapse onset. In addition, a positive result from the sample of a patient who failed to reach remission showed the marker was stable in patients with resistant disease.

4-194
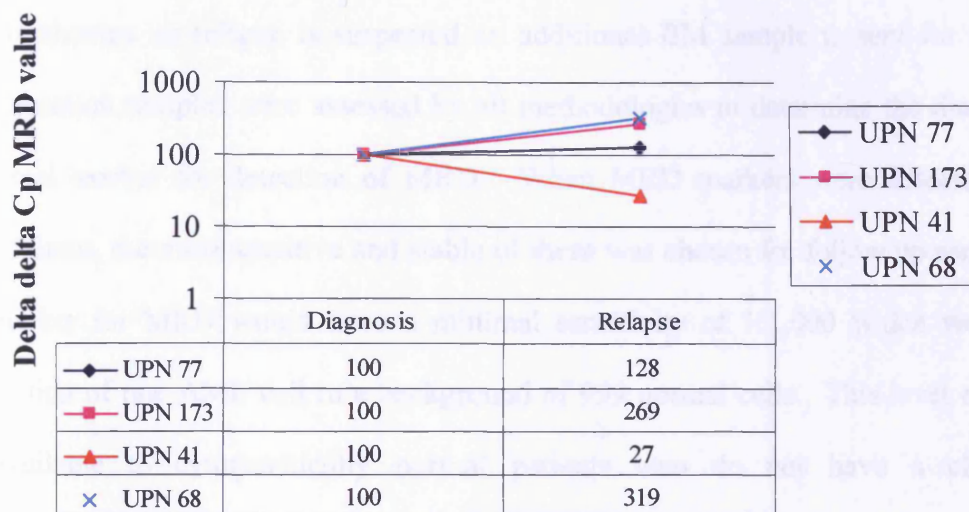
Figure 4.12 MLL PTD Expression Level at Diagnosis and Relapse
MRD values of expression levels of patients at diagnosis and relapse. The level of MRD was
calculated using the $\Delta\Delta Cp$ MRD method therefore the diagnostic level is set at 100%.

4.2.18 Patient MRD Data

The MRD status of eight AML patients with MLL PTD was monitored longitudinally
to assess the value of RQ-PCR for detecting levels of disease and predicting relapse. The RQ-
PCR assay was compared against the "standard" measures of disease available to the patients:
BM morphology and, if applicable, immunophenotyping and cytogenetic analysis.

4.2.19 Timepoints for MRD Analysis

The time points used to determine the progress of the patient are aligned to the
administration of the treatment regime. The standard time points for measurement are at
diagnosis and then following induction and consolidation treatments subsequent to the
recovery of the WBC. The frequency of follow up samples after consolidation is reduced for
patients in complete remission (CR). A BM sample is collected for analysis at the each of

these points, however, if at any time in a patient's treatment there is a lack of response to chemotherapy or relapse is suspected an additional BM sample is sent for analysis. The presentation samples were assessed by all methodologies to determine the diagnosis and the optimal marker for detection of MRD. When MRD markers were detected by multiple disciplines, the most sensitive and stable of these was chosen for follow up samples. Ideally, a marker for MRD would have a minimal sensitivity of 1:1,000 which would allow the detection of one AML cell in a background of 999 normal cells. This level of sensitivity is unavailable to cytogenetically normal patients who do not have a clearly aberrant immunophenotype; a patient with this profile is limited to a sensitivity of around 5%, obtained from morphological assessment of BM (Liu Yin 2002b). Follow up samples were sent for analysis to the departments determined to offer the best MRD method available to the patient. Due to the lack of a standard molecular marker for PCR at diagnosis, the samples at some time points were not sent for molecular analysis so the data sets are incomplete.

4.2.20 Patient Characteristics

The samples from 8 MLL PTD patients, selected for the availability of follow up time points, were analysed to evaluate the MLL PTD MRD assay. The patient characteristics at diagnosis are listed in Table 4.7. The criteria given in the 2001 WHO classifications were used to diagnose 7/8 patients as *de novo* AML (Vardiman, Harris, & Brunning 2002). One patient was diagnosed with MDS transforming to AML and was treated according to the AML 15 protocol as described previously (Chapter 2 Section 2.3.2). All 8 patients were determined to have the MLL PTD by the qualitative PCR as described in Chapter 3 and this was confirmed by RQ-PCR. There was an average of 2.8 samples received per patient (n=22). There were 4 male and 4 female patients with a median age of 60 years (range 35-72). The patients had a number of similarities; of the seven patients with available data, all

but one presented with pancytopenia and the blast counts, as determined by BM morphology, of the AML samples were tightly grouped with a median of 56% (range 56-63%). All of the samples except for one were cytogenetically normal and were therefore categorised as intermediate prognosis, using cytogenetic guidelines (Grimwade *et al.* 1998). The one sample with an aberrant cytogenetic karyotype was a 7q(del), which is also classified as intermediate prognosis. None of the molecular markers used in the AML RT-PCR screen (*PML-RAR*, RUNX1-RUNXT1, or CBFB/*MYH11*) were observed. These characteristics were in line with the results presented in Chapter 3 and with published reports of MLL PTDs association with cytogenetically normal samples and a lack of association with favourable risk cytogenetic translocations (Caligiuri *et al.* 1998; Schnittger *et al.* 2000). Surprisingly, none of the samples were positive for the *FLT3* ITD, which can co-occur with the MLL PTD (Chapter 3) (Olesen *et al.* 2005a; Steudel *et al.* 2003). The MLL PTD transcripts were the two most commonly observed types, e9/e3 and e10/e11 + e3/e11, with 5/8 (63%) and 3/8 (37%) respectively. All patients received intensive treatment.

UPN	Sex	Age	WBC (x10^9/L)	FAB	BM Morphology Blast (%)	Cytogenetic Marker	Cytogenetic Prognosis	MLL PTD Type	FLT3-ITD Status	Clonal Evolution
77	M	61	0.8	M0	60	No	Intermediate	e10/e3+e11/e3	WT	No
173	F	42	1.0	M2	56	No	Intermediate	e9/e3	WT	Yes
41	M	63	60.0	M2	63	No	Intermediate	e9/e3	WT	Yes
188	F	72	1.9	M6	50	No	Intermediate	e9/e3	WT	No
68	M	37	n/d	M2	n/d	Yes (del 7q)	Intermediate	e10/e3+e11/e3	WT	Yes
49	M	60	3.0	M2	56	No	Intermediate	e9/e3	WT	Yes
150	F	35	3.6	MDS	10.0	No	Intermediate	e9/e3	WT	Yes
17	F	60	2.6	M1	57	No	Intermediate	e10/e3+e11/e3	WT	Yes

Table 4.7 Diagnostic Characteristics of MLL PTD
Characterisation of the MLL PTD positive patients assayed for MRD monitoring. UPN: unique patient number; L: litre; ITD internal tandem duplication; M: male; F: female; n/d: not determined; WT: wild type.

The number of samples exhibiting clonal evolution either at relapse or in refractory samples was higher than expected. When both relapsing and refractory patients were included clonal evolution was detected in 75% (6/8) of the samples. 71% (5/7) percentage of CN MLL PTDs showed clonal evolution and despite the low number of patients tested this was substantially higher than the range of 25-51% reported in previous studies (Estey *et al.* 1995; Garson *et al.* 1989; Kern *et al.* 2002; Testa *et al.* 1979). Comparison of karyotypes at diagnosis and relapse revealed evidence of clonal evolution (Table 4.8). All of the changes involved unbalanced translocations and mutations considered to be secondary abnormalities when found at diagnosis. The cytogenetic classification of two of the six patients changed from an intermediate to an adverse prognostic classification as a result of >5 cytogenetic aberrations, while the other four patients remained intermediate, despite changing from normal to aberrant karyotypes. One of the six patients that exhibited clonal evolution was considered to have refractory disease; supporting the theory that karyotype instability may promote the acquisition of a phenotype that is resistant to anti-leukaemic treatment (Kern *et al.* 2002).

UPN	Karyotype at diagnosis	Prognosis	Karyotype at Relapse	Prognosis	Δ from Diag. Prognosis
173	46,XX[30]	Intermediate	47,XX,+21[5]	Intermediate	No
41	46,XY[20]	Intermediate	46,XY,add(1)(q4?1),add(17)(q24)[3]/46,XY[47]	Intermediate	No
68	46,XY,del(7)(q23q22)[8]/46,XY[2]	Intermediate	47,XY,add(5)(p13),-6,-9,-17,+4mar[3]/46XY[3]	Adverse (5)	Yes
49	46,XX	Intermediate	47,XX,+8[1]/46,XX[29]	Intermediate	No
150	46,XX	Intermediate	46,XX,?add(6)(q2?5)[3]/46,XX[17]	Intermediate	No
17	46,XX	Intermediate	47,XX,-4,inv(7)(q22q36),-17,+22,+mar1,+mar2[2]/46,XX,-4,t(5;9)(q13;p2?2),inv(7)(q22q36),add(12)(q24),t(13;22)(q1?2;q13),-	Adverse (5)	Yes

Table 4.8 Evidence of Clonal Evolution (Karyotypic) Instability) iin Cytogenetically Normal Patients
Karyotypes of patients with clonal evolution. The diagnostic karyotype and relapsed or refractory karyotypes are shown with prognosis based on cytogenetic analysis. Column 6 indicates whether the change in karyotype changes the cytogenetic prognosis group. Δ:Change; "Adverse (5)" signifies adverse prognosis due to 5 or more cytogenetic aberrations. UPN: unique patient number.

4.2.21 Correlation of MRD Methods

The results of the MLL PTD RQ-PCR assay were compared against the standard measures of MRD that were available for each patient (BM morphology, immunophenotyping and cytogenetic analysis), to prove correlation with existing methods, to compare sensitivity and to determine if early detection of relapse was possible. The results of the BM morphology report were used to determine remission using the WHO criteria of <5% blasts to define morphological remission. Relapse was also determined from analysis of the BM morphology. Concordant results were observed between the MLL PTD RQ-PCR and all standard MRD methods when diagnostic, relapsed and refractory samples were analysed (Table 4.9). All presentation samples exhibited a high level of expression with a ΔCp_K median of 9.3 (range, 8-10.7, n=7). Follow up samples from each patient were normalised relative to the value at diagnosis. The samples that were designated as relapse or refractory by BM morphology, immunophenotyping or cytogenetic analysis all had substantial MLL PTD expression levels and were within 1 log of the diagnostic value. In four samples where morphological remission was indicated standard MRD markers did not detect disease while MLL PTD RQ-PCR was able to detect a decreased amount of transcript (between 1 and 2 logs below the diagnostic level). The alternative values were a result of the higher sensitivity of the MLL PTD RQ-PCR. The level of MRD was determined in three of the samples with values of 2, 3 and 4%; the fourth was found to be a DPP. The detection of MRD by MLL PTD RQ-PCR in samples designated as remission samples by standard methods demonstrated that a higher level of sensitivity was obtained using the new RQ-PCR assay. There were no instances where the MLL PTD RQ-PCR was negative and a standard MRD method showed

evidence of disease. The only sample that showed a discordant result was the 2 month sample for UPN 77. However, the marrow was described in the morphology report as a hypocellular sample and, as such, it may not have been representative of the whole and the fraction of blasts was calculated as 2% by morphology, while the level of MRD calculated was 52%. The quality of the RNA isolated was acceptable in both concentration and quality as determined by spectrophotometry and all of the criteria for the MLL PTD assay were met. Two subsequent marrows that were not received by molecular diagnostics substantiated the original MLL PTD MRD result with increased levels of blasts seen in each sample. In this example, it appeared that the RQ-PCR assay detected the underlying level of disease despite the poor quality and low quantity of the sample received, while the standard BM morphology result was ambiguous and had to be repeated. Concordance between the results of the RQ-PCR assay and the standard methods of MRD occurred in 77% (17/22) of samples. In the remaining 23% (5/22) of samples, the higher sensitivity of the RQ-PCR assay detected MRD in samples declared to be free of disease by other methods.

UPN	77				173				41		
Time from Diagnosis (months)	0	2*	4	16	0	3	5	8	0	1	5
WBC (x10^9/L)	0.8	0.8	5.9	3.4	1	0.7	1.6	1.2	60	4.7	3.7
BM Blast Count (%)	60	2	4	22	56	59	4	n/d**	63	2	21
Immunophenotyping Blast Count (%)	58	n/d	<0.5	10	30	65	0	n/d***	65	n/d	22
Cytogenetic Analysis Cell Count	n/a	n/a	n/a	n/a	n/a	5/5	0/200	17/20	n/a	n/a	3/50
MLL PTD QPCR - Delta Delta Cp (%)	100	52	2	128	100	165	4	269	100	DPP	27

UPN	188			68		49		150		17	
Time from Diagnosis (months)	0	1	2	0	27	0	4	0	2	10	12
WBC (x10^9/L)	1.9	0.5	n/d	n/d	8	3	1.7	3.6	n/d	2.9	6.1
BM Blast Count (%)	50	25	4	n/d	73	56	52	10	n/d	85	n/d
Immunophenotyping Blast Count (%)	40	20	5	n/d	n/d***	n/d***	n/d	34	n/d	n/d***	n/d***
Cytogenetic Analysis Cell Count	n/a	n/a	n/a	8/10	3/6	n/a	1/30	n/a	3/20	7/10	n/d
MLL PTD QPCR - Delta Delta Cp (%)	100	464	3	100	319	100	582	100	15	100****	130

Table 4.9 Comparison of Standard Methods of MRD for MLL PTD Patients Over the Course of Treatment
All results in red are refractory disease or relapse. UPN: unique patient number; WBC: white blood count; L: litre; BM: bone marrow; n/d: not determined; Cp: crossing point. *Hypocellular sample, may not be representative, BM and flow done at 2.5 and 2.8 months and showed increasing blasts; **BM trephine observed infiltration of myeloblasts - no percentage given; ***Majority of cells are blasts - no percentage given; ****Relapse sample, assigned 100%;

4.2.22 Prediction of Relapse and Attainment of Remission

Although neither molecular remission nor prediction of relapse was demonstrated with the MLL PTD RQ-PCR, this is likely to stem from inadequate sampling frequency rather than lack of sensitivity of the assay. The assay has clearly shown healthy donor samples to have no detectable levels of MLL PTD transcript, but none of the patients positive for the MLL PTD reached this point during their course of treatment. The molecular MRD data for all 8 patients are shown in Figure 4.13. The sample closest to reaching negativity was the one month time point of UPN 41, which reached a >2 log reduction to DPP, only to relapse 4 months later. No samples showed clear evidence of relapse prior to detection of relapse by other methods. However, the failure to clear the disease was apparent in four patients who subsequently relapsed and this, combined with the higher sensitivity of the RQ-PCR assay, suggests the ability to detect early relapse if samples were collected at more frequent time points. The paucity of longitudinal patient monitoring at regular intervals limits the interpretation of this study's findings. A greater number of samples from patients with longer event free survival (EFS) may have demonstrated whether the MLL PTD transcript became undetectable and whether the assay was capable of predicting relapse. Many of the samples received from patients with longer EFS were sent for molecular diagnostics only because the relapse was already suspected so the opportunity to predict an early relapse was lost.
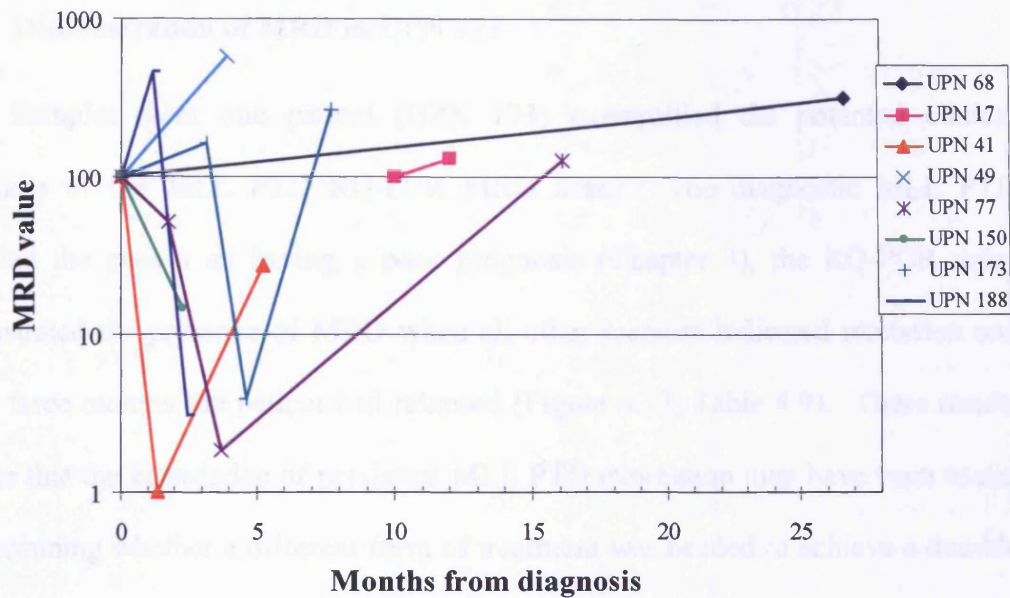
Figure 4.13 Time Course Monitoring of MRD in Eight Patients with AML
Longitudinal MRD monitoring of MLL PTD positive patients over the course of
treatment. MRD values were calculated using the $\Delta\Delta Cp$ MRD method. Patients 188,
49 and 173 had refractory disease. Patients 41, 77 and 173 had initial reductions of
greater than 10% before relapse. The level of 1 represents DPP for patient 41. The
two samples plotted for patient 17 are successive relapse samples, the diagnostic
sample was not available, UPN: unique patient number.

4.2.23 Demonstration of MRD in UPN 173

Samples from one patient (UPN 173) exemplified the potential clinical usefulness of the MLL PTD RQ-PCR MRD assay. The diagnostic MLL PTD identified the patient as having a poor prognosis (Chapter 3), the RQ-PCR assay demonstrated the presence of MRD when all other markers indicated remission and within three months the patient had relapsed (Figure 4.13, Table 4.9). These results indicate that the knowledge of persistent MLL PTD expression may have been useful in determining whether a different form of treatment was needed to achieve a durable remission.

The patient was a 42 year old woman who was pancytopenic with a WBC of 1.0×10^9 cells/l, a blast level of 56% and 30% respectively for BM and immunophenotyping: she was diagnosed as AML FAB type M2. The MLL PTD RQ-PCR analysis detected a diagnostic level of disease with a ΔCp_K of 10.7 (diagnostic range 8-12.4). No markers were found for cytogenetics or molecular diagnostics, therefore immunophenotyping was recommended as the most useful method for the detection of MRD. On the basis of cytogenetic analysis, the patient was classified as intermediate prognosis. The patient was entered into the AML 15 trial and received standard treatment. Following the first course of chemotherapy, the patient was in CR with less than 3% blasts by morphology and less than 1% with flow cytometry. The next sample for molecular diagnostics was received 3 months post diagnosis with the clinical details "?relapse" and WBC of 0.7×10^9 cells/l. The BM morphology showed normal cellularity and 59% blasts; the conclusion was relapsed AML. The immunophenotyping detected blasts comprising 65% of the nucleated cells and the loss of the TdT marker seen at diagnosis. Cytogenetic analysis provided confirmation

of relapse, with the observation of trisomy 21 in 5/5 cells, confirmed with TEL/*AML1*

fluorescent *in situ* hybridisation (FISH). Due to the poor quality of the cells, the

possibility of further cytogenetic abnormalities could not be ruled out. MLL PTD

RQ-PCR indicated definitive relapse with a $\Delta\Delta$Cp MRD higher than that found at

diagnosis at 165%. The next sample followed in 2 months, after the patient had

undergone a third course of therapy. The WBC was 1.6×10^9cells/l and morphology

on the BM showed normal cellularity and 4% blasts; the patient had responded well to

treatment and was in complete remission. Immunophenotyping found no blasts with

the presentation phenotype and stated, "If there are any leukaemic blasts they are

beyond our limit of detection." TEL/*AML1* FISH found no evidence of trisomy 21 in

200 nuclei examined. In contrast, the RQ-PCR assay detected MLL PTD transcripts

with the $\Delta\Delta$Cp MRD at 4%, less than a 2 log reduction from the diagnostic level.

Unfortunately, no further samples were received for molecular analysis until the final

sample was received 3 months later; 8 months post diagnosis. The clinical details

were, "AML second remission. Still pancytopenic." All of the methods used for

MRD detection found evidence of relapse. The WBC was 1.2×10^9cells/l, with an

observation of circulating blasts, and an infiltration of myeloblasts was seen in the

analysis of the BM trephine sample. The majority of cells observed by

immunophenotyping were TdT positive blasts as seen at diagnosis. Cytogenetic

analysis of the BM found that 17/20 cells had an abnormal karyotype with trisomy 21,

as observed at first relapse and FISH analysis confirmed this result. Additionally, 4 of

these cells showed monosomy 7, an additional marker on the short arm of

chromosome 11 and a marker chromosome of unknown origin. The remaining 3/20

cells showed no cytogenetic abnormalities. These results confirmed relapse, along

with the possibility of further clonal evolution. The MLL PTD RQ-PCR confirmed the relapse with the $\Delta\Delta Cp$ MRD rising clearly above the diagnostic level to 269%. No further samples were received for molecular diagnostics. The patient went on to receive palliative treatment and died of septicaemia, liver failure and AML 11 months after diagnosis.

The greater sensitivity of the MLL PTD RQ-PCR and the stability of the MLL PTD as a marker compared to the other MRD methods available to the patient suggest it would be a clinically useful assay. The standard methods of MRD, flow cytometry and cytogenetic analysis, were not sensitive enough to detect MRD in the 5 month post diagnosis sample from patient UPN 173. Analysis by immunophenotyping showed the phenotype changing from TdT +ve at diagnosis to TdT -ve and back again, while clonal evolution was observed in the cytogenetic analysis. Whether this was due to two separate cell lines populating the marrow at different time points or a phenotypic switch within one cell line was unclear. Clonal evolution was observed as the karyotype changed from cytogenetically normal to trisomy 21 at first relapse, and then progressed to trisomy 21, monosomy 7 with additional aberrations at second relapse. The disadvantage to using markers of MRD prone to change is that at relapse the new phenotype may not be detected or may be detected but not recognised as leukaemia. The stability of the MLL PTD marker was evident from detection of MRD at every time point tested, from diagnosis through morphological remission to relapse. Increasing the frequency of molecular testing for MRD, *i.e.* following every course of chemotherapy, has the potential to predict impending relapse and to identify the patient as a poor responder who may be benefit from a different treatment strategy. Furthermore, the presence of the MLL PTD throughout patient monitoring

suggested that it may be necessary for maintaining the leukaemic phenotype and was possibly an early event in the transformation of the cell.

4.3   Discussion

The measurement of MRD has become increasing important in the treatment of patients with AML. Studies have proven that MRD levels are associated with the prognosis and clinical outcome of patients and can be used to guide the course of treatment. Unfortunately, as the majority of MRD markers detected by PCR are based on fusion genes, 63-79% of all AML patients are left with no sensitive method of following the course of disease after reaching morphological remission. The MLL PTD mutation occurs in approximately 6%-10% of CN patients and was tested as a potential marker for MRD using RQ-PCR.

The optimisation of the RQ-PCR assay was paramount because the MLL PTD transcript has been shown previously to be detectable in healthy donors and in MLL PTD negative AML patients (Chapter 3) (Schnittger *et al.* 2000), (Marcucci *et al.* 1998). The experiments conducted for the optimisation used MLL PTD negative AML samples and the two most commonly occurring MLL PTD mutations. Optimised concentrations of $MgCl_2$, cDNA template and the primer annealing temperature created a separation of nearly 3 logs between the positive and negative samples. With conditions optimised, healthy donor samples showed no specific amplification of MLL PTD, thereby eliminating the possibility of misinterpreting a healthy normal as a false positive.

The effect of different expression levels at diagnosis has been studied by a number of groups with contrasting conclusions; diagnostic levels were shown to have a significant effect on prognosis for the fusion genes, RUNX1-*RUNX1T1*, PML/RARA and CBFB/*MYH11*, (Leroy *et al.* 2005; Schnittger *et al.* 2003) and no correlation in a separate study on PML/RARA in APL (Gallagher *et al.* 2003). The diagnostic range of expression level for the MLL PTD was reported as being >2log, with the median value approximately 3 log above the level detected in both MLL PTD negative AML and healthy donors (Weisser *et al.* 2005b). In this study the diagnostic range was narrower, all samples were within 1.2 logs, and the donor samples showed no amplification. The discrepancy in the diagnostic level may be due to the lower number of samples used, or to the relatively narrow range of blast cell levels in the samples, (all within 50-63%, apart from one MDS sample at 10%). The Weisser (2005) study used the same amplification primers as the current study in combination with hybridisation probes to demonstrate the specific amplification of the MLL PTD transcript, whereas this study used Syber Green to identify the amplification of any double stranded product independent of specificity. The ability to visualise samples that produced non-specific amplification provided additional information, unavailable if using specific hybridisation probes, used to exclude them from further analysis. This may explain the negative results for the healthy donors in this study, since false positives are avoided. Another possibility is that the difference was due to the higher stringency for amplification in this study compared to the Weisser (2005) study, arising from the higher primer annealing temperature, 69°C vs. 64°C, and the smaller concentration of cDNA template. Evidence of this was the increase in the Cp of the MLL PTD positive samples as annealing temperature and template volume were increased (Table 4.3 and 4.4) and the increase in the Cp difference in positive and

negative samples when annealing temperatures were increased (Table 4.5). The result would be a lower overall sensitivity for the RQ-PCR assay, offset by the fact that donor samples were not amplified. The median difference of just over 3 log between MLL PTD positive samples and healthy donor samples reported in the Weisser (2005) study adds support to this theory, as that would place the transcript level of donor samples just below the sensitivity of the present assay.

Despite the alternative strategies of detection and the resulting differences, the sensitivity range for both assays are broadly similar with the sensitivity of this study at 2-3 log and a 2-4 log sensitivity, dependent upon diagnostic levels, reported by Weisser (2005). Likewise, samples presenting with a higher diagnostic ΔCp_K level would also be expected to have a higher sensitivity as previously outlined (Figure 4.10). The stability of the MLL PTD was demonstrated by the detection of the marker in 4 paired samples tested at diagnosis and at relapse. The stability of MLL PTD, even in cases with clonal evolution, may indicate that the marker is an early event in leukaemogenesis. However, as the mutation is found in healthy donors as well as the progenitor sub-fraction of cord blood samples the likelihood that it is independently leukaemogenic is low. MLL PTD positive cells may need further mutations or an increase in transcription level before leukaemic progression begins (Basecke et al. 2006b). All of the relapse samples had substantial levels of MLL PTD transcript at relapse, within 1 log of the diagnostic level. Interestingly, three of the four relapse samples relapsed at a higher level than that detected at diagnosis; this has previously been reported for both NPM1 and WT-1 (Barragan et al. 2008; Tamaki et al. 1996). The data shown here demonstrate that the sensitivity, stability and specificity of the assay make it a suitable candidate for further evaluation as a method for measuring MRD.

The MLL PTD RQ-PCR measurements calculated using the $\Delta\Delta Cp$ MRD formula showed correlation with the "traditional" MRD methods down to the level of morphological remission. Only one of the patients with follow up data available had a cytogenetic abnormality available for use as a MRD marker, the remainder were monitored using morphology and occasionally an aberrant immunophenotype, however, none of these phenotypes were exceptionally sensitive. No correlation was shown in four patients where standard methods had declared the patient to be in remission, yet the higher sensitivity achieved with the RQ-PCR demonstrated lingering disease in these patients. Overall, the data show that the RQ-PCR assay for detection of MLL PTD is capable of measuring MRD in patients harbouring the MLL PTD and may offer greater clinical relevance due to the higher level of sensitivity compared to other available methods.

Leukaemia-associated aberrant phenotypes (LAIPs) are being used more frequently for MRD because they are useful in a large proportion of AML patients, especially those that are CN. In this study, the LAIPs identified were not as sensitive as the MLL PTD RQ-PCR, but this may not always be the case as LAIP MRD analysis has a reported sensitivity of up to 4 logs (Campana & Coustan-Smith 1999; Freeman, Jovanovic, & Grimwade 2008). LAIPs have been used to detect MRD in MLL PTD positive patients and were able to detect MRD in patients in morphological remission, however the sensitivity of the assay was not determined (Munoz *et al.* 2003). In contrast to RQ-PCR analysis, which measures the level of transcript expression from an individual gene, LAIP analysis is a measurement of the number of leukaemic cells present in a sample. Approximately 25% of LAIPs present at diagnosis may not be found at relapse and studies have shown variation in the intensity of the antigen at relapse (Oelschlagel *et al.* 2000; Voskova *et al.* 2004). Due

4-221

to the different information provided by LAIPs and the possibility of marker instability, AML patients may be best served by a combination of the two techniques presented in a multidisciplinary report.

None of the patients' follow up samples reached molecular remission, however the principle was demonstrated clearly in the sensitivity test where the 0.1% dilution was negative. Further studies are needed to evaluate whether the achievement of molecular remission is significant to the long term disease free survival of MLL PTD positive patients. The data from previous studies are not conclusive. Several reports have demonstrated that long term complete remission was obtainable despite PCR positivity in the fusion genes RUNX1-*RUNX1T1*, PML/RARA and CBFB/*MYH11* (Gallagher *et al.* 2003; Krauter *et al.* 2003; Schnittger *et al.* 2003). It has been proposed that immunological intervention may able to hold low levels of LSCs in balance, preventing proliferation that would lead to the patient's eventual relapse (Grimwade & Lo 2002). This in possible combination with the concept that the translocation itself is not enough to cause leukaemia, could explain positive MRD results for patients in long term remission (Grimwade & Lo 2002). On the other hand, it was also reported that molecular remission was a requirement for long-term complete remission in CBFB/*MYH11* and PML/RARA (Chillon *et al.* 2004; Martin *et al.* 2000). The most likely explanation for the conflicting findings is that all groups are correct within the context of their respective tests, but the inconsistencies are a result of different sensitivities between the assays.

An improvement in the sensitivity of the assay may be possible by an adaptation of the hybridisation probe methodology used in the Weisser (Weisser *et al.* 2005b). The difference in amplification stringency due to the higher primer annealing temperature may provide a greater separation between the MLL PTD positive samples

and the healthy donor samples. This would be helpful in increasing the range of sensitivity to include samples with low initial diagnostic levels, which currently have a sensitivity of 2 logs. While this would give the advantage of making the assay more objective by removing the requirement for Tm analysis, it would also remove the benefit of being able to identify the amplification of non-specific products thereby preventing false positives. Another disadvantage is the considerable cost of synthesising hybridisation probes. While the current sensitivity is less than might be considered ideal for a molecular MRD marker, the patient cohort in question is populated with individuals with normal karyotype who often lack any markers at all beyond BM morphology so this assay provided additional useful information.

The importance of MRD monitoring in predicting the clinical outcome in AML has meant an increase in the study of mutations that occur preferentially in CN AMLs as potential markers of MRD. Cases of the MLL PTD and the *FLT3* ITD occurring in the same patient have been shown in Chapter 3, as well as in previously published reports (Libura *et al.* 2003; Steudel *et al.* 2003). *FLT3* ITD is often used as an MRD marker because of its high frequency of 40% in AMLs with a normal karyotype and 25% in AMLs (Schnittger *et al.* 2002). However, the instability of FLT ITD at relapse may lead to false negative results. In addition, unless patient and mutation specific primers are used, the sensitivity remains 2-3 log, so this MLL PTD assay may be an alternative method of MRD detection in patients with both mutations (Schnittger *et al.* 2004).

Clear evidence supporting the early prediction of relapse was not possible due to the paucity of follow up samples available. However, the remission samples showed the presence of MLL PTD transcripts in patients that later relapsed suggesting that a higher frequency of sampling time points may allow such predictions. The

4-223

ability to predict early relapse in APL coupled with an early treatment response in APL resulted in an improved clinical outcome (Diverio *et al.* 1998; Lo *et al.* 1999b). Additional studies are needed to see if this would prove to be true with MLL PTD patients.

The previously reported fraction of CN patients undergoing clonal evolution is substantially lower than that of karyotypically normal patients with the MLL PTD mutation in this study (Estey *et al.* 1995; Kern *et al.* 2002). Although *FLT3* ITD and MLL PTD status was determined in a previous study of karyotypic instability, only four MLL PTD patients were found and the results were given in terms of the two mutations together (Kern *et al.* 2002). There is no previous data published on the observation of higher clonal evolution in MLL PTD patients. The suggestion that clonal evolution increased the risk of acquiring a mutation that confers resistance to leukaemia chemotherapy has been made previously (Kern *et al.* 2002). This hypothesis was supported by the poor response of the patients in this study and warrants further studies into the high level of clonal instability in MLL PTD patients.

4.4 Conclusion

In conclusion, the hypothesis that the expression level of the *MLL* PTD as determined by RQ-PCR can be used as a marker for MRD was supported by the results shown in this study. The RQ-PCR would be beneficial to the 6%-10% of CN AML patients who present with the *MLL* PTD mutation and for whom there is often no sensitive method of MRD available. The assay is potentially a beneficial tool, despite its relatively low sensitivity, because the refractoriness and short duration of remission seen in many *MLL* PTD positive patients would make monitoring the response to therapy and the prediction of relapse in poor responders clinically useful. As *MLL* PTD patients have a high risk of relapse, long term MRD monitoring could also be improved by implementing this assay for detection of impending relapse, however, it may be less useful in this group of patients due to the low sensitivity level. Early identification of poor responders may enable intervention through more targeted treatment or an allogeneic bone marrow transplant, resulting in an improved clinical outcome.

The data in this study show a higher than expected level of clonal evolution from diagnostically normal karyotypes and may be a factor in the high level of relapse seen in *MLL* PTD positive patients.

Although the results of the study suggest the ability to predict relapse and determine molecular remission, future studies using established time points for MRD analysis and a much larger cohort of patients are needed to confirm this. The inclusion of *MLL* PTD MRD testing in a major AML trial would allow threshold

levels, and the relapse kinetics, to be studied so that the important time points for MRD analysis could be determined. Early prediction of relapse or failure to reach a threshold level could be followed by studies to determine whether a change in treatment could influence the outcome of these patients.

5 Gene Expression Profiling of *MLL*

PTD

5.1   Introduction

The use of gene expression profiling (GEP) for investigating the pathobiology of AML is well established. GEP is used to discover new categories of AML, to develop predictors (*i.e* good versus poor outcome) and to understand the biological processes associated with the disease using the three respective applications: class discovery, class prediction and class comparison. The technique has advanced from the milestone study by Golub *et al.* (1999), where GEP demonstrated a gene signature that could distinguish ALL from AML samples, and is now used to distinguish smaller less well defined subgroups within the disease(Golub *et al.* 1999). AML subdivision has developed in step with technical advancement, from morphology to cytogenetics to molecular genetics; the capacity for global analysis of gene expression now allows previously unrecognised characteristics to be identified by unique gene signatures.

New subgroups have been identified from the large heterogeneous group of cytogenetically normal AML samples using the technique of class discovery, whereby similarities and differences between the profiles of AML samples are sought. Unsupervised hierarchical analysis, which is analysis performed without consideration of sample characteristics, found gene signatures coinciding with established cytogenetic groups and cytogenetically normal AML (CN AML). The cytogenetically normal samples were divided into two subsets with unique gene signatures suggestive of different methods of AML progression. Outcome prediction used 133 genes significant to the duration of patient survival and found the

designation of good or poor outcome in the CN AMLs correlated with the 2 gene signature groups (Chapter 1.4.2) (Bullinger *et al.* 2004).

GEP analysis has been used to determine the expression patterns of *MLL* chimeric fusion genes in paediatric and adult AML leading to a better understanding of the biological features of this subgroup. The group of leukaemias defined by chimeras at the 11q23 locus show unique gene signatures, which allow them to be segregated from other AML translocations (Kohlmann *et al.* 2005). When paediatric AML and ALL patient samples were combined, signatures derived from *MLL* translocations grouped the *MLL* leukaemias by translocation instead of lineage (Ross *et al.* 2004; Zangrando *et al.* 2009). This indicated that similar transcriptionally-linked processes, regardless of the cell lineage involved, could establish *MLL* leukaemogenesis. Genes commonly expressed in both lineages included homeobox A (*HOXA)* cluster genes, in addition to the *HOX* co-regulators, myeloid ecotropic viral integration site 1 homeobox (*MEIS1*) and pre B cell leukaemia transcription factor 3 (*PBX3*) (Bullinger *et al.* 2004; Kohlmann *et al.* 2005). However, unsupervised analysis of combined acute leukaemia showed the *MLL* translocations grouping by lineage (Kohlmann *et al.* 2005; Ross *et al.* 2004; Zangrando *et al.* 2009). Unsupervised analysis conducted exclusively on *MLL* chimeric samples clearly separated the AML and ALL lineages, yet failed to further segregate the cases into groups according to the fusion partner or, for AML cases, according to FAB classification (Kohlmann *et al.* 2005). Subsequent attempts to identify unique gene signatures and predict class according to the fusion partner failed. In summary, gene expression signatures derived from microarray technology increased the understanding of the relationship between *MLL* fusion genes within AML and between acute leukaemias.

The *MLL* PTD occurs primarily in AML and affects the same region of the MLL gene as the translocations yet no partner gene is involved. The overarching hypothesis of this chapter is that *MLL* PTD has a gene signature that is unique to other AMLs and can be used to add understanding to the aetiology of the disease.

The objectives of the chapter were:

1) to determine whether *MLL* PTD alters the overall gene signature from AML *MLL* PTD WT and can be used to predict for *MLL* PTD through GEP

2) to establish if molecular pathways or biological function correlations between the identified gene lists add insight to the pathogenesis of *MLL* PTD and can be used to discover potential therapeutic options

3) to validate the microarray data by correlation with RQ-PCR analysis

4) to ascertain if identified genes can be used as markers of minimal residual disease

5.2 Materials and Methods

5.2.1 Patients and Control Samples

In this study bone marrow aspirates from 181 samples were analysed; 162 were presentation samples from AML patients and 19 were obtained from transplant samples provided by healthy control subjects . All AML samples were from patients enrolled in MRC-UK AML clinical trial protocols and were processed at University Hospital Wales, Cardiff. Clinical characteristics of the patients are listed in Appendix A. RNA was isolated as described in Chapter 2 Materials and Methods (Section 2.4).

5.2.2 Affymetrix Microarray Analysis

The process of sample preparation, hybridisation and scanning of the Affymetrix GeneChips is covered in the Materials and Methods (Section 2.10). Affymetrix numerical cell (.cel) files were imported into Partek Genomics Suite (Partek Inc., St. Louis, MO, USA), a statistical program designed for microarray analysis, using the default normalisation parameters. Robust multi-chip averaging (RMA) was used to pre-process the perfect match probe data using background correction, quantile normalisation across all chips, $\log_2$ transformation and median polish summarisation. RMA normalised the background intensity of the perfect match probe values across each GeneChip and then normalised data across GeneChips with quantile normalisation. Two different GeneChips were used in this study, the HG-U133A and the HG-U133 Plus 2. The HG-U133A array is a human genome microarray containing >22,000 probe sets, analysing the expression level of

over 18,400 transcripts and variants that code for over 14,500 genes. The HG-U133 Plus 2 is an improved higher density GeneChip that evolved from the original HG-U133A, enabling the analysis of an additional 32,000 probe sets, 28,600 transcripts and 24,000 genes.

Analysis of variance (ANOVA) ($p<0.05$) was used to determine differentially expressed probe sets due to the *MLL* PTD with a false discovery rate (FDR) of less than 5%. The resulting list was filtered to exclude all sets that changed within a 2-fold change window in *MLL* PTD expressing cells when compared to AML *MLL* PTD WT (Figure 5.1).

To increase the statistical power, the sample number was increased by combining data from the common probe sets from the HG-U133A and the HG-U133 Plus 2 cohorts. To prevent batch effects created by the different GeneChips and protocols from obscuring the biological effects of the *MLL* PTD the Partek Batch Remover programme was used. For information on protocol differences between HG-U133A and HG-U133 Plus 2 see Chapter 2 Material and Methods (Section 2.10). The results of Partek Batch Remover are demonstrated in Figure 5.2, which shows the samples from both chips combining into one common cluster.

For the validation of microarray data the log_2 values of the genes of interest (GOIs) were plotted against the ΔCp values. The log_2 data for each gene of interest was retrieved from the combined sample group data set. When multiple transcripts identified a single gene, the transcripts with significant ($p<0.05$) and >2 fold expression levels up or down between MLL PTD and MLL WT were used and the mean taken. This ensured expression levels of variant genes, if present, were taken into account. All of the data sets showed Gaussian distribution so Pearson's correlation coefficient was used. GraphPad Prism version 3.02 (GraphPad Software

5-232

Inc., La Jolla, CA, USA) was used to perform the statistical analysis, and the null hypothesis was rejected when p<0.05.

5.2.3 Statistical Analysis for Determining Significant Probe Sets

The following strategy was applied to each of the GeneChip cohorts. To identify statistically significant changing genes due to the presence of the *MLL* PTD mutation, supervised differential expression analysis (an analysis that considers external factors) was performed using ANOVA followed by exclusion of probe sets that exhibited <2-fold change in expression level up or down (Figure 5.1). The list of probe sets was generated with a false discovery rate (FDR) of less than 5%. Principal component analysis (PCA) and hierarchical clustering was used to show how the increase in the stringency of the analysis caused a decrease in the number of transcripts and increasingly separated the *MLL* PTD samples from the *MLL* PTD WT samples. The 2-fold change in expression level was chosen to maximise the chance that the gene product had an appreciable affect via its function and was not exhibiting a normal level of variation. If an MRD marker is constitutively expressed in normal patients, it must be significantly up-regulated in patients with AML in order to be clinically useful. If it is not expressed highly enough above the normal background range the lack of sensitivity raises the possibility of false positives thereby limiting the clinical relevance. The sensitivity level for identification is dependent upon the MRD target and the technique used, RQ-PCR is up to 2 logs more sensitive than multiparameter flow cytometry (Liu Yin 2002b). However, the sensitivity required is determined by the clinical application of the assay, for example in ALL a sensitivity of 10^{-4} is required to detect low risk patients and 10^{-2} may be sufficient for detection

of high risk patients (Van, V *et al.* 2003b). Given that many *MLL* PTD patients will

be cytogenetically normal and have no other marker available, sensitivity greater than

that of morphology (approximately 1:20) would be an improvement(Liu Yin 2002b).
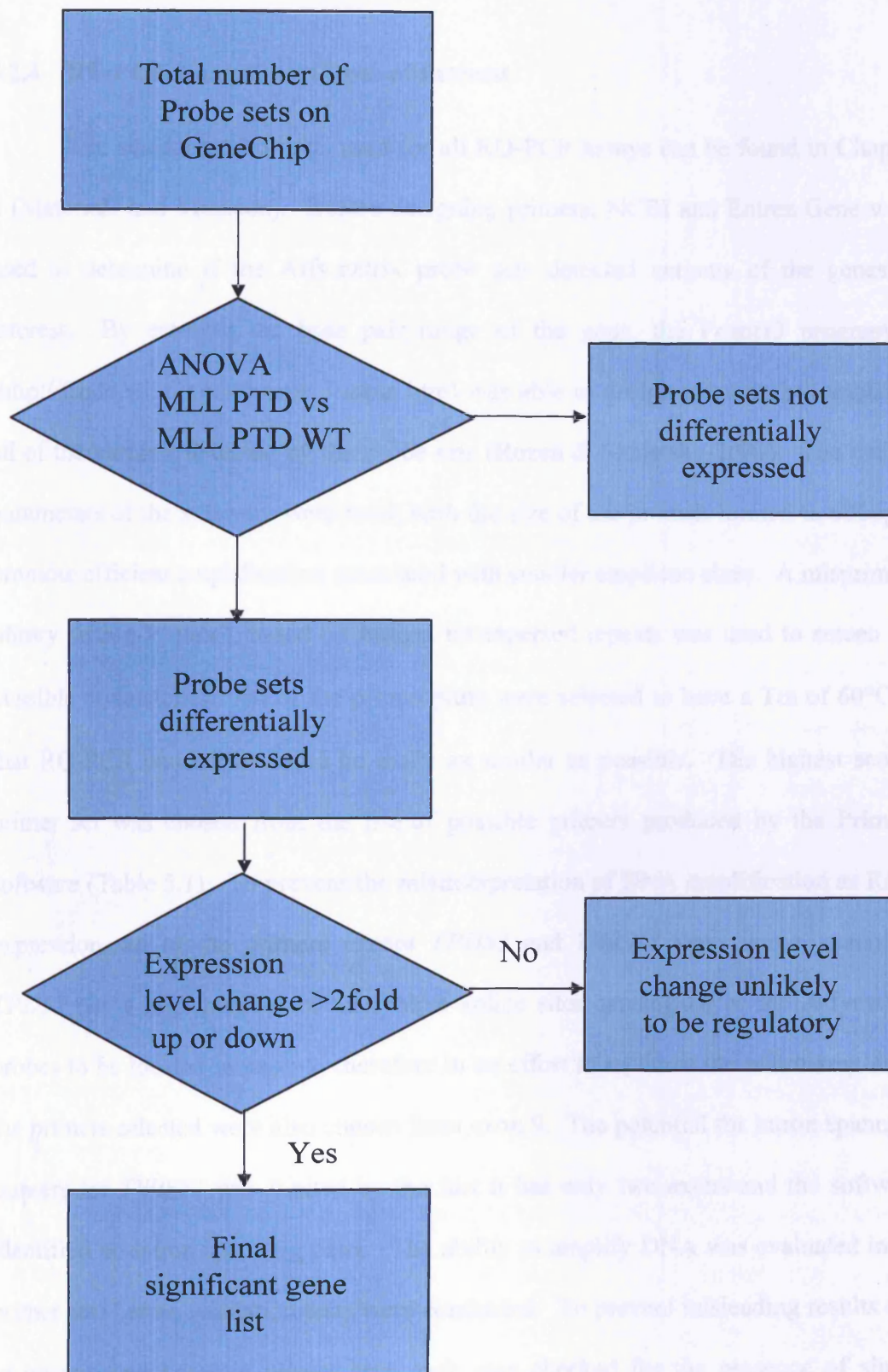
Figure 5.1 Data Analysis Flow Chart
Decision tree for analysing Affymetrix microarray data. Data are processed through ANOVA (MLL PTD vs. AML WT) (significance defined by p<0.05) and then expression level change to identify a significant gene list.

5.2.4 RT-PCR Analysis of Genes of Interest

The standard conditions used for all RQ-PCR assays can be found in Chapter 2 (Materials and Methods). Before designing primers, NCBI and Entrez Gene were used to determine if the Affymetrix probe sets detected variants of the genes of interest. By entering the base pair range of the gene, the Primer3 programme (http://frodo.wi.mit.edu/primer3/input.htm) was able to design primers that amplified all of the variants detected by the probe sets (Rozen & Skaletsky 2000). The default parameters of the software were used, with the size of the product limited to 600bp to promote efficient amplification associated with smaller amplicon sizes. A mispriming library within Primer3, based on human interspersed repeats was used to screen out possible mismatches. All of the primer pairs were selected to have a Tm of 60°C so that RQ-PCR conditions could be made as similar as possible. The highest scored primer set was chosen from the list of possible primers produced by the Primer3 software (Table 5.1). To prevent the misinterpretation of DNA amplification as RNA expression, all of the primers except *TPD52* and *TWIST1* were intron spanning. *TPD52* has a high number of alternative splice sites causing all of the Affymetrix probes to be located in exon 9, therefore in an effort to replicate the microarray data, the primers selected were also chosen from exon 9. The potential for intron spanning primers for *TWIST1* was limited by the fact it has only two exons and the software identified no intron spanning pairs. The ability to amplify DNA was evaluated in all primer sets before validation tests were conducted. To prevent misleading results due to mismatched bases in primer sets, each was checked for the presence of single nucleotide repeats (SNPs) using the web based software SNPCheck

http://ngrl.manchester.ac.uk/ SNPCheckV2/snpcheck.htm). All primers were found to be free of SNPs.

GenBank Acc. No.	Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')	Amplicon Length (bp)
NM_001233.3	CAV2	acgactcctacagccaccac	cgtcctacgctcgtacacaa	402
NM_078626.2	CDKN2C	tgcacaaaatggatttggaa	gggcaggttcccttcattat	223
NM_014608.2	CYFIP1	atgcatgtggacgagtgtgt	cagaaatccagcacagcaaa	184
NM_006209.3	ENPP2	caccagctgtctggatttca	gtccaggctggtgagatgtt	196
NM_001025253.1	TPD52	ttgaattcggctgcaaat	tcatggcaatgcatgttg	158
NM_000474	TWIST1	tccgcagtcttacgaggag	ggaggacctggtagaggaa	196
NM_007313.2	ABL	cccaaccttttcgttgcactgt	cggctctcggaggagacgtaga	386

Table 5.1 Primers for Validation RQ-PCR
The forward and reverse primers sequences for each gene selected for validation. The GenBank accession number and amplicon length are also provided. Acc: accession; no: number; bp: basepairs.

The specificity of the selected PCR primers to the genes of interest and their variants was confirmed by testing each primer with the NCBI Blast programme (http://blast.ncbi.nlm.nih.gov/Blast.cgi)(Altschul *et al.* 1990). Every primer returned the genes and gene variants expected and all amplicons entered produced the base pair size predicted by Primer 3. *ABL* was chosen for use as the control gene for the RQ-PCR validation experiments and is shown as the seventh gene in Table 5.1. The *ABL* primer sequences chosen were used to amplify *ABL* as the control gene for the quantification of *BCR-ABL* in the estimation of disease in CML patients (Emig *et al.* 1999).

To prevent batch variation affecting the results of RQ-PCR analysis, the total amount of cDNA needed for all assays was calculated and a single batch of each cDNA was produced. Due to the limited capacity of the light cycler (32 capillaries), the 28 samples used for gene validation of each GOI were assayed in groups of 14 with a positive and negative control, equalling 32. This enabled all cases to be

analysed in two assays. Every assay for one gene, *i.e.* 2 sets of 32 capillaries each, was assayed consecutively using the same master mix.

The RQ-PCR data showed a Gaussian distribution therefore ANOVA was used with the Tukey method applied used as a post test to determine significant differences between the means of the groups.

5.3   Results

5.3.1   Identification and Analysis of Significant Gene Lists

5.3.1.a   Patient and sample characteristics

GEP was used to identify significant genes associated with the *MLL* PTD for their potential as MRD markers, identifiers of pathways for therapeutic intervention and predictors of *MLL* PTD positive AML. The AML subtypes in the 162 bone marrows analysed included normal karyotype, t(8;21), t(15;17), inv(16), *MLL* chimeric fusion genes and complex cytogenetic karyotypes. The clinical characteristics of the seven *MLL* PTD positive samples analysed are given in Table 5.2. The cytogenetic karyotypes represented were varied with the expected normal karyotypes (3/7), a trisomy 11 (1/7) previously associated with the *MLL* PTD (Caligiuri *et al.* 1996; Caligiuri *et al.* 1997), a complex karyotype (1/7), a monosomy 7 (1/7) and a surprising co-occurrence with t(8;21) (1/7) which is an infrequent event (Ishikawa *et al.* 2009; Schnittger *et al.* 2000). The resulting prognoses based on cytogenetic analysis were as follows: one favourable, four intermediate and two adverse (Grimwade *et al.* 1998). Three of the samples had co-occurrence of the *MLL* PTD and *FLT3* mutations, two with *FLT3* ITD and one with *FLT3* TKD. All healthy control samples were wild type for the *MLL* PTD.

5-238

UPN	Sex	Age	WBC (x109/L)	BM Blast (%)	FAB subtype	Cytogenetics	Cytogenetics Prognosis	MLL PTD	FLT3 Mutant	FLT3 ITD/TKD
282	F	65	2.6	63	M2	-7	Adverse	e10/e3+e11/e3	WT	WT
296	M	76	46.2	84	M1	Normal	Standard	e9/e3	Mutant	ITD
370	M	60	1	87	M1	Complex	Adverse	e9/e3	WT	WT
320	F	41	79.9	71	M2	t(8;21)	Favourable	e9/e3	WT	WT
41	M	64	60	63	M2	Normal	Standard	e10/e3+e11/e3	WT	WT
262	F	72	125	90	M5	Normal	Standard	e9/e3	Mutant	ITD
301	M	71	50.5	92	M1	+11	Standard	e9/e3	Mutant	TKD

Table 5.2 Clinical and Molecular Characteristics of *MLL* PTD Positive Patients.
Characterisation of *MLL* PTD positive samples in relation to different clinical variables. UPN: unique patient number; M: male; F: female; BM: bone marrow; WBC; white blood count; FAB , French-American-British; e: exon; PTD: partial tandem duplication; ITD: internal tandem duplication; TKD: tyrosine kinase domain; WT: wild type.

5.3.2 Patient Cohorts and Microarray Chips

To maximise the likelihood of identifying novel genes differentially expressed due to the *MLL* PTD mutation, data from both microarray GeneChips designed for transcriptional regulation studies were used (HG-U133A and the HG-U133 Plus 2). The Department of Haematology at University Hospital Wales has an extensive set of GEP data that can be analysed for the determination of gene expression in AML. There are 328 different AML GEPs analysed on the 3' expression arrays currently available, however RNA for some of the samples had been expended, thus preventing *MLL* PTD analysis. Data was collected from two Affymetrix GeneChip platforms including the human genome HG-U133A and the HG-U133 Plus 2.0. Initially, diagnostic AML samples were tested on the HG-U133A, however as the technology evolved and the feature size of the GeneChips increased, analysis was subsequently performed using the HG-U133 Plus 2 because of the greater gene coverage.

The 93 diagnostic AML samples and the 7 healthy samples assayed using the HG-U133A chips were tested for the presence of the *MLL* PTD (Table 5.3). Three of the samples were positive, giving a frequency of 3.2%. The *MLL* PTD PCR detected 4 positive samples from the 69 diagnostic AML samples assayed with the HG-U133 Plus 2 chip, giving a frequency of 5.8%. The combined frequency of *MLL* PTD from both data sets was 4.3%, which is consistent with the incidence seen in Chapter 1 and that reported in other studies (Bacher *et al.* 2005; Schnittger *et al.* 2000). There was no duplication of samples between the two cohorts.

Affymetrix GeneChip	Patient Groups		
	AML Samples	MLL PTD Positive	MLL PTD Frequency (%)
HG-U133A	93	3	3.2
HG-U133 Plus 2	69	4	5.8
Total	162	7	4.3

Table 5.3 Gene Chip Cohorts.
Total numbers of diagnostic samples tested on the different Affymetrix chips and the incidence of *MLL* PTD. The numbers of healthy donors tested on HG-U133A and HG-U133 Plus 2 were 7 and 12 respectively.

5.3.3   Cohort Merging for Increased Sample Number

Data from the HG-U133A and the HG-U133 Plus 2 cohorts were combined to increase the sample number and then analysed together. This set of samples was referred to as the "combined cohort". Partek Batch Remover™ was used in order to reveal the effects of the *MLL* PTD mutation on gene expression in isolation from any batch effects. Such batch effects can be large relative to the effect of the condition being studied and this may cause the experimental effects to be obscured. The Batch Remover effectively normalises any differences between batches while leaving the effects of the experimental condition unaffected. The PCA of the combined cohort prior to analysis is shown in Figure 5.2A. Although each set of samples clustered in a parallel plane, the two groups were separate from one another. The batch effect caused by the different types of GeneChips and protocols were removed using the software and the two sets of data merged into one congregation on PCA (Figure 5.2B). The success of the procedure was demonstrated by the tight clustering of *MLL* PTD and healthy control samples from the two different cohorts using ANOVA. This allowed subsequent analysis to be performed treating the data as one uniform group.
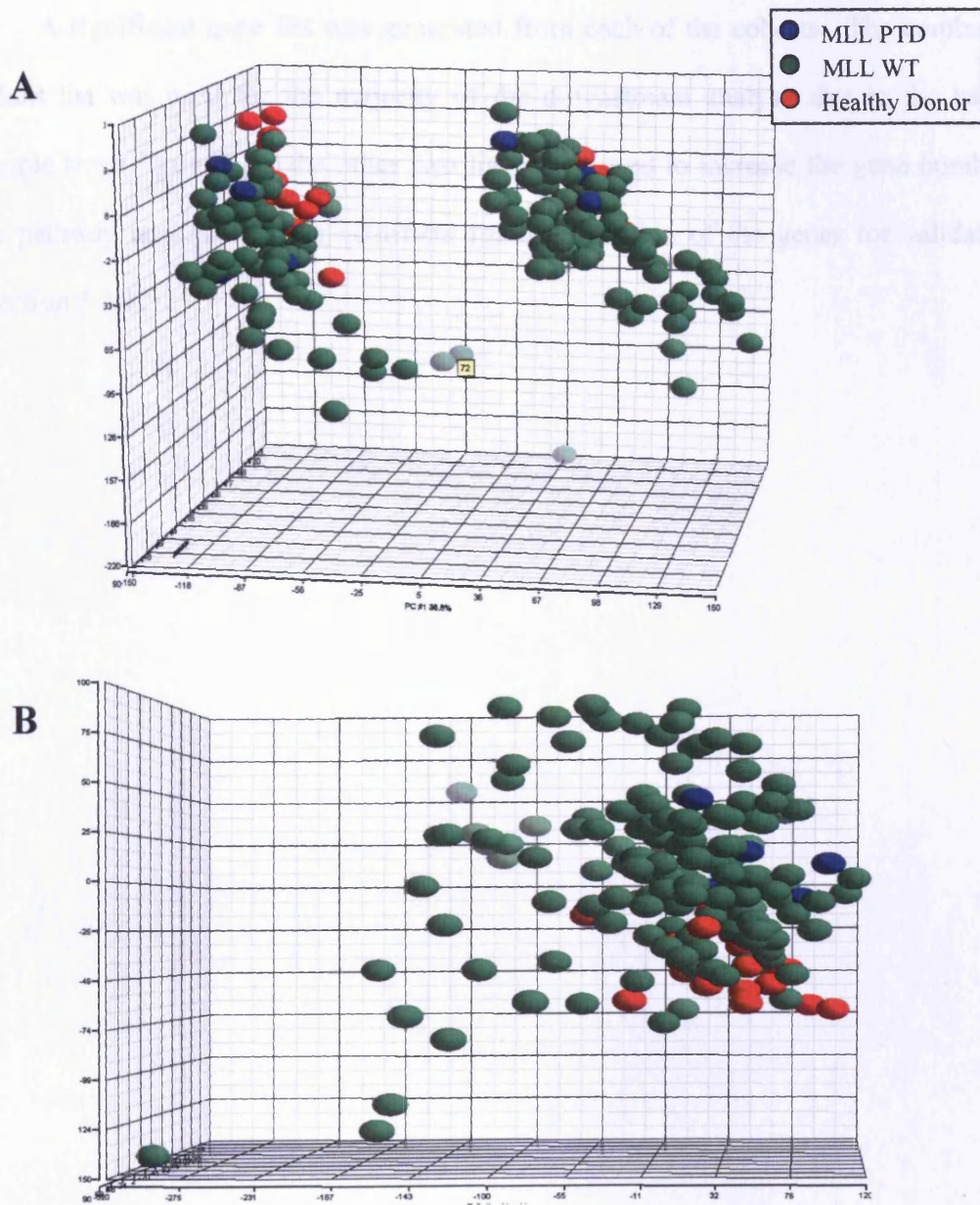
Figure 5.2 Principal Component Analysis of Batch Effect on the Combined Cohort.
The samples are projected in a three-dimensional space, plotted on the basis of the three principal components that best capture the variance of the data. Each case is represented by a coloured sphere. A) HG-U133A (right) and HG-U133 Plus 2 (left) are separated due to batch effects. B) After application of Batch Remover the two groups are merged into one common cluster. Samples were colour coded as indicated in legend. PCA, principal component analysis; pos, positive; wt, AML WT; Dwt, healthy control.

A significant gene list was generated from each of the cohorts. The combined cohort list was used for the majority of the downstream analysis due to the larger sample size (Figure 5.3); the other two lists were used to increase the gene numbers for pathway analysis and as qualifiers for the selection of the genes for validation (Section 5.3.12).
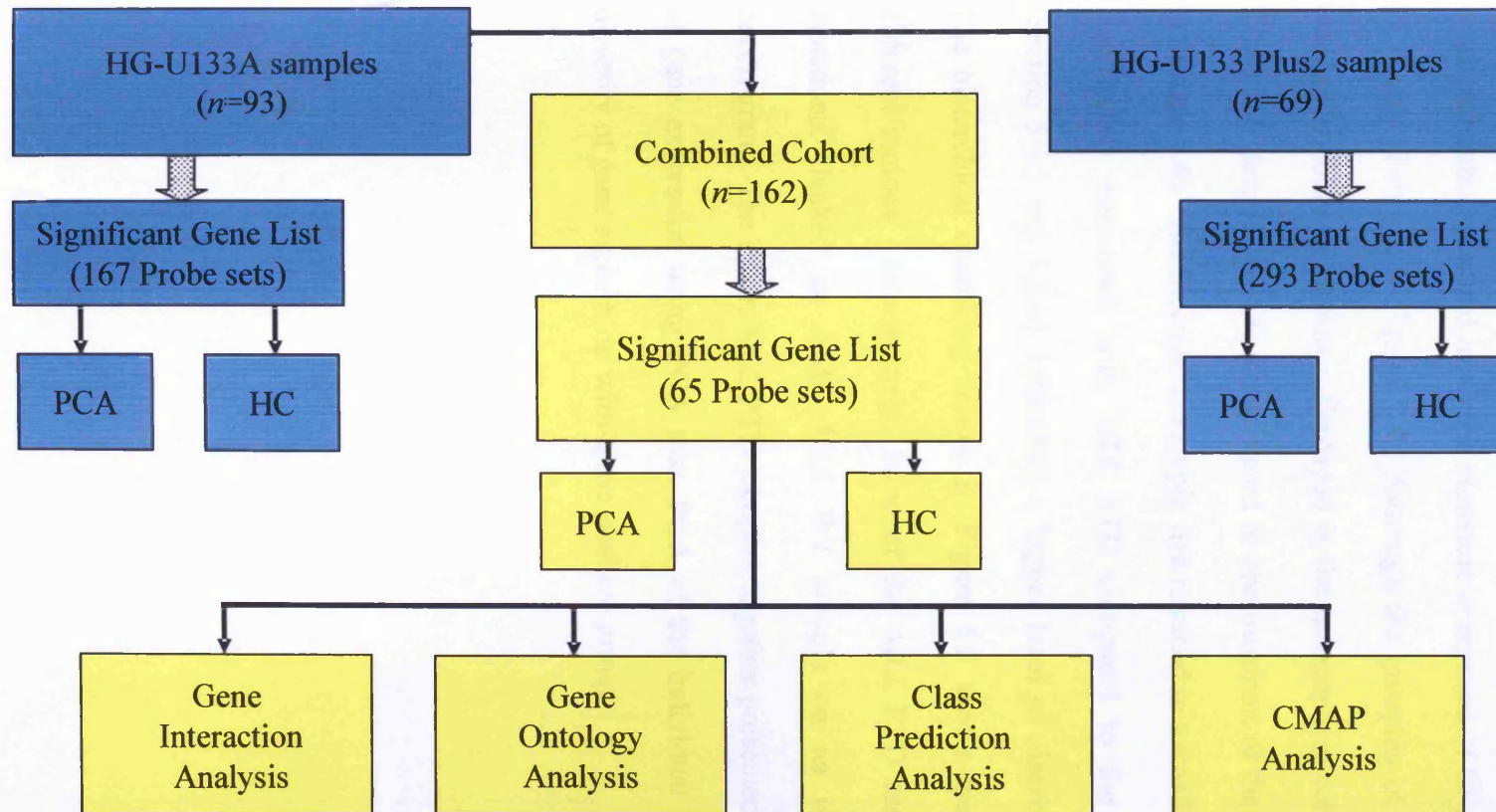
Figure 5.3 Combined Cohort Analysis Flow Chart Illustrating Combined Cohort Origin and Downstream Analysis
The flow chart demonstrates the sample composition of the combined cohort and illustrates the analyses performed on the different cohorts.
PCA: principal component analysis, HC: hierarchical cluster analysis; CMAP: connectivity map.

5.3.4 Significant Probe Sets Identified from Combined Cohort

The application of the above analysis strategy to the 162 samples in the combined cohort resulted in a gene signature composed of 65 probe sets; the resulting PCA map is shown in Figure 5.4. Although the grouping of the *MLL* PTD samples was not as cohesive as when displayed in the independent cohorts (Section 5.3.5 and 5.3.6), the samples are clearly grouped in one quadrant of the map (discussed further in Section 5.4). The increase in sample size resulted in a smaller number of probe sets significantly associated with *MLL* PTD compared to the individual GeneChips (Section 5.3.5 and 5.3.6), indicating a higher level of discrimination was obtained. The hierarchical clustering shown in Figure 5.5 shows the samples occurring in different regions. Interestingly, three of the *MLL* PTD samples, two with *FLT3* mutations, flanked an AML *MLL* WT sample on an isolated branch of the dendrogram. The seven *MLL* PTD samples together presented a more diverse picture of gene expression using PCA than PCA of the individual GeneChips, due to the diversity of gene expression within the positive group.
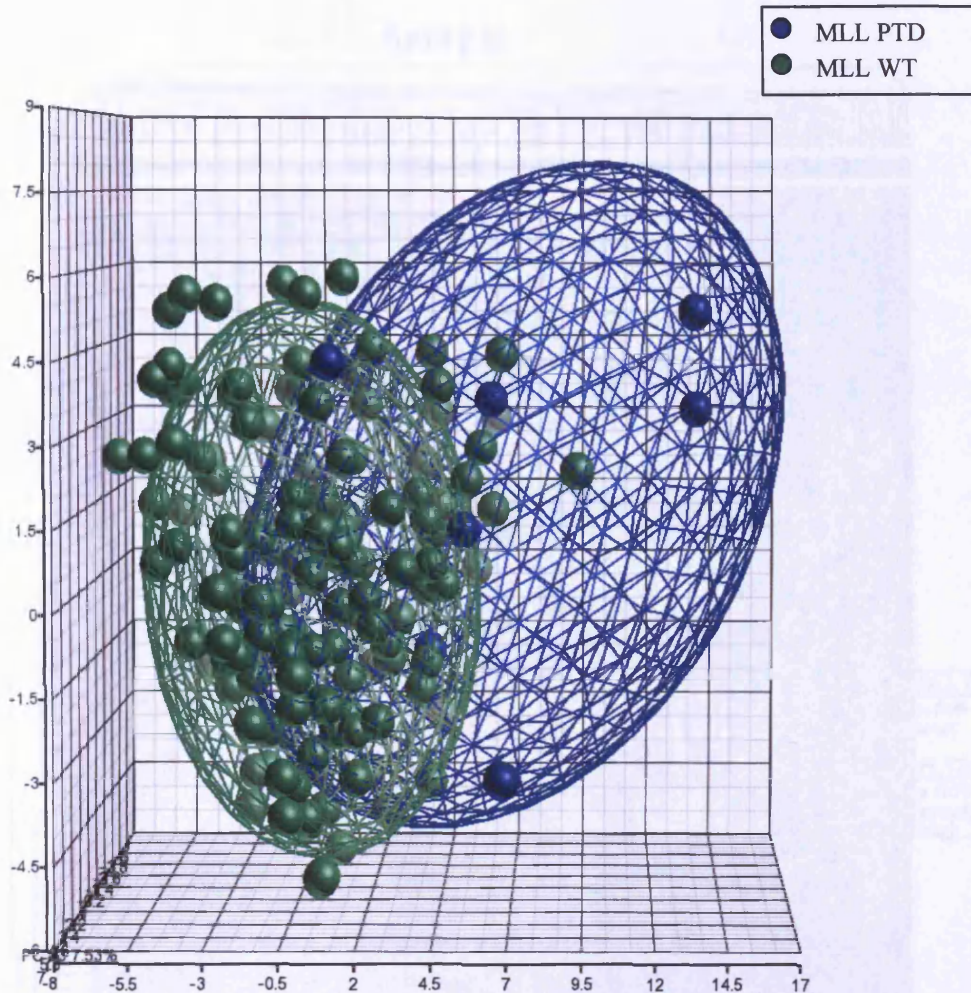
Figure 5.4 Principal Component Analysis of the Combined Cohort.
The samples are projected in a three-dimensional space, plotted on the basis of the three principal components that best capture the variance of the data. Each case is represented by a coloured sphere. The *MLL* PTD group (n=7) showed accurate separation from AML *MLL* WT (n=162) using 65 discriminating probe sets. Samples were colour coded as indicated in legend. PCA: principal component analysis; pos: positive; wt: AML WT.
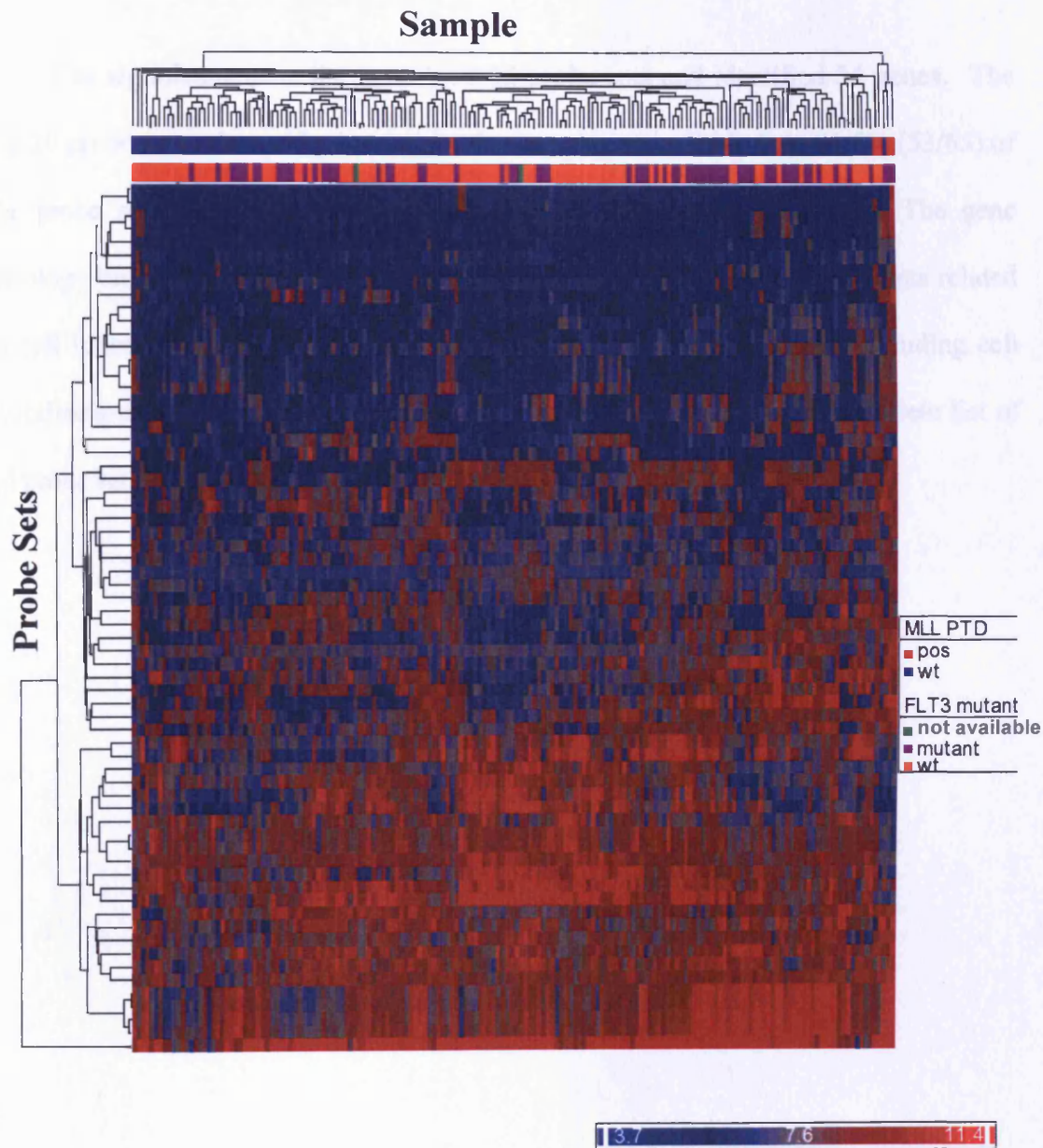
Figure 5.5 Hierarchical Cluster Analysis of Combined Cohort.
Hierarchical cluster analysis of AML (n=162) versus 65 *MLL* PTD discriminating probe sets. Probe sets are shown by rows and samples by columns. The normalised expression value of the probe sets are shown, with colour coding representing deviation from the mean, with the higher expression in red and lower expression in blue. Indicated below the samples is the *MLL* PTD and *FLT3* ITD classification. The green colourings for *FLT3* indicate that no data were available. *MLL* PTD abbreviations pos: *MLL* PTD; wt: AML *MLL* WT. *FLT3* mutant abbreviations, wt: *FLT3* wild type.

The significant gene list contained 65 probe sets and identified 54 genes. The top 20 probe sets selected by lowest p value are shown in Table 5.4. 81.5% (53/65) of the probe sets differentially regulated by *MLL* PTD are up-regulated. The gene ontology biological processes links genes with operations or molecular events related to cell function. The genes identified were associated with processes including cell mobilisation, proliferation as well as transcriptional regulation. The complete list of all probe sets is compiled in Appendix C.

Affymetrix ID	Common Name	Gene Title	GenBank Acc No.	Gene GO:BP	p-value(pos vs. wt)	FoldChange(pos vs. wt)
205048_s_at	PSPH	phosphoserine phosphatase	NM_004577	L-serine metabolic process	1.98E-06	4.7
211122_s_at	CXCL11	chemokine (C-X-C motif) ligand 11	NM_005409	chemotaxis	3.98E-06	2.7
210839_s_at	ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2	NM_001040092	phosphate metabolic process	7.61E-06	2.4
204751_x_at	DSC2	desmocollin 2	NM_004949	cell adhesion	8.03E-06	2.1
203324_s_at	CAV2	caveolin 2	NM_001233	negative regulation of endothelial cell proliferation	1.68E-05	3.2
219479_at	KDELC1	KDEL (Lys-Asp-Glu-Leu) containing 1	NM_024089	---	5.54E-05	2.3
203323_at	CAV2	caveolin 2	NM_001233	negative regulation of endothelial cell proliferation	7.44E-05	2.8
201689_s_at	TPD52	tumor protein D52	NM_001025252	anatomical structure morphogenesis	7.62E-05	2.5
209392_at	ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2	NM_001040092	phosphate metabolic process	9.47E-05	3.3
204779_s_at	HOXB7	homeobox B7	NM_004502	transcription	1.08E-04	2.1
210163_at	CXCL11	chemokine (C-X-C motif) ligand 11	NM_005409	chemotaxis	1.60E-04	2.0
204159_at	CDKN2C	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4	NM_001262	regulation of cyclin-dependent protein kinase activity	2.73E-04	2.6
213943_at	TWIST1	twist homolog 1 (Drosophila)	NM_000474	negative regulation of transcription from RNA polymerase II promote	2.91E-04	2.3
212097_at	CAV1	caveolin 1, caveolae protein, 22kDa	NM_001753	MAPKKK cascade	5.99E-04	2.6
201688_s_at	TPD52	tumor protein D52	NM_001025252	anatomical structure morphogenesis	6.91E-04	2.1
208923_at	CYFIP1	cytoplasmic FMR1 interacting protein 1	NM_001033028	multicellular organismal development	7.43E-04	2.1
201690_s_at	TPD52	tumor protein D52	NM_001025252	anatomical structure morphogenesis	1.25E-03	2.4
215017_s_at	FNBP1L	formin binding protein 1-like	NM_001024948	endocytosis	1.48E-03	2.1
202833_s_at	SERPINA1	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase	NM_000295	response to hypoxia	1.51E-03	-4.7
203761_at	SLA	Src-like-adaptor	NM_001045556	---	1.93E-03	-2.0

Table 5.4 Genes Differentially Regulated In Combined Cohort *MLL* PTD Samples

The top 20 dysregulated probe sets representing 15 genes are listed according to their significance (p value). Fold change is given with respect to AML *MLL* WT mean expression level. Acc: accession; no: number; GO:BP: Gene ontology biological process; pos: *MLL* PTD; wt: AML *MLL* WT.

5.3.5 Significant Probe Sets Identified from HG-U133A

A significant gene list from the HG-U133A GeneChip, differentiating *MLL* PTD from *MLL* PTD WT was identified by ANOVA and >2 fold difference in expression level for further analysis. The successive PCAs show how the variance from the decreased number of genes increasingly separated the *MLL* PTD samples from the *MLL* PTD WT and donor samples. In the sequence of PCA diagrams (Figure 5.6, A-C), the *MLL* PTD samples clearly become further separated from both the AML *MLL* WT and the healthy controls as the stringency of analysis is increased and the number of significant genes reduced. The ability of the 167 significant probe sets to segregate the *MLL* PTD samples demonstrated the uniqueness of their expression levels compared to the other groups. In Figure 5.7, hierarchical clustering shows the *MLL* PTD samples grouped together on the right side of the diagram in a small discrete set separated from the bulk of the other samples. The samples grouped together despite the fact that 2 of the 3 samples had a co-occurring *FLT3* mutation in addition to the *MLL* PTD. This may indicate that the *MLL* PTD expression profile was preserved despite the presence of the *FLT3* mutation or a similarity in the profiles of the two mutations.
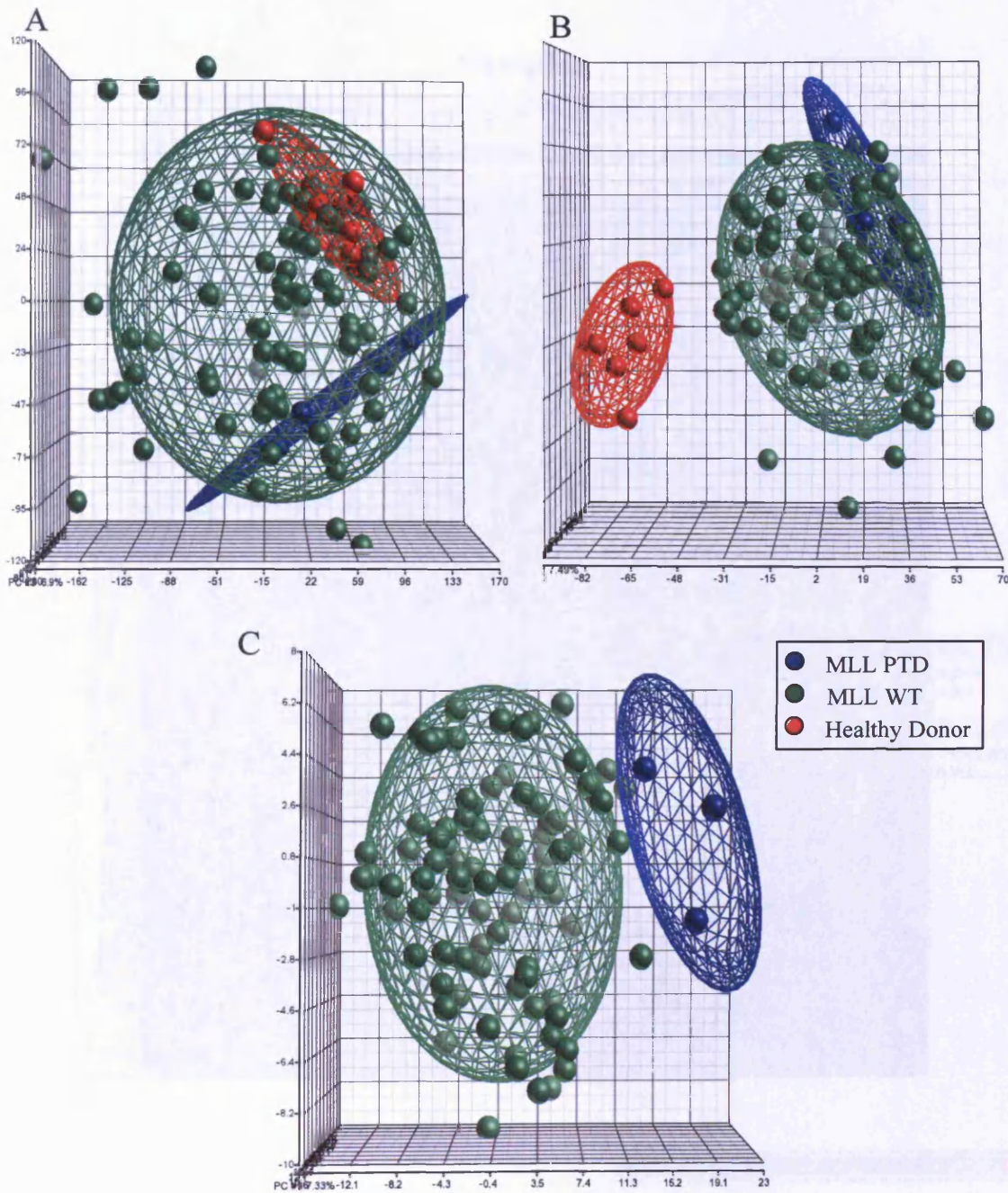
Figure 5.6 Principal Component Analysis of HG-U133A Samples.
The samples are projected in a three-dimensional space, plotted on the basis of the three principal components that best capture the variance of the data. Each case is represented by a coloured sphere. The *MLL* PTD group (n=3) show further separation from AML *MLL* WT (n=93) and healthy controls (n=7) as the number of probe sets is reduced through statistical analysis. A) PCA with compete gene list, probe sets=22,283, B) PCA with genes discriminating *MLL* PTD from AML WT probe set=4731, and C) PCA with genes discriminating *MLL* PTD positive from AML WT and greater than 2 fold expression, probe sets=167. Samples were colour coded as indicated in legend. PCA, principal component analysis; pos, positive; wt, AML WT; Dwt, healthy control.

5-251
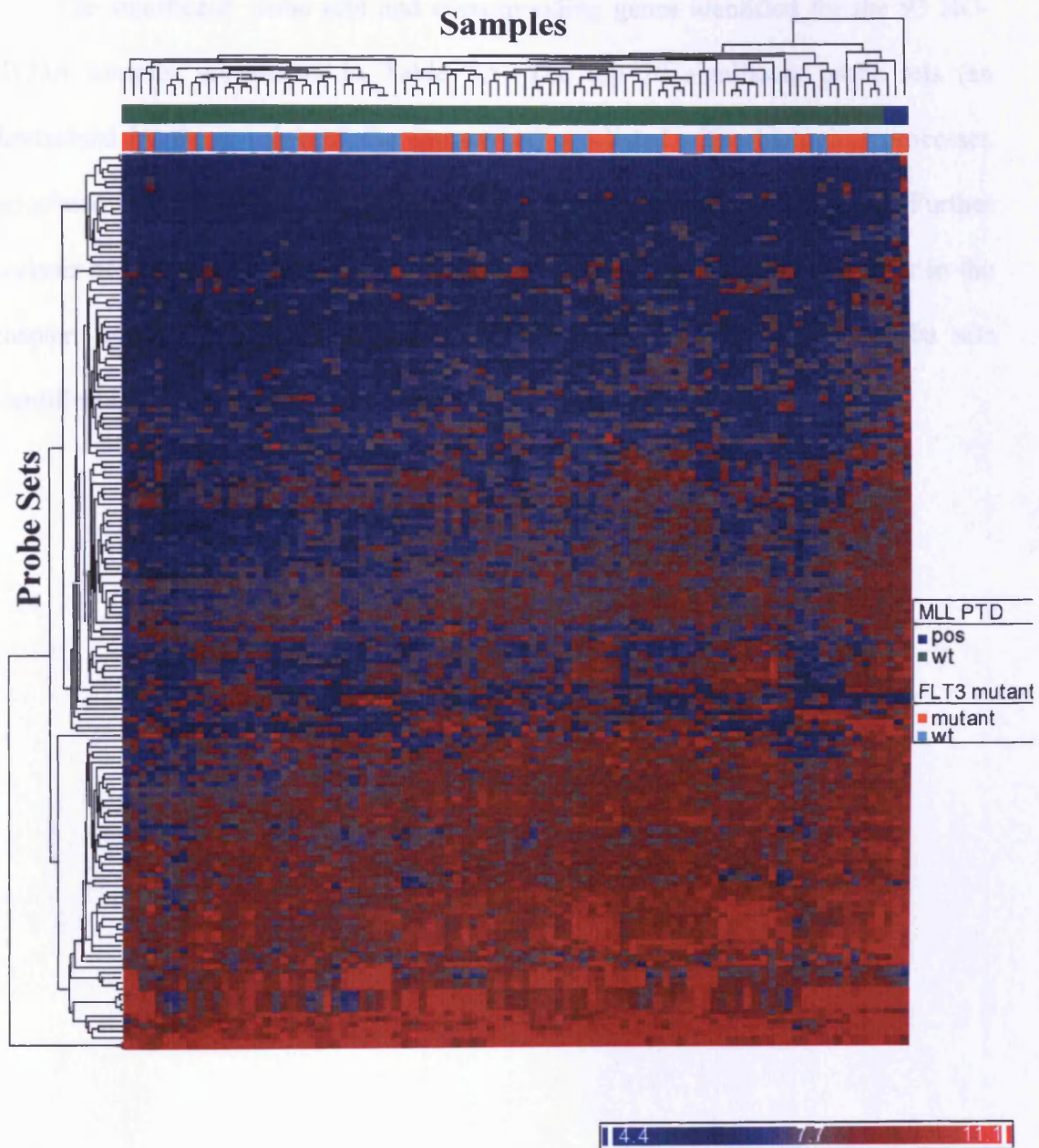
Figure 5.7 Hierarchical Cluster Analysis of HG-U133A Samples.
Hierarchical cluster analysis of AML (n=93) samples versus 167 *MLL* PTD
discriminating probe sets. Probe sets are shown by rows and samples by columns.
The normalised expression value of the transcripts are shown, with colour coding
representing deviation from the mean, with the higher expression in red and lower
expression in blue. Indicated below the samples is the *MLL* PTD and *FLT3* mutant
classification. *MLL* PTD abbreviations: pos, *MLL* PTD; wt, AML *MLL* WT. *FLT3*
mutant abbreviations, wt *FLT3* wild type.

The significant probe sets and corresponding genes identified for the 93 HG-U133A samples are shown in Table 5.5. The top 20 significant probe sets (as determined by the p value of the transcripts) are listed. The biological processes associated with the genes identified include transcription and mobilisation. Further analysis of the genes identified by the significant probe sets is addressed later in the chapter, Sections 5.3.7-9 and 5.12. Additional lists of the complete probe sets identified are compiled in Appendix C.

Affymetrix ID	Common Name	Gene Title	GenBank Acc No.	Gene GO:BP	p-value(pos vs. wt)	FoldChange(pos vs. wt)
211122_s_at	CXCL11	chemokine (C-X-C motif) ligand 11	NM_005409	chemotaxis	1.69E-08	4.6
210163_at	CXCL11	chemokine (C-X-C motif) ligand 11	NM_005409	chemotaxis	2.85E-08	3.9
203915_at	CXCL9	chemokine (C-X-C motif) ligand 9	NM_002416	chemotaxis defense respo	4.13E-08	3.5
201083_s_at	BCLAF1	BCL2-associated transcription factor 1	NM_001077440	transcription	1.18E-07	2.4
209392_at	ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2	NM_001040092	phosphate metabolic process	1.44E-07	8.1
210839_s_at	ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2	NM_001040092	phosphate metabolic process	1.45E-06	3.1
221530_s_at	BHLHE41	basic helix-loop-helix family, member e41	NM_030762	transcription	1.55E-06	2.2
216016_at	NLRP3	NLR family, pyrin domain containing 3	NM_001079821	apoptosis; induc	5.32E-06	2.1
204779_s_at	HOXB7	homeobox B7	NM_004502	transcription	9.68E-06	2.8
206500_s_at	C14orf106	chromosome 14 open reading frame 106	NM_018353	cell cycle; mito	1.60E-05	2.2
203305_at	F13A1	coagulation factor XIII, A1 polypeptide	NM_000129	blood coagulation; blood	2.32E-05	8.5
216973_s_at	HOXB7	homeobox B7	NM_004502	transcription	3.38E-05	2.4
204438_at	MRC1 /// MRC1L	mannose receptor, C type 1 /// mannose receptor, C type 1-like 1	NM_001009567	endocytosis; rec	3.99E-05	5.3
213943_at	TWIST1	twist homolog 1 (Drosophila)	NM_000474	negative regulation of transcription from RNA polymerase II promo	1.84E-04	2.5
201462_at	SCRN1	secernin 1	NM_001145513	exocytosis	2.12E-04	3.6
222154_s_at	SPATS2L	spermatogenesis associated, serine-rich 2-like	NM_001100422	---	2.91E-04	2.6
213734_at	WSB2	WD repeat and SOCS box-containing 2	NM_018639	intracellular signaling cascade	3.41E-04	2.4
202314_at	CYP51A1	cytochrome P450, family 51, subfamily A, polypeptide 1	NM_000786	steroid biosynthetic process	4.61E-04	2.4
201951_at	ALCAM	activated leukocyte cell adhesion molecule	NM_001627	cell adhesion	8.54E-04	2.8
203324_s_at	CAV2	caveolin 2	NM_001233	negative regulation of endothelial cell proliferation	9.84E-04	3.5

Table 5.5 Genes Differentially Regulated In HG-U133A *MLL* PTD Samples
The top 20 dysregulated probe sets representing 18 genes are listed according to their significance (p value). Fold change is given with respect to AML *MLL* WT mean expression level. Acc: accession; no: number; GO:BP: Gene ontology biological process; pos: *MLL* PTD; wt: AML *MLL* WT.

5.3.6 Significant Probe Sets Identified from HG-U133 Plus 2

The same analysis rationale was applied to the 69 samples analysed with the HG-U133 Plus 2 chip to identify a significant gene list associated with the *MLL* PTD. The results of the analysis reduced the number of probe sets from >54,000 to 293; the final results are shown on the PCA map in Figure 5.8. The isolated grouping of the *MLL* PTD samples showed that the 293 transcripts were able to discriminate these samples from the AML *MLL* PTD WTs. The result is similar to that observed with the greater number of samples in the HG-U133A experiment. When the samples were displayed using hierarchical clustering, the *MLL* PTD samples associate on one half of the dendrogram with three of the four located together on an independent branch (Figure 5.9). One of these three clustered samples had a co-occurring *FLT3* mutation yet still matched the expression pattern of the *MLL* PTD positives. The PCA and the hierarchical dendrogram demonstrated separation of *MLL* PTD from the *MLL* WT samples using the significant list of probe sets.
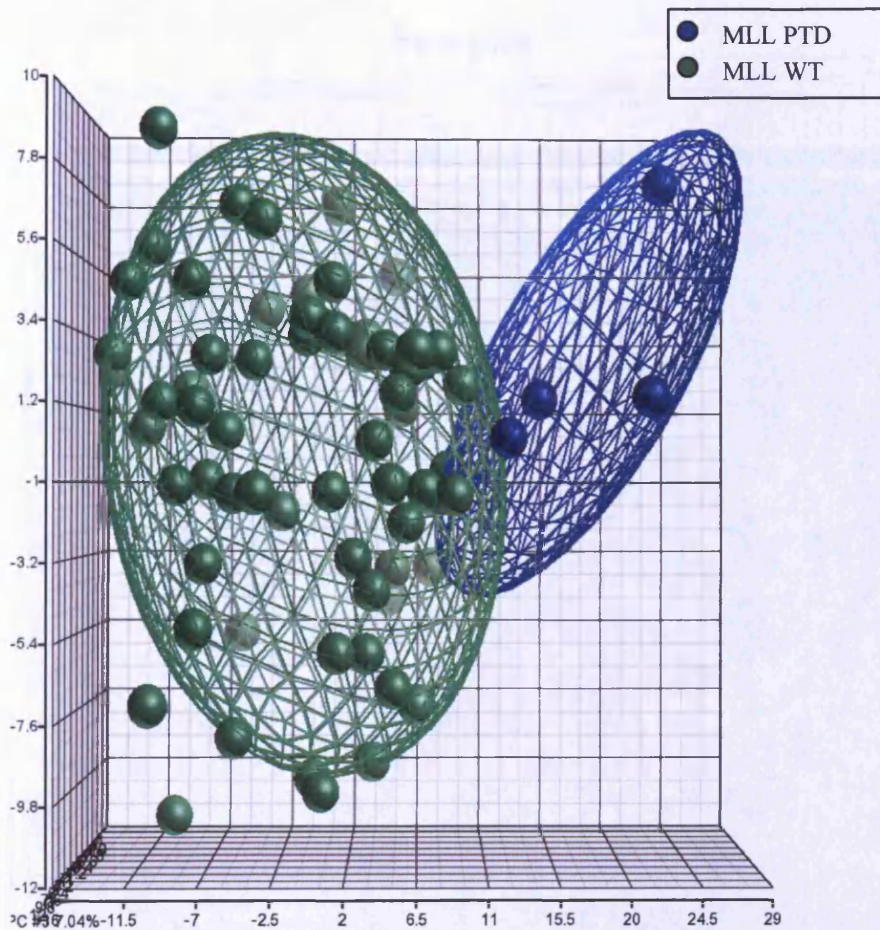
Figure 5.8 Principal Component Analysis of HG-U133 Plus 2 Samples.
The samples are projected in a three-dimensional space, plotted on the basis of the three priniciple components that best capture the variance of the data. Each case is represented by a coloured sphere. The *MLL* PTD group (n=4) showed accurate separation from AML *MLL* WT (n=69) using 293 discriminating probe sets. Samples were colour coded as indicated in legend. PCA: principal component analysis; pos:, positive; wt: AML WT.
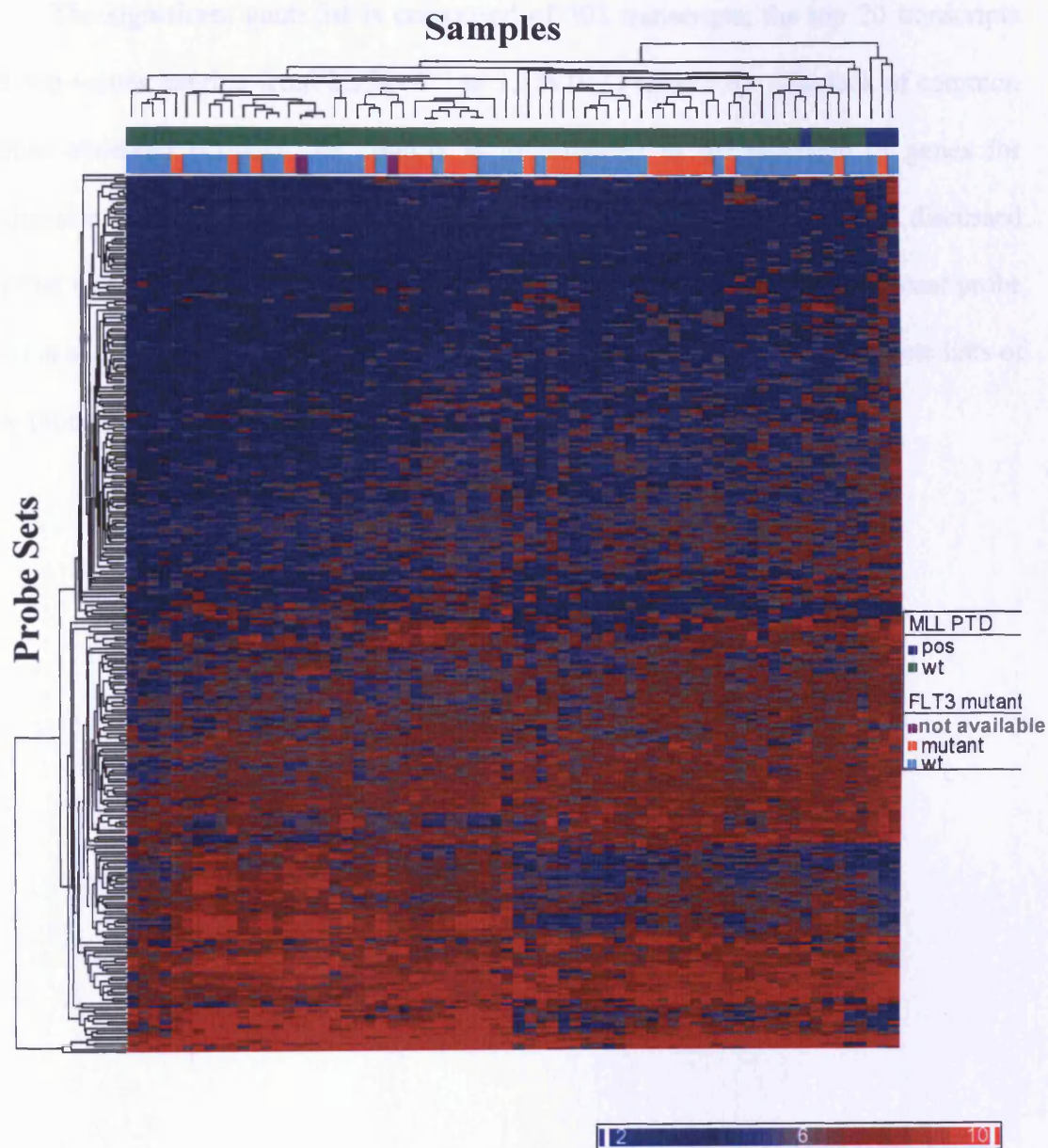
Figure 5.9 Hierarchical Cluster Analysis of HG-U133 Plus 2 Samples.
Hierarchical cluster analysis of AML (n=69) samples versus 293 *MLL* PTD discriminating transcripts. Transcripts are shown by rows and samples by columns. The normalised expression value of the transcripts are shown, with colour coding representing deviation from the mean, with the higher expression in red and lower expression in blue. Indicated below the samples is the *MLL* PTD and *FLT3* mutant classification. The purple colourings for *FLT3* indicate that no data were available. *MLL* PTD abbreviations: pos, *MLL* PTD; wt, AML *MLL* WT. *FLT3* mutant abbreviations, wt, *FLT3* wild type.

The significant gene list is composed of 293 transcripts; the top 20 transcripts have p-values ranging from 2.95×10^{-11} to 3.11×10^{-4} (Table 5.6). The lack of common genes observed between the cohorts is investigated in the selection of genes for validation where common genes are identified (Section Venn 5.3.12.a) and discussed further in Section 5.4. Further analysis of the genes identified by the significant probe sets is addressed later in the chapter (Sections 5.10-5.13 and 5.16). Complete lists of the probe sets are compiled in Appendix C.

Affymetrix ID	Common Name	Gene Title	GenBank Acc No.	Gene GO:BP	p-value(pos vs. wt)	FoldChange(pos vs. wt)
242054_s_at	---	Transcribed locus	---	---	2.92E-11	5.09
239552_at	FLJ14712	hypothetical protein FLJ14712	NM_001135924	---	8.19E-08	2.94
205048_s_at	PSPH	phosphoserine phosphatase	NM_004577	L-serine metabolic process	7.90E-06	15.19
1557280_s_at	ITSN1	intersectin 1 (SH3 domain protein)	---	---	1.40E-05	2.10
225619_at	SLAIN1	SLAIN motif family, member 1	NM_001040153	---	1.59E-05	4.30
206634_at	SIX3	SIX homeobox 3	NM_005413	eye development	3.36E-05	3.12
1555564_a_at	CFI	complement factor I	NM_000204	proteolysis	3.62E-05	2.17
205830_at	CLGN	calmegin	NM_001130675	protein folding	5.20E-05	3.03
222462_s_at	BACE1	beta-site APP-cleaving enzyme 1	NM_012104	proteolysis	5.43E-05	2.20
227036_at	RASAL2	RAS protein activator like 2	NM_004841	signal transduction	6.85E-05	2.02
204751_x_at	DSC2	desmocollin 2	NM_004949	cell adhesion	7.29E-05	2.55
206693_at	IL7	interleukin 7	NM_000880	T cell lineage commitment	8.13E-05	2.69
244740_at	MGC9913	Hypothetical protein MGC9913	---	---	8.69E-05	3.55
243918_at	---	Transcribed locus	---	---	8.81E-05	2.18
218898_at	FAM57A	family with sequence similarity 57, member	NM_024792	---	9.46E-05	2.17
220669_at	OTUD4	OTU domain containing 4	NM_001102653	---	1.66E-04	2.48
231094_s_at	MTHFD1L	methylenetetrahydrofolate dehydrogenase	NM_015440	one-carbon metabolic process	1.81E-04	2.09
222463_s_at	BACE1	beta-site APP-cleaving enzyme 1	NM_012104	proteolysis	2.67E-04	2.15
244741_s_at	MGC9913	Hypothetical protein MGC9913	---	---	2.67E-04	4.53
211429_s_at	SERPINA1	serpin peptidase inhibitor, clade A	NM_000295	response to hypoxia	3.11E-04	-11.12

Table 5.6 Genes Differentially Regulated in *MLL* PTD HG-U133 Plus 2 Samples
The top 20 dysregulated probe sets representing 19 genes are listed according to their significance (p value). Fold change is given with respect to AML *MLL* WT mean expression level. Acc: accession; no: number; GO:BP: Gene ontology biological process; pos: *MLL* PTD; wt: AML *MLL* WT.

5.3.7 Analysis of Significant Gene Interactions

A pathway algorithm was utilised to evaluate direct interactions between the fifty-four genes from the combined cohort. The genes were inspected for interactions within the list of 54 genes, between functionally related gene groups, as well as for co-operation within established canonical biological pathways to gain insight into the molecular mechanisms contributing to *MLL* PTD pathogenesis. To give a biological context to the gene list, the pathway function of GenesSpring GX version 10 (Agilent Technologies Inc., Santa Clara, CA, USA) was used to search for direct gene/gene product interactions from within the 54 genes identified. Ten genes from the list had direct interactions as defined by links identified through connections in either literature or pathway databases. (Figure 5.10). The two up-regulated caveolin genes, *CAVI* and *CAV2*, encode for scaffold membrane proteins that provide a compartment for signalling molecules (Scherer *et al.* 1996). Further analysis of *CAVI* and *CAV2* function can be found in the Discussion Section. The most interesting interaction haematologically is between the brain acid soluble protein 1 (*BASP1*) and AML linked gene Wilms Tumour 1 (*WT-1*). *BASP1* protein acts as a co-suppressor of *WT-1* through interaction with the suppression domain on the N-terminus of the *WT-1* gene, therefore the down regulation of *BASP1* observed in *MLL* PTD may be associated with the *WT-1* over expression (Carpenter *et al.* 2004). The six contiguous proteins did not have strong connections with oncology, although CD59 antigen and tissue factor pathway inhibitor (*TFPI*) were associated with haematology. The down-regulated gene C5a anaphlylatoxin chemotactic receptor (C5Ar1), is a neutrophil chemoattractant that transduces signals through intracellular GTP-binding proteins. Deficiency of C5Ar1 has been linked to airway hyper-responsiveness in mouse

asthma models which suggests the promotion of an anti-inflammatory phenotype

(Karp *et al.* 2000). The significance of the dysregulation of this gene to AML is

uncertain.
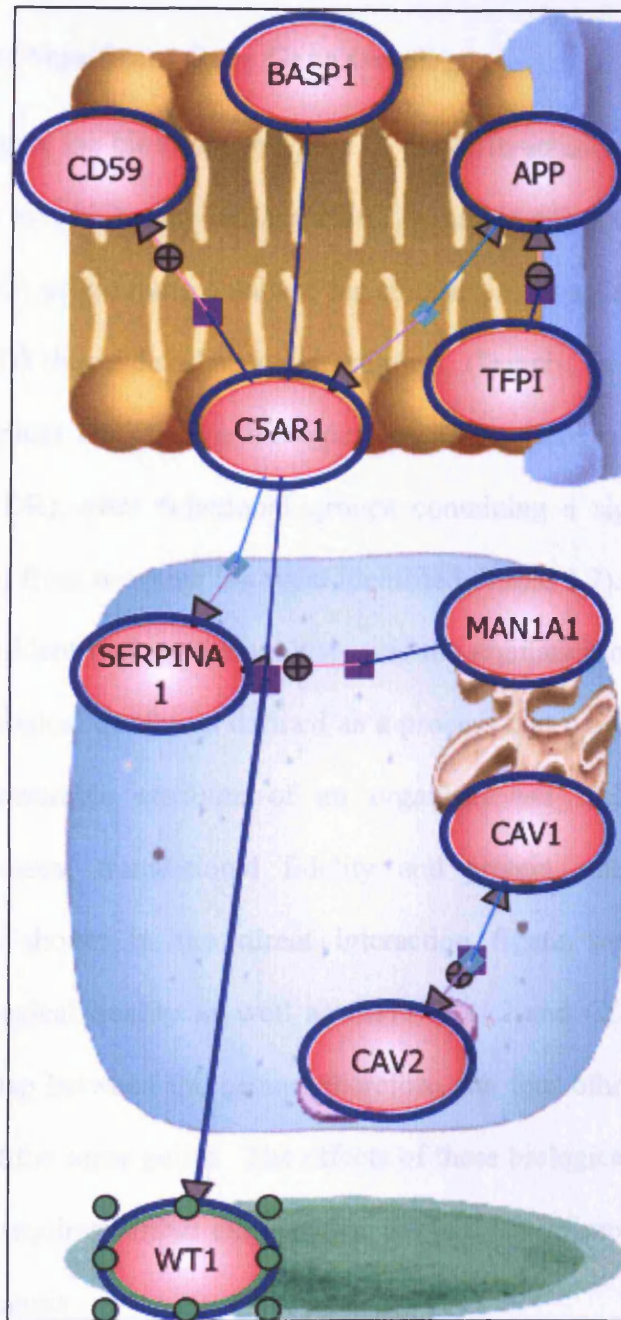
Figure 5.10 Directly Interacting Proteins from Combined Cohort
The genes/gene products are displayed as nodes and the relationships between them as
edges. The 10 genes have direct interaction with another gene from the combined
cohort gene list.

5-262

5.3.8 Analysis of Significant Gene Gene Ontology

To investigate the biological purpose of the differentially regulated genes, the functional groups to which they are associated were identified using the web based gene ontology (GO) programme database for annotation, visualisation and integrated discovery (DAVID) (http://david.abcc.ncifcrf.gov/) (Dennis, Jr. *et al.* 2003; Huang, Sherman, & Lempicki 2009). The biological processes were ranked using the false discovery rate (FDR), after functional groups containing a significant number of genes (FDR<10%) from the gene list were identified (Table 5.7). The most common biological process identified by the analysis was the regulation of biological quality. Regulation of biological quality is defined as a process that modulates the frequency or rate of a measurable attribute of an organism and includes regulation of homeostatic processes, translational fidelity and protein stability. 5/6 of the contiguous genes shown in the direct interaction figure were associated with regulation of biological quality as well as *CAV1*, *CAV2* and *CDKN2C*. There was considerable overlap between the groups therefore, the four other functional groups contained many of the same genes. The effects of these biological processes, and the genes implicated, requires further examination to clarify their impact on the aetiology of *MLL* PTD leukaemia.

Gene Ontology ID	Biological Process	Genes Identified (%)	P-Value	Fold Enrichment	FDR
GO:0065008	regulation of biological quality	15 (24.6)	8.60E-08	5.81	<0.01
GO:0009605	response to external stimulus	12 (19.7)	1.40E-06	6.36	<0.01
GO:0007626	locomotory behaviour	7 (11.5)	2.00E-05	12.07	0.04
GO:0009611	response to wounding	9 (14.8)	2.63E-05	7.13	0.05
GO:0007610	behaviour	8 (13.1)	3.69E-05	8.33	0.07

Table 5.7 Gene Ontology Analysis of Biological Processes from Combined Cohort List
The top five biological processes according to FDR. GO, Gene Ontology; FDR, false discovery rate.

To enhance the ability to identify biologically important functional gene groups through interactions, the output (gene lists) from the three GeneChip cohorts were combined. In summary the GO:biological processes included processes identified previously, and indicated a link with regulation of cell proliferation; negative regulation of cellular processes, negative regulation biological processes and cell proliferation. Some of the processes identified were previously linked to leukaemia (Zangrando *et al.* 2009), however there were no common functional processes between the GO terms identified by the *MLL* PTD signature and the paediatric *MLL* translocation specific signature (Ross *et al.* 2004). The KEGG pathway database (http://www.genome.jp/kegg/pathway.html) (Kanehisa *et al.* 2010a) identified pathways related to the regulation of cell proliferation and cell cycle, but the findings were not statistically significant. These results and further discussion can be found in the appendix.

5.3.9 Expression of Genes Related to *MLL* Rearrangements

Members of the *HOX* family, specifically members of the *HOX* A cluster, along with *MEIS1* and *PBX3* are consistently up-regulated in AML and ALL leukaemic blasts with *MLL* translocations ((Armstrong *et al.* 2002; Kohlmann *et al.*

2005). Therefore expression levels of (*HOX*) genes and genes that complex with the

HOX family to regulate gene expression were evaluated in this study. The *MLL* PTD

significant gene list derived from the combined cohort did not align with *HOX* genes

normally identified with *MLL* translocations, and contained only one member of the

*HOX*A cluster. The three genes identified, *HOX*A10, *HOX*B6 and *HOX*B7 were up-

regulated. Furthermore, the genes *MEIS1* and *PBX3*, which interact with the *HOX*

pathways, were not differentially expressed. *HOX*A9 is at least partially necessary for

the leukaemic potential of some *MLL* translocations, specifically *MLL-ENL* and *MLL-*

AF10 (Ayton & Cleary 2003). Whether the difference in the *HOX* expression

suggests an alternative pathway in the leukaemogenesis of *MLL* PTD or is the result

of redundancy within the *HOX* gene family remains to be determined. The lack of

MEIS1 and *PBX3* expression may be evidence that a different pathway is used.

5.3.10 Class Prediction

Class prediction was used to determine if gene expression profiling could

accurately predict the *MLL* PTD status of a sample. Accurate prediction of *MLL* PTD

by microarray analysis could lead to its inclusion in the armoury of AML diagnostic

tests. The model selection application from Partek Genomic Suite tested the

effectiveness of four models of class prediction with the differentially regulated sets

of identified genes from the three cohorts. None of the tests revealed a gene signature

that could be used to accurately predict *MLL* PTD samples in any of the prediction

models tested (results from the support vector machine method (SVM) are shown in

Table 5.8). This was most likely due to the small number of positives samples and the

corresponding heterogeneity of expression. A similar failure of class prediction was

reported previously in a paediatric AML study where the gene signature was established from a combined group of *MLL* PTD and *MLL* translocation samples (Ross *et al.* 2004).

	# Per Class	Proportion	# Correct	# Errors	% Correct	% Error	Std. Error
pos	7	0.04	0	7	0	100	0.00%
wt	163	0.96	163	0	100	0	0.00%
The Average					50	50	
Total	170	1	163	7	95.88	4.12	1.52%
Kappa=0.0000							

Table 5.8 Classification Summary for Support Vector Machine Model
An example of class prediction results made using the SVM model with the combined cohort samples.

5.3.11 Connectivity Map Analysis

The differentially expressed genes from the combined cohort were used as a template to identify compounds that may be potentially beneficial to *MLL* PTD AML patients. The dysregulated genes associated with a mutation can identify potential target areas for therapeutic intervention by revealing unique pathways necessary for leukaemogenesis. Therapeutic drugs that may reverse or repress *MLL* PTD were identified using the web based tool, Connectivity Map (CMAP) (http://www.broadinstitute.org/CMAP/) (Lamb *et al.* 2006). CMAP connects up- and down-regulated genes of interest with a database of gene expression profiles derived from human cell lines treated with over 1,300 compounds; this enables the identification of agents that produce the same or the inverse of the expression signature (Lamb *et al.* 2006). When administered to human cell lines, three drugs, tanespimycin, trichostatin A and vorinostat, effectively produced an inverse of the *MLL* PTD gene signature (Table 5.9). The composite analysis of all instances

examined for these three drugs showed negative enrichment scores, meaning they produced the inverse of the gene signature template and may be useful for the reversal of AML. The one instance of significant positive enrichment, (i.e. a treated cell line with a similar gene signature) for tanespimycin involved the cell line HL60, an *MLL* PTD negative promyelocytic cell line. However, the CMAP specificity table, which determines how many of 312 published gene signatures scored higher (if positively enriched) than the gene signature of interest, demonstrated that the majority of AML derived GEPs were associated with the negatively-enriched cell lines, therefore the results bear further investigation.

Rank	Drug Name	Cell Line	Mean	n	Enrichment	P-value	Specificity
1	tanespimycin	HL60	0.405	12	0.726	<1.00E-05	0.0756
2	trichostatin A	MCF7	-0.433	92	-0.421	<1.00E-05	0.4654
3	wortmannin	MCF7	0.549	10	0.708	2.00E-05	0.0915
4	dapsone	MCF7	0.82	2	0.99	1.40E-04	0
5	tanespimycin	MCF7	-0.349	36	-0.355	1.60E-04	0.4098
6	phenoxybenzamine	MCF7	-0.797	3	-0.955	2.20E-04	0.0092
7	vorinostat	MCF7	-0.562	7	-0.733	2.20E-04	0.2136
8	trichostatin A	PC3	-0.309	55	-0.276	3.40E-04	0.7356
9	decamethonium bromide	MCF7	0.771	2	0.985	4.00E-04	0
10	khellin	MCF7	0.8	2	0.98	7.00E-04	0

Table 5.9 Connectivity Map Analysis of Combined Cohort Significant Gene List
The top ten identified drugs are listed according to their p-value. n: number.

Tanespimycin, also known as 17-AAG, is an antineoplastic antibiotic that inhibits the protective chaperone activity of the heat shock protein 90 (*HSP90*), which binds to many known oncogenic signalling proteins (Al *et al.* 2008). Therefore, tanespimycin promotes degradation of the oncogenic proteins over expressed in tumour cells. Trichostatin A and vorinostat inhibit HDACs resulting in hypereacetylated histones which alter chromatin configuration and ultimately gene transcription. HDAC inhibitors (HDACi) cause differentiation, growth arrest and apoptosis of cells *in vitro* and inhibit tumour proliferation in animals (Marks *et al.* 2001). Tanespimycin and the HDAC inhibitors are compounds with potential to be effective in a leukaemic setting (Section 5.4).

5.3.12 Identification and Validation of GOI

5.3.12.a Genes of Interest Chosen by Venn Diagram Intersection

Due to the scatter of the *MLL* PTD samples seen in the PCA and dendrogram analysis, the genes demonstrating reproducible differential expression in the significant probe lists of all three cohorts were chosen for further analysis. To accomplish this, a Venn diagram was created using the three significant probe set lists generated from the

three groups of samples. Nine probe sets were found in the overlap between the three

cohorts (Figure 5.11). These nine probe sets were reproducibly detected through two

independent batches of samples and on separate Affymetrix GeneChip types. The

nine probe sets identified by all three groups were marked as potential validation

genes and candidates for further study as markers for MRD (Figure 5.12).
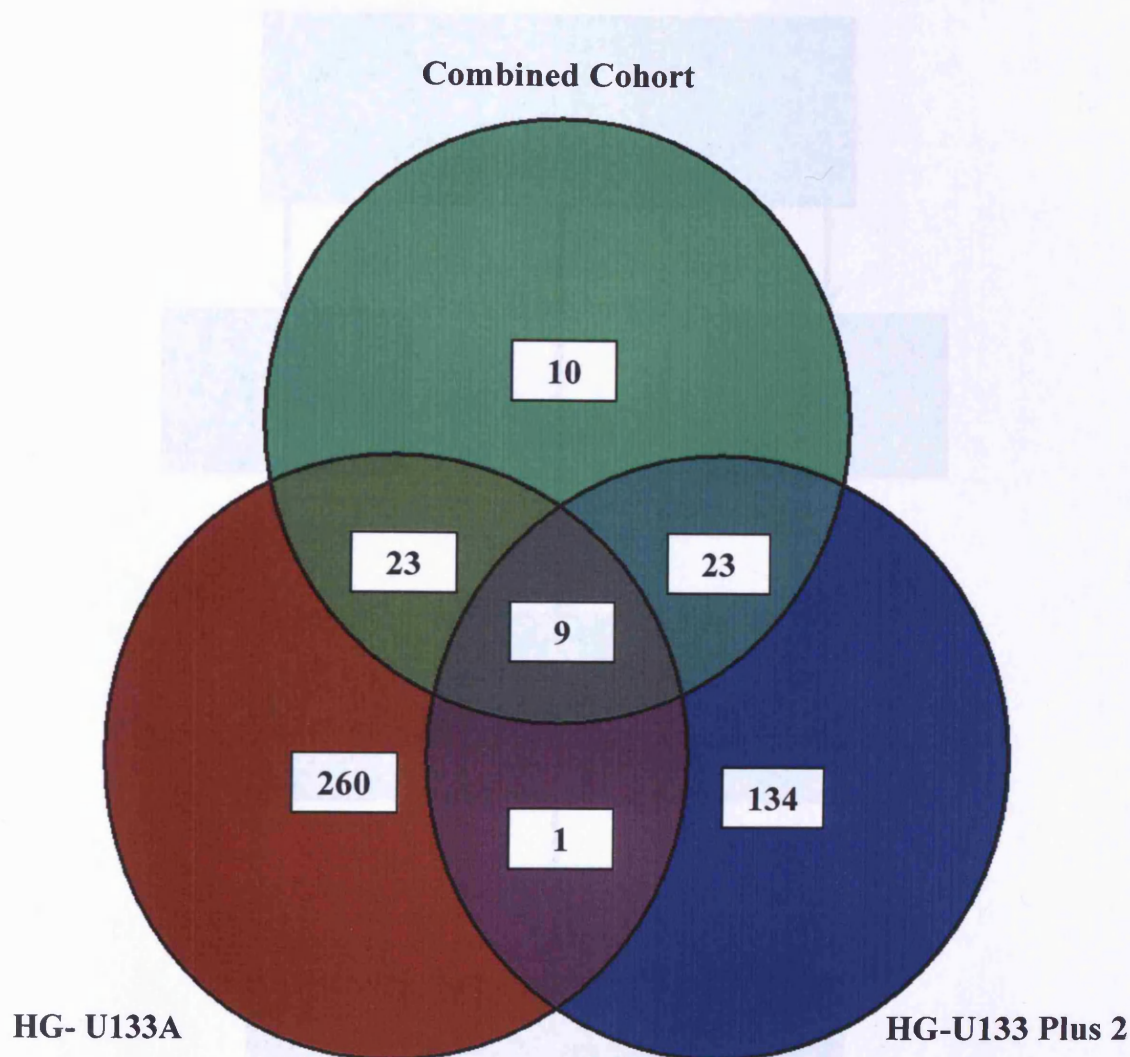
Figure 5.11 Identification of *MLL* PTD Dysregulated Probe Sets
Probe sets identified from HG-U133A, HG-U133 Plus 2 and the combined cohort.
Numbers indicate how many probe sets were identified in the intersection. Nine
discriminating probe sets were common to all three cohorts. The intersections are not
to scale.
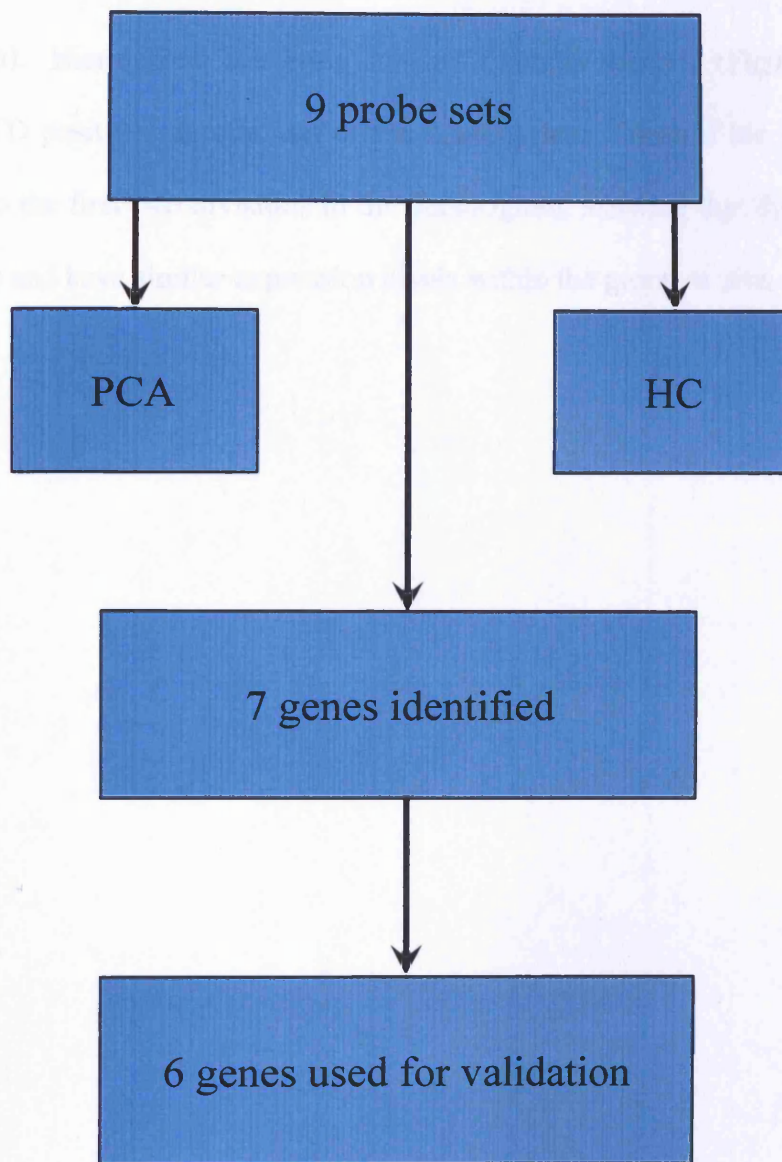
Figure 5.12 Flow Chart Depicting Gene of Interest Analysis Process
The nine probe sets were complementary to seven genes, six of which were used for
validation of gene expression data. PCA: principle component analysis; HC:
hierarchical component analysis.

PCA mapping using these nine transcripts was performed to determine if *MLL*

PTD samples could be discriminated from the AML *MLL* PTD negative samples.

Although not as compact as in previous PCA maps, the *MLL* PTD samples were

separated from the majority of the other samples and all occurred in one quadrant

(Figure 5.13). Hierarchical clustering showed a similar scenario (Figure 5.14) with the *MLL* PTD positives on one half of the dendrogram. Three of the samples were grouped into the first two divisions in the dendrogram, showing that they are unique to the others and have similar expression levels within the group of nine.
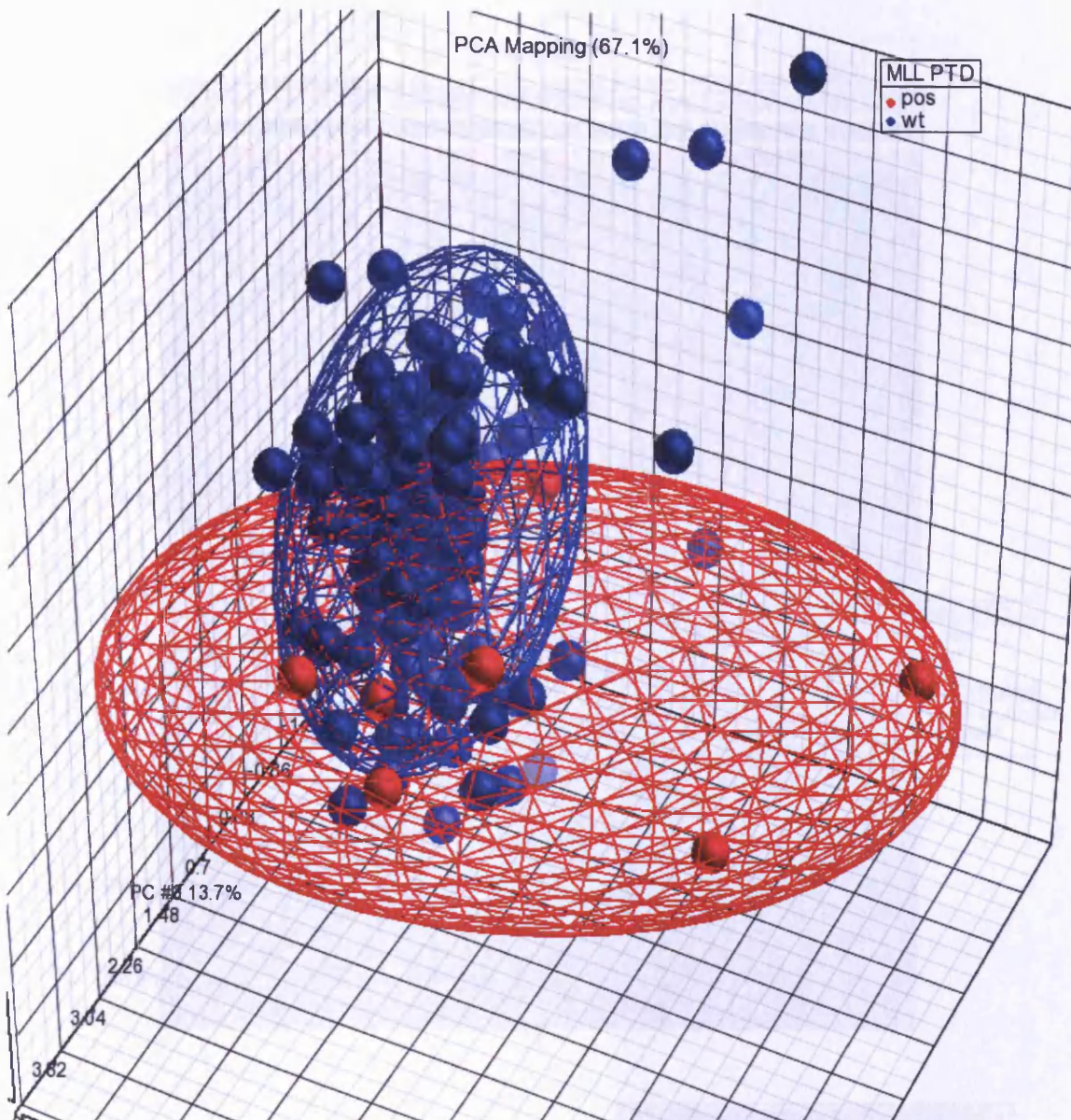
Figure 5.13 Principal Component Analysis of the Combined Cohort Using the Significant Probe Sets Identified
The samples are projected in a three-dimensional space, plotted on the basis of the three principal components that best capture the variance of the data. Each case is represented by a coloured sphere. The *MLL* PTD group (n=7) showed accurate separation from AML *MLL* WT (n=162) using the 9 discriminating probe sets identified as common between HG-U133A, HG-U133 Plus 2 and the combined cohort. Samples were colour coded as indicated in legend.
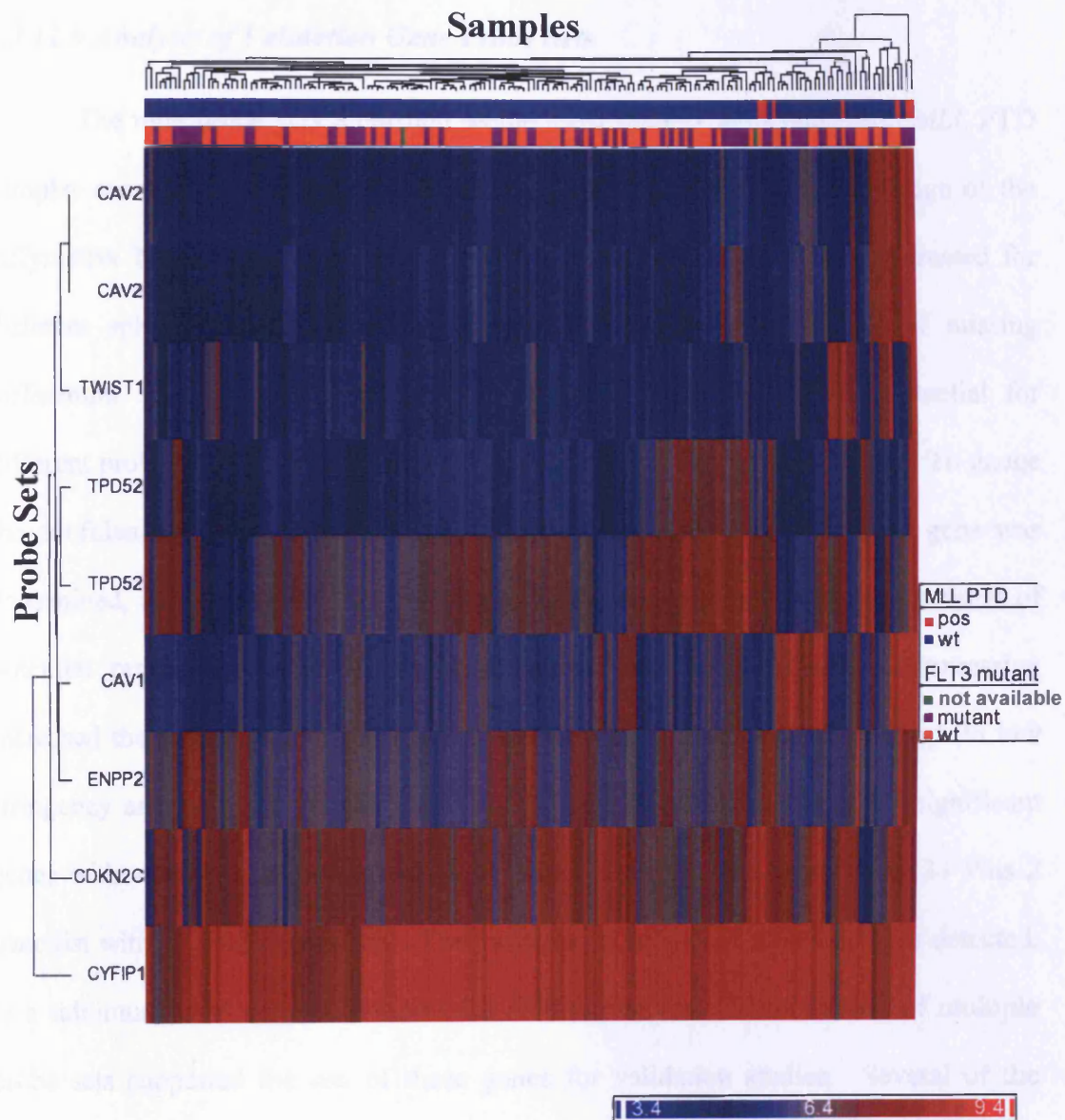
Figure 5.14 Hierarchical Cluster Analysis of Combined Cohort Using 9 Significant Probe Sets.

Hierarchical cluster analysis of AML (n=162) samples versus 9 *MLL* PTD discriminating probe sets. Probe sets are shown by rows and samples by columns. The normalised expression value of the probe sets are shown, with colour coding representing deviation from the mean, with the higher expression in red and lower expression in blue. Indicated below the samples is the *MLL* PTD and *FLT3* ITD classification. The green colourings for *FLT3* indicate that no data were available.

5.3.12.b Analysis of Validation Gene Probe Sets

The nine probe sets identified as the most reliably associated with *MLL* PTD samples are complementary to 7 different genes (Table 5.10A). The design of the Affymetrix human genome chips resulted in multiple probe sets being created for different splice variants of the same gene. This reduced the risk of missing differential expression due to gene variation, but also created the potential for different probe sets of the same gene to record different expression levels. To gauge the confidence of the result, the number of potential probe sets for each gene was determined, along with whether it was detected in a gene list. A greater number of potential probe sets identified in combination with high differential expression increased the confidence of the result (Table 5.10B). The groups from high to low stringency are the Venn intersection list, the combined cohort list of 65 significant genes with >2 fold expression and lastly either the HG-U133A or HG-U133 Plus 2 gene list with <2 fold expression. The multiple probe sets of all genes were detected, to a minimum level of $p<0.05$ and >1.2 fold expression. Identification of multiple probe sets supported the use of these genes for validation studies. Several of the genes are reported to have an association with the promotion of an oncogenic phenotype including *CAV1*, *CAV2*, *CDKN2C*, *ENPP2*, *TPD52* and *TWIST1*. Expanded examination of the genes identified can be found in the Discussion.

A.

Affymetrix ID	Common Name	Gene Title	GenBk AcNo.	Gene GBP	p-value(pos/swt)	FoldChange(pos/swt)
212097_at	CAV1	caveolin1, caveoprotein 22kDa	NM00173	MAPKK cascade	4.0E-04	2.64
203324_s_at	CAV2	caveolin2	NM00123	negative regulation of endothelial cell proliferation	1.5E-05	3.16
203323_at	CAV2	caveolin2	NM00123	negative regulation of endothelial cell proliferation	5.0E-05	2.78
204159_at	CDKN2C	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	NM00162	regulation of cyclin-dependent protein kinase activity	2.4E-04	2.60
208923_at	CYFP1	cytoplasmic FMR1 interacting protein 1	NM0103028	multicellular organismal development	8.9E-04	2.15
210839_s_at	ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase2 (autotaxin)	NM0040092	phosphate metabolic process	6.8E-06	2.35
201690_s_at	TPD52	tumor protein D52	NM00025252	anatomical structure morphogenesis	1.7E-03	2.35
201689_s_at	TPD52	tumor protein D52	NM00025252	anatomical structure morphogenesis	7.5E-05	2.46
213943_at	TWIST1	twist homolog 1 (acrocephalosyndactyly 3; Saethre-Chotzen syndrome) (Drosophila)	NM00044	negative regulation of transcription from RNA polymerase II promoter	2.0E-04	2.33

B.

Affymetrix ID	Common Name	Venn Intersection	Expression >2 Fold	Expression >2 Fold
212097_at	CAV1	Y	Y	n/a
203065_s_at		N	N	Y (1.77)
203324_s_at	CAV2	Y	Y	n/a
203323_at		Y	Y	n/a
204159_at	CDKN2C	Y	Y	n/a
211792_s_at		N	N	Y (1.3)
208923_at	CYFIP1	Y	Y	n/a
210839_s_at	ENPP2	Y	Y	n/a
209392_at		Y	Y	n/a
201690_s_at	TPD52	Y	Y	n/a
201689_s_at		Y	Y	n/a
201688_s_at		N	Y	n/a
201691_s_at		N	N	Y (1.2)
213943_at	TWIST1	Y	Y	n/a

Table 5.10 Differential Expression of Genes and Assessment of Value as Discriminator for *MLL* PTD
A) The nine dysregulated probe sets representing seven genes are listed according to (p value). Fold change and p-value are given with respect to the combined cohort values calculated. Fold change is given with respect to AML *MLL* WT mean expression level. B) The probe sets for each gene are indicated as present or absent for three categories of decreasing stringency; the Venn intersection of probe sets for all three cohorts, the combined cohort list of probe sets > 2 fold expression and the combined cohort list of probe sets < 2 fold expression.

5.3.12.c Selection of Positive Controls Cell Lines for Validation Genes

Six of the seven genes common to all cohorts were selected to validate the Affymetrix expression data by correlation with expression levels determined by RQ-PCR analysis. *CAV1* was not used for validation due to its similarity in expression and function to *CAV2*. Affymetrix HG-U133A data were used to locate cell lines as positive controls for the RQ-PCR validation. The probe sets for the GOIs were screened for increased expression levels in seven leukaemic cell lines previously analysed on the microarray GeneChip. After normalisation the expression levels given for each cell line were rated as either absent, marginal or present for the probe set in question. The present or absent call was based on the comparison of sample hybridisation levels to the matched and mismatched probes that make up the probe set. This rating prevents non-specific hybridisation from being misinterpreted as a high expression level. An example of the results is shown for the gene *TWIST1* (Figure 5.15). In this example, six of the cell lines were rated as absent for the gene and one as present. The transcript for *TWIST1* was evident in the cell line NB4-R2. Interestingly its expression was significantly higher (p=0.028) than the expression levels in the other cell lines which included NB4, the cell line from which NB4-R2 is derived. NB4 is a commonly used cell line in the study of APL, and has been used to derive a number of drug resistant cell lines (e.g. NB4-R2) (Roussel & Lanotte 2001). ATRA (all-trans retinoic acid) causes NB4 to undergo granulocytic maturation whereas NB4-R2 cells are unresponsive due to a point mutation in the retinoid-binding domain of *PML-RARα* (Duprez *et al.* 2000). *TWIST1* is a transcription factor associated with drug resistance in breast cancer, so could potentially be connected with the drug resistance acquired by the NB4-R2 cell line. Three other cell lines also

showed up-regulation of one or more of the genes of interest (Table 5.11). These four

cell lines were chosen as the controls for the RQ-PCR validation.
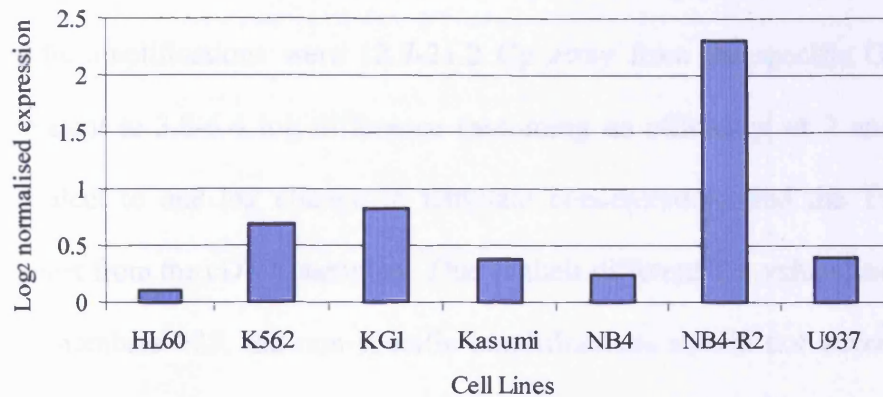
Figure 5.15 Differential *TWIST1* Expression in a Selection of Leukaemic Cell Lines.
Expression of *TWIST1* is significantly up-regulated in NB4-R2 (p=0.028) relative to other cell lines.

Cell Line	Gene(s)
KG1	ENPP2
NB4-R2	TWIST1
K562	CAV2, CYFIP1, CDKN2C
HL60	TDP52

Table 5.11 Positive Control Cell Lines for RQ-PCR Validation
Gene expression analysis results identified four leukaemic cell lines with up regulation of the six genes of interest.

5.3.12.d RQ-PCR Optimisation of Validation Genes

The cell lines identified as positive controls were used to confirm the efficacy of the PCRs and optimisation of PCR conditions. All six of the genes expressed in these cells successfully amplified PCR products using the initial reaction conditions, with sufficient levels of quality as determined by slope, Cp value and fluorescent intensity to make further optimisation unnecessary. The Cps of all of the genes are shown in Table 5.12. The range of the Cps in the cell lines varies between 18.9 for *CAV2* and 24.9 for *TWIST1* and is a result of the expression level of the genes and the

efficiency of the RQ-PCR assays. All of the assays, apart from *TWIST1*, showed low levels of non-specific amplification in H_2O with a minimum Cp of 35. The non-specific amplifications were 12.7-21.2 Cp away from the specific GOI Cp values, equivalent to 3.8-6.4 log difference (assuming an efficiency of 2 and 3.3 cycles is equivalent to one log change in template concentration) and the Tm values were different from the cDNA samples. Due to their different Tm values and appearance at cycle numbers >35, the non-specific amplifications should not adversely affect the performance or analysis of the RQ-PCR assays. The difference in the level of expression for *TWIST1* between the cell lines NB4 and NB4-R2 observed in the Affymetrix data was supported by the RQ-PCR results. *TWIST1* expression is calculated to be 3.2 fold higher in NB4-R2 than NB4 using ΔCp values. The identity of the amplicons was confirmed by size comparison. Gel electrophoresis sizing of all of the amplicons, including *ABL*, matched the predicted sizes from primer 3 (Figure 5.16).

Gene	Cell Line	Avg Cp	Avg Tm
CAV2	K562	18.9	90.5
CDKN2C	K562	20.5	85.4
CYFIP1	K562	21.1	87.9
ABL	K562	17.4	88.9
ENPP2	KG1	24.3	88.0
ABL	KG1	20.0	89.3
TPD52	HL60	21.1	86.8
ABL	HL60	19.8	88.9
TWIST1	NB4-R2	24.9	92.3
ABL	NB4-R2	20.6	88.8
TWIST1	NB4	33.8	92.3
ABL	NB4	19.6	88.8

Table 5.12 Gene Expression Levels in Positive Controls (RQ-PCR)
Average crossing points and Tm of validation genes (in duplicate) as well as control gene *ABL* for each positive control cell line. Avg: average; Cp: crossing point; Tm: temperature of melting.
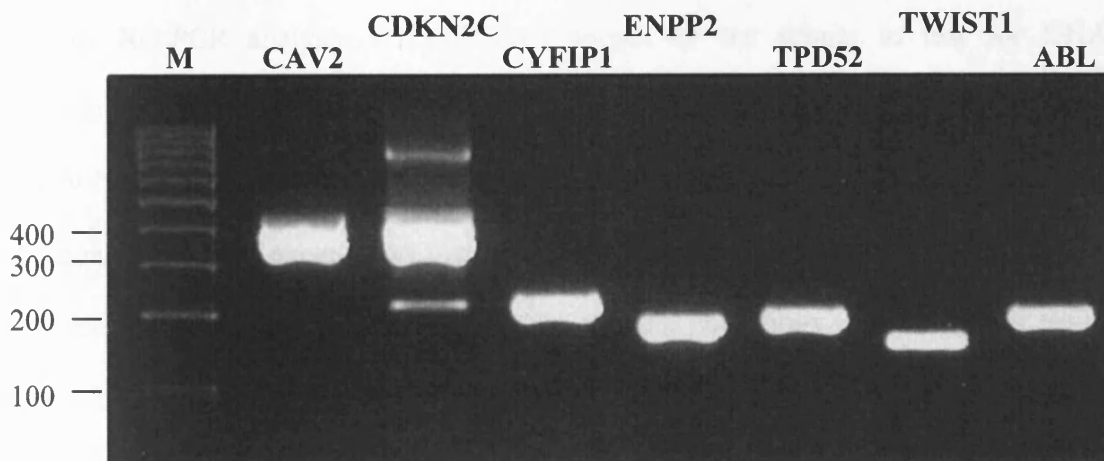
Figure 5.16 Analysis of RQ-PCR Amplicon Size.
The amplified RQ-PCR products were run on an agarose gel to confirm base pair length. All products were the size predicted by the primer design and Blast analysis. M: marker.

5.3.12.e Test for Non-Specific Amplification of DNA

To prevent the possibility of DNA contamination being misinterpreted as RNA expression, all primer pairs were tested for the ability to amplify DNA. The standard conditions of amplification were used to amplify 5ng of DNA. All samples amplified the DNA sample with different levels of efficiency; the lowest, most efficient value was *CAV2* at Cp=20 and the highest was *ABL* at Cp=33.9 (Figure 5.17). Two of the six genes, *CAV2* and *CDKN2C*, produced DNA Tm values that were different than the values produced from the RNA samples, which would make any DNA amplification recognisable. The remaining four genes could potentially have DNA contamination misinterpreted as expression of RNA. To prevent mistakes of this sort, all samples were pre-screened for the presence of DNA contamination before RQ-PCR analysis. *CAV2* was chosen as the primer to test for DNA contamination in all RNA samples due to its higher comparable efficiency at amplifying DNA. A "mock" RT reaction was performed with all conditions and reagents as normal except that the reverse transcription enzyme was excluded, thus preventing the production of cDNA. The standard conditions of the RQ-PCR assay were followed and the results were analysed for the presence of DNA. Although some amplification was observed, it was non-specific, irreproducible and had Cp values of 35 and higher. All samples used in the RQ-PCR assays were tested prior to analysis and found to be free from DNA contamination.
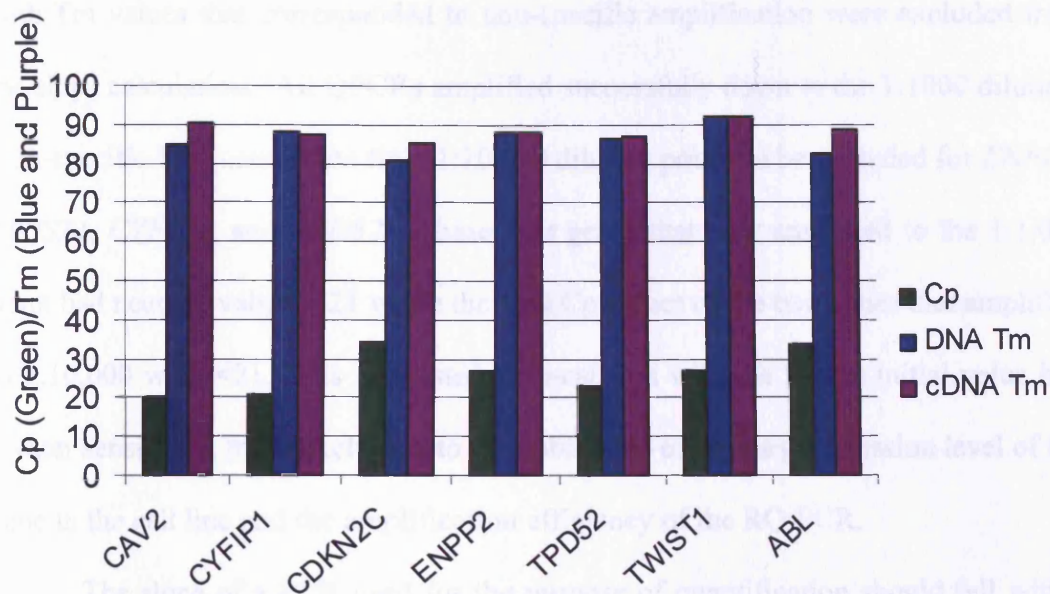
Figure 5.17 Quantification of DNA Amplification
The Cp of DNA amplification are indicated by green bars, DNA Tm by blue bars and cDNA Tm by purple bars. The value of the Cp is inversely related to the amount of RQ-PCR product amplified, therefore *CAV2* and CYFIP1amplify DNA most readily. ; Cp: crossing point; Tm: melting temperature.

5.3.12.f Establishing RQ-PCR Efficiency

An accurate assessment of gene expression level by RQ-PCR requires comparable efficiencies between the target gene and the control gene, as discussed previously in Chapter 3. In order to compare the efficiencies, the cDNAs of all of the control cell lines were analysed neat and with 4 serial dilutions of 1:10 to establish the slope of amplification over a range of template concentrations (Figure 5.18). Differences in amplification due to the concentration of the PCR reagents were prevented by diluting the cDNAs into a solution of RT master mix ensuring that the concentration was the same for the neat sample as it was for the dilutions. Each cell line was amplified using the primers for the control gene *ABL* and the appropriate GOI and then the resulting slopes were compared (Figure 5.19 a-f). The Tm values of all amplifications were checked to verify the specificity of the amplification; samples

with Tm values that corresponded to non-specific amplification were excluded from the slope calculation. All QPCRs amplified successfully down to the 1:1000 dilution. Non-specific Tms caused the final 1:10,000 dilution points to be excluded for *ENPP2*, *TWIST1*, *CYFIP1*, and *TPD52*. These four genes that only amplified to the 1:1,000 point had neat Cp values >21 while the neat Cp values of the two genes that amplified to 1:10,000 were <21. This suggested the neat Cps with the lowest initial value had greater sensitivity, most likely due to a combination of primary expression level of the gene in the cell line and the amplification efficiency of the RQ-PCR.

The slope of a PCR used for the purpose of quantification should fall within the range of –3.9 and -3.0 and have a coefficient of determination (i.e. a measure of the association between the dilution points and the calculated trend line) of >0.95 (Van, V *et al.* 2003b). The R^2 values for the GOIs and *ABL*s were 0.98 or higher and the slopes of the GOIs ranged from a low of –4.0 for *CAV2*, *CYFIP1* and *TPD52* to a high of –3.5 for *CDKN2C*. The lack of variation in the slope of *ABL*, 3.6-3.7, in the four different cell lines demonstrated reproducibility, as the efficiency would not be expected to change due to the origin of the cDNA. The difference in values of slope between the target genes and the control gene were acceptable for the purposes of quantifying RNA expression levels for validation.
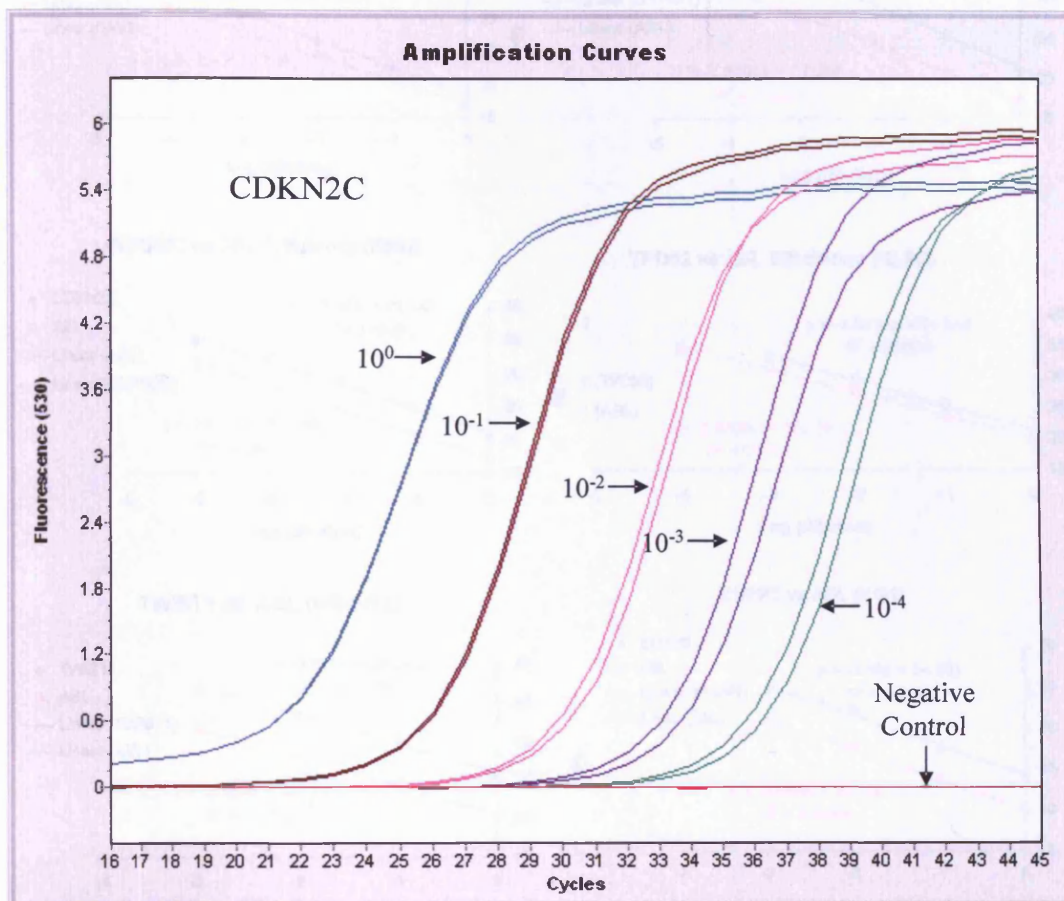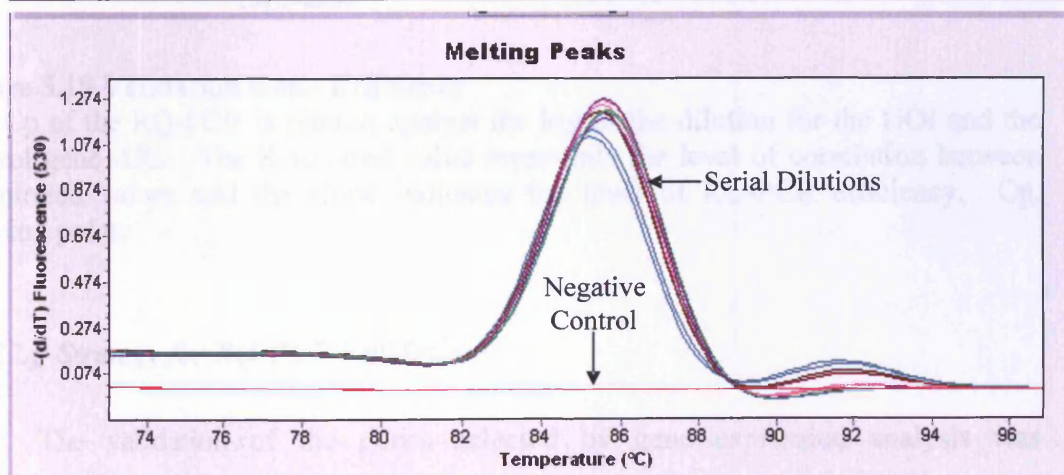
Figure 5.18 Light Cycler Amplification of CDKN2C Serial Dilution
A) Real time Syber Green fluorescent history was plotted against the cycle number of serially diluted *CDKN2C* cDNA (10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}). The duplicate dilutions amplified reproducibly across the range of dilutions. B) Fluorescent history versus temperature. The negative derivative of fluorescence over time was plotted against temperature to calculate the melting temperature of the PCR products. All products recorded a consistent Tm and the negative control did not amplify.
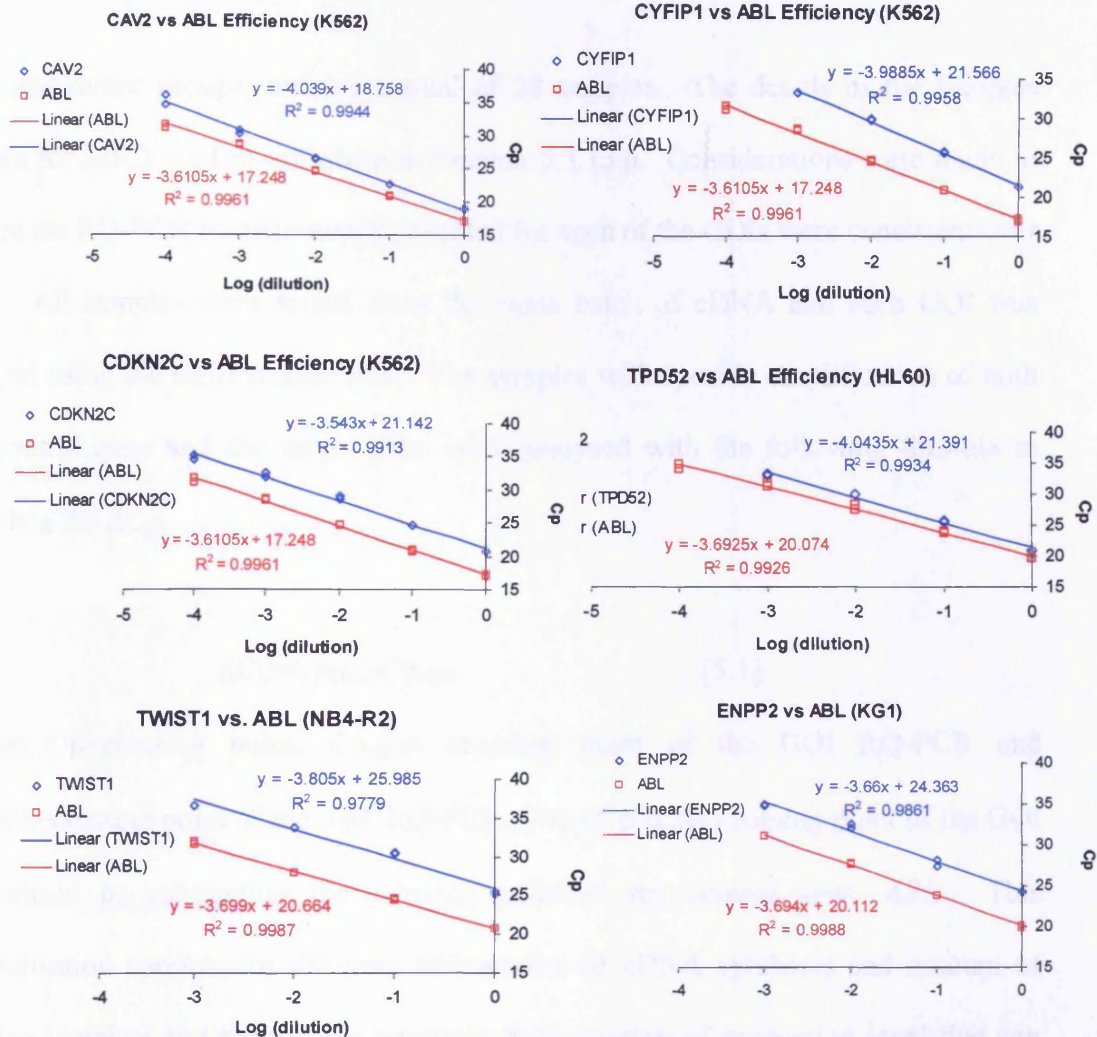
Figure 5.19 Validation Gene Efficiency
The Cp of the RQ-PCR is plotted against the log of the dilution for the GOI and the control gene *ABL*. The R squared value represents the level of correlation between the plotted values and the slope indicates the level of RQ-PCR efficiency. Cp, crossing point.

5.3.12.g Strategy for RQ-PCR Validation

The validation of the genes selected by gene expression analysis was performed by comparing the relative expression levels of the microarray results with results from RQ-PCR analysis. All microarray samples with available material were selected for analysis: 3 *MLL* PTD, 9 AML *MLL* PTD WT (CN), and 4 healthy control samples. Twelve more samples were analysed at the same time as the validation studies to screen for potential MRD markers and to increase the numbers in the *MLL*

5-286

PTD and donor groups, making a total of 28 samples. The details of the samples chosen for MRD studies are given in Section 5.3.13.a. Considerations were made to ensure the RQ-PCR measurements recorded for each of the GOIs were consistent over time. All samples were tested from the same batch of cDNA and each GOI was assayed using the same master mix. The samples with specific amplification of both the control gene and the target gene were analysed with the following formula to calculate the ΔCp.

$$\Delta Cp = Cp_{GOI} - Cp_{ABL} \qquad [5.1]$$

Where Cp=crossing point, Cp_{GOI}= crossing point of the GOI RQ-PCR and Cp_{ABL}=crossing point of the *ABL* RQ-PCR. The ΔCp is the crossing point of the GOI normalised by subtracting the crossing point of the control gene, *ABL*. This normalisation corrects for different efficiencies of cDNA synthesis and amount of starting template and produces a normalised description of expression level that can be used to compare different samples.

The *ABL* RQ-PCR was the first assay to be performed from the cDNA and then used for all subsequent calculations. To ensure that the first *ABL* assay was valid for use over the time period that the tests took place, it was repeated after all the assays were completed. The mean difference between the Cp of the samples of the first *ABL* run and the last *ABL* run was 0.29 ± 0.23, (n=28). The small difference supported the hypothesis that the expression levels calculated by RQ-PCR were stable over time.

5.3.12.h Correlation of RQ-PCR and Microarray Expression Levels

The expression levels determined from the microarray analysis were compared with the ΔCp values determined by RQ-PCR, to validate the expression levels of the genes of interest. The six genes chosen for validation were all up-regulated; with ΔCp value inversely related to the amount of expression, the correlation between the RQ-PCR and microarray data should also be inversely correlated, giving a negative slope. Five of the six genes had an r value of less than -0.637 and showed significant correlation ($p<0.05$) between RQ-PCR and microarray expression levels (Table 5.13). The correlation plots for all of the GOIs are shown in Figure 5.20. Only *CDKN2C* showed a lack of correlation, with an r value of -0.416 and non-significant p value. The r and p values of the correlation studies indicated overall agreement between the RQ-PCR and microarray expression levels thus validating the microarray analysis.

CAV2 Gene Expression Correlation

$y = -0.5375x + 6.0529$
$R^2 = 0.4049$

CDKN2C Gene Expression Correlation

$y = -0.4416x + 4.8889$
$R^2 = 0.1742$

CYFIP1 Gene Expression Correlation

$y = -1.1421x + 12.986$
$R^2 = 0.6629$

ENPP2 Gene Expression Correlation

$y = -1.2915x + 11.74$
$R^2 = 0.4685$

TPD52 Gene Expression Correlation

$y = -1.0039x + 11.083$
$R^2 = 0.5293$

TWIST1 Gene Expression Correlation

$y = -2.3667x + 21.708$
$R^2 = 0.455$
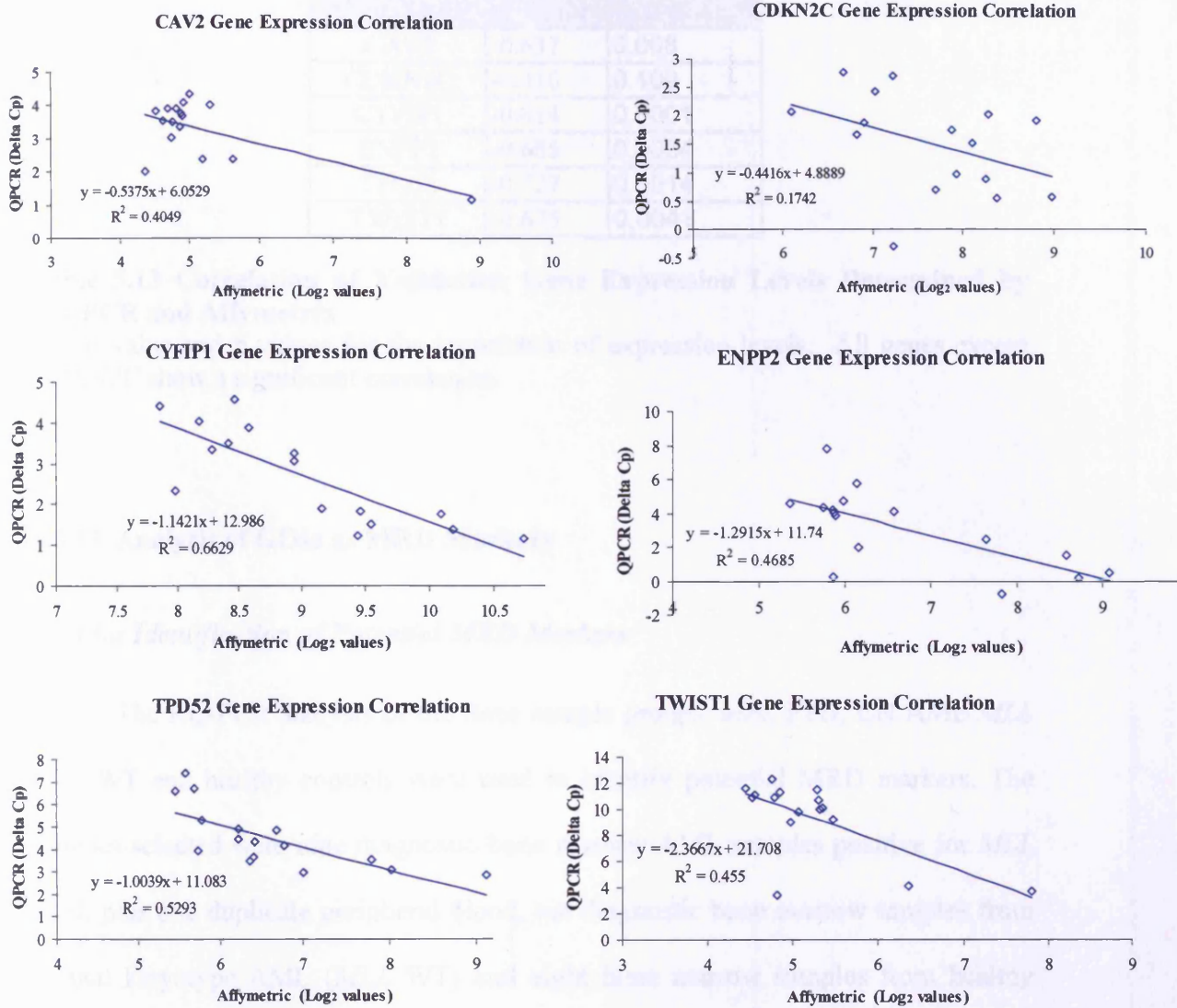
Figure 5.20 Correlation of Validation Gene Expression Levels Determined by RQ-PCR and Affymetrix Microarray
Correlation between the level of gene expression calculated by RQ-PCR and microarray. RQ-PCR delta Cp values are plotted against the Affymetrix microarray results. The R squared represents the level of correlation between the values. The Delta Cp value is inversely related to the microarray data therefore correlation should be represented by a negative slope. Cp, crossing point.

5-289

Gene Name	r Value	P value
CAV2	-0.637	0.008
CDKN2C	-0.416	0.109
CYFIP1	-0.814	0.0001
ENPP2	-0.685	0.0034
TPD52	-0.727	0.0014
TWIST1	-0.675	0.0041

Table 5.13 Correlation of Validation Gene Expression Levels Determined by RQ-PCR and Affymetrix
The R value and p values for the correlation of expression levels. All genes except *CDKN2C* show a significant correlation.

5.3.13 Analysis of GOIs as MRD Markers

5.3.13.a Identification of Potential MRD Markers

The RQ-PCR analysis of the three sample groups, *MLL* PTD, CN AML *MLL* PTD WT and healthy controls were used to identify potential MRD markers. The samples selected were nine diagnostic bone marrow AML samples positive for *MLL* PTD, plus one duplicate peripheral blood, ten diagnostic bone marrow samples from normal karyotype AML (*MLL* WT) and eight bone marrow samples from healthy donors. Where possible, the normal karyotype AML (*MLL* PTD WT) samples which had follow up samples available were selected preferentially to enable MRD studies.

The representation of GOI expression level as ΔCp meant that the lower points on the y-axis indicated higher expression of the gene. Only two genes *CYFIP1* and *TPD52,* showed a significant difference in expression (p=0.0001 and 0.0007 respectively). *CYFIP1* was significantly up-regulated in the *MLL* PTD group when compared to the AML *MLL* WT (p<0.05) and the healthy control group (p<0.001) (Figure 5.21 and 5.22). The difference between the *MLL* PTD and the healthy donor group indicated *CYFIP1* was a candidate MRD marker gene. *TPD52* was down-

regulated in the AML *MLL* WT compared to the healthy controls (p<0.05) and *MLL*

PTD (p <0.01). Interestingly, there was no significant difference between the *MLL*

PTD and the donor groups (p>0.05). Based on the significant transcript levels

determined from the RQ-PCR results, *CYFIP1* was the only gene that showed

potential as a marker for MRD for the *MLL* PTD positive samples.
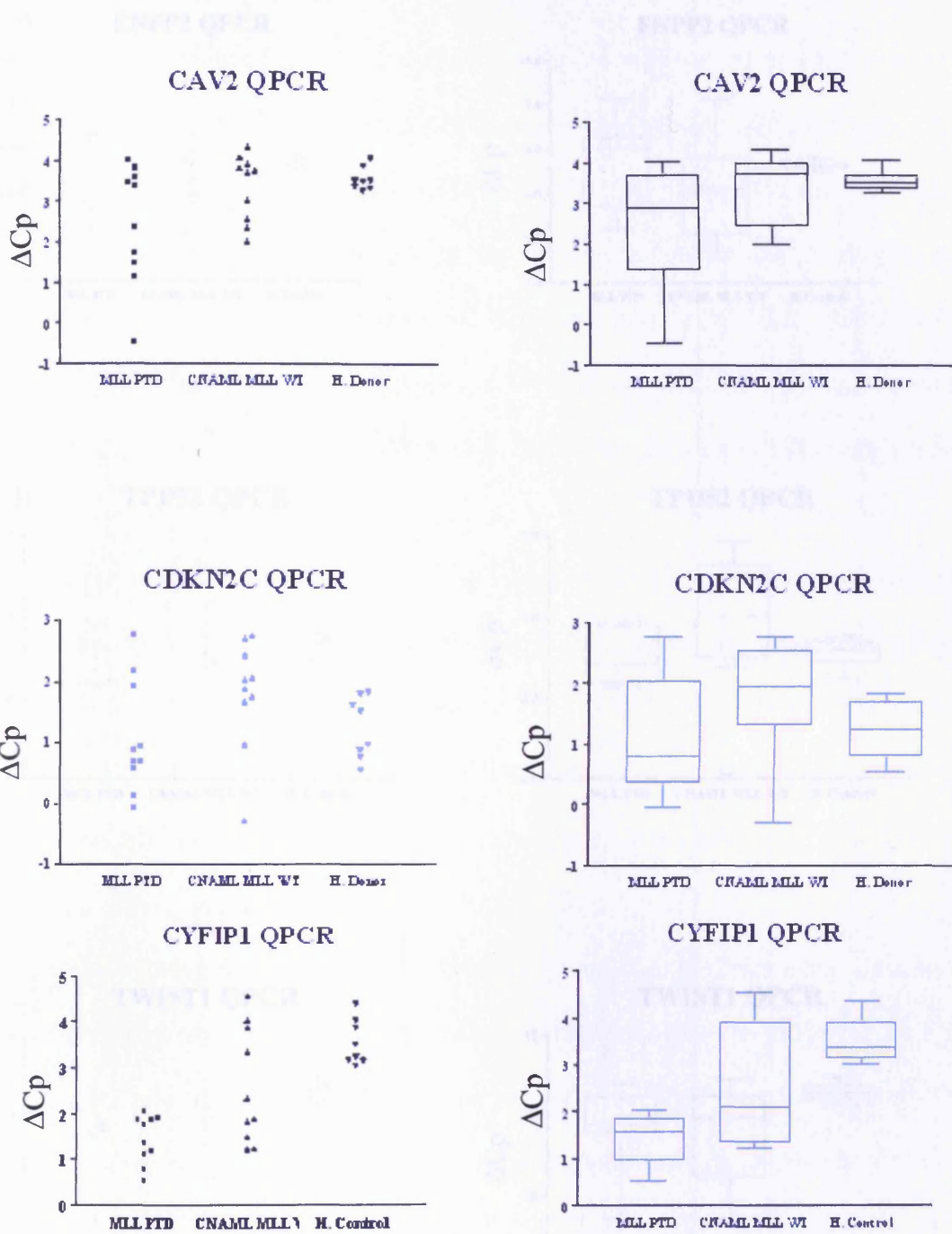
Figure 5.21 Gene Expression of Validation Genes in *MLL* PTD, CN AML *MLL* WT and Healthy Controls

The gene expression results from the RQ-PCR assay are plotted as a column scatter plot and with a box and whiskers graph indicating the mean and 95% confidence interval. The delta Cp is inversely related to the level of gene expression detected, therefore the lower the point the higher the expression. A) *CAV2*, B) *CDKN2C*, C) CYFIP1. See following page., D) *ENPP2*, E) *TPD52*, F) *TWIST1*.
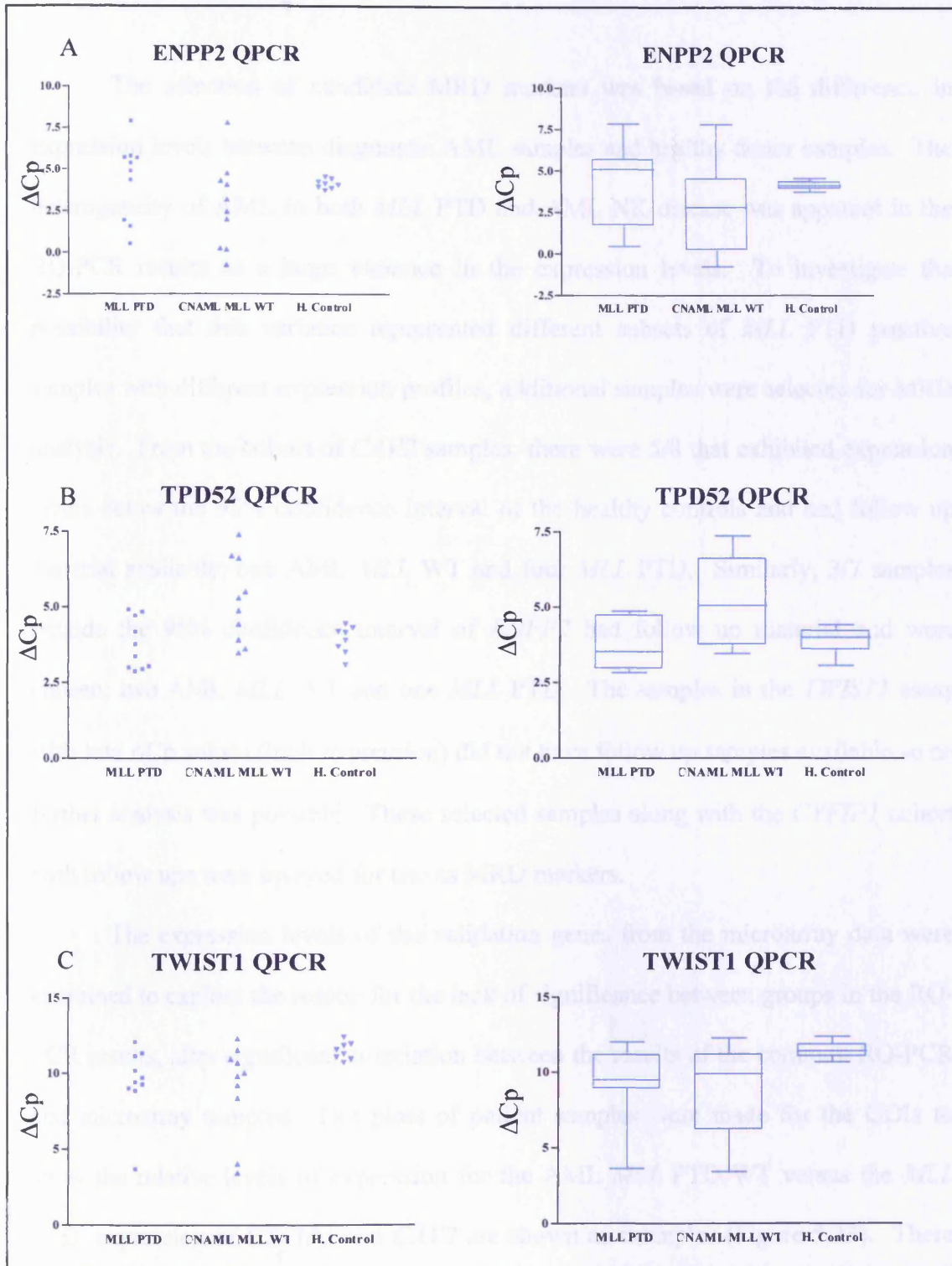
Figure 5.22 Gene Expression of Validation Genes in *MLL* PTD, CN AML *MLL* WT and Healthy Controls
The gene expression results from the RQ-PCR assay are plotted as a column scatter plot and with a box and whiskers graph indicating the mean and 95% confidence interval. The ΔCp is inversely related to the level of gene expression detected, therefore the lower the point the higher the expression. A) *ENPP2*, B) *TPD52*, C *TWIST1*.

The selection of candidate MRD markers was based on the difference in expression levels between diagnostic AML samples and healthy donor samples. The heterogeneity of AML in both *MLL* PTD and AML NK disease was apparent in the RQ-PCR results as a large variance in the expression levels. To investigate the possibility that this variance represented different subsets of *MLL* PTD positive samples with different expression profiles, additional samples were selected for MRD analysis. From the cohort of *CAV2* samples, there were 5/8 that exhibited expression levels below the 95% confidence interval of the healthy controls and had follow up material available; one AML *MLL* WT and four *MLL* PTD. Similarly, 3/7 samples outside the 95% confidence interval of *ENPP2* had follow up material and were chosen; two AML *MLL* WT and one *MLL* PTD. The samples in the *TWIST1* assay with low ΔCp values (high expression) did not have follow up samples available so no further analysis was possible. These selected samples along with the *CYFIP1* cohort with follow ups were assayed for use as MRD markers.

The expression levels of the validation genes from the microarray data were examined to explore the reason for the lack of significance between groups in the RQ-PCR results, after significant correlation between the results of the common RQ-PCR and microarray samples. Dot plots of patient samples were made for the GOIs to show the relative levels of expression for the AML *MLL* PTD WT versus the *MLL* PTD; expression of *CYFIP1* and *CAV2* are shown as examples (Figure 5.23). There was less overlap between the two groups for CYFIP1 than the other GOIs, which is comparable to the result seen in the RQ-PCR assay. The plots for selected probe sets showed that the RQ-PCR accurately represented the varied expression levels of the *MLL* PTDs, with the overlap between the two groups clearly seen in all examples.
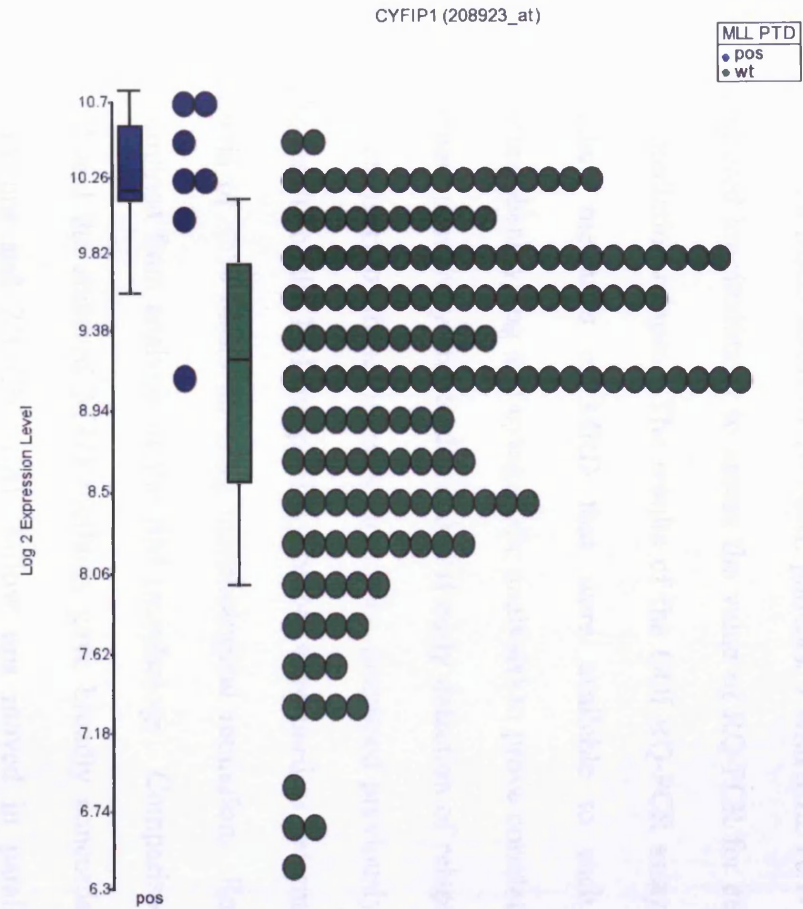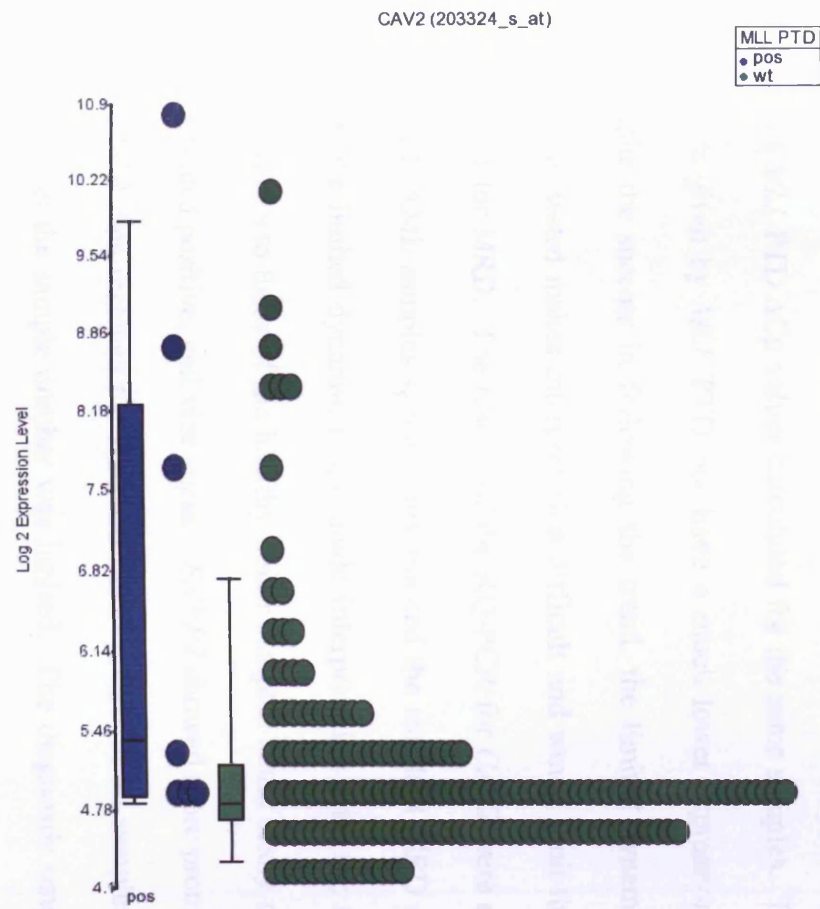
Figure 5.23 Dot Plot Comparison of *MLL* PTD and AML *MLL* WT Microarray Gene Expression Levels
The log 2 of the gene expression level for *MLL* PTD and AML *MLL* WT samples are plotted with each point representative of a patient sample. The scales of the two plots are different. The mean and 95% confidence interval are indicated on the box and whisker plot. Pos: *MLL* PTD; wt: AML *MLL* WT

5.3.13.b Correlation of MRD Methods

The MRD status of 10 AML patients, 7 with *MLL* PTD and 3 normal karyotype was monitored longitudinally to assess the value of RQ-PCR for detecting low levels of disease and predicting relapse. The results of the GOI RQ-PCR assays were compared against the standard measures of MRD that were available to each patient (BM morphology, immunophenotyping and cytogenetic analysis) to prove correlation with existing methods, to compare sensitivity and to determine if early detection of relapse was possible. The clinical characteristics of these patients have been discussed previously (Chapter 4, Section 4.2.20). The results of the BM morphology report were used to determine remission with the WHO criteria of <5% blasts defining morphological remission. Relapse (>5% blasts) was also determined from analysis of the BM morphology. Comparison of *CYFIP1* data from RQ-PCR and the standard MRD methods gave broadly concordant results; 11/12 *MLL* PTD follow ups and 2/3 CN AML follow ups moved in parallel to the alternative MRD measurement given. Figure 5.24 shows an example of ΔCp values for *CYFIP1* in comparison to the *MLL* PTD ΔCp values calculated for the same samples. The *CYFIP1* values mirror the results given by *MLL* PTD but have a much lower expression level than *MLL* PTD. So, despite the success in following the trend, the limited dynamic range of the assay in the samples tested makes interpretation difficult and would limit the usefulness of *CYFIP1* as a marker for MRD. The results of the RQ-PCR for *CAV2* were equivocal; 4/7 *MLL* PTD and 0/1 CN AML samples successfully tracked the standard MRD results. As with the *CYFIP1* data, the limited dynamic range made interpretation difficult; the proximity of the positive ΔCp values to those of the healthy donor samples could easily cause a negative sample to be designated positive, and vice versa. *ENPP2* showed more promise as an MRD marker as the RQ-PCR data matched the standard MRD results in all samples (2/2 *MLL* PTD and 2/2 CN AML) yet the sample number was limited. The diagnostic range of *ENPP2* was wider than

either *CYFIP1* or *CAV2*, but narrower than *MLL* PTD, reaching a >1log difference between the *MLL* PTD and healthy control samples. Unfortunately, this larger difference between the ΔCp values was only observed for a subset of the samples tested making it unhelpful as a universal marker. Despite the potential promise of new MRD markers, none of the genes tested were deemed acceptable due to limitations in sensitivity or numbers of target patients.
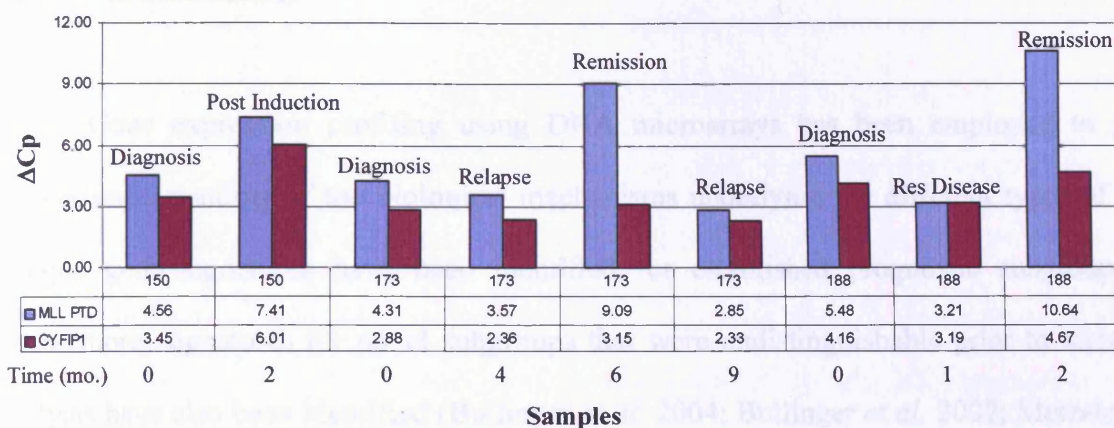
Figure 5.24 Timecourse MRD Analysis of Patients Using CYFIP1 and *MLL* PTD
The expression levels of three patients are for *MLL* PTD (blue) and CYFIP1 (purple) plotted over the time course of treatment. ΔCp is inversely related to gene expression so lower values indicate higher levels of transcript. For a better comparison between the expression levels between the genes, all CYFIP1 values have been multiplied by 2.

150	150	173	173	173	173	188	188	188
MLL PTD 4.56	7.41	4.31	3.57	9.09	2.85	5.48	3.21	10.64
CYFIP1 3.45	6.01	2.88	2.36	3.15	2.33	4.16	3.19	4.67
Time (mo.) 0	2	0	4	6	9	0	1	2

Samples

5-298

5.4 Discussion

Gene expression profiling using DNA microarrays has been employed to gain a greater understanding of the biological mechanisms underlying the different types of AML. Unique gene signatures have been identified for established prognostic subgroups, and furthermore, signatures for novel subgroups that were undistinguishable prior to microarray analysis have also been identified (Bullinger *et al.* 2004; Bullinger *et al.* 2007; Metzeler *et al.* 2008). Cases of AML with *MLL* rearrangements were found to have distinct gene signatures that could be used to discriminate this population, however the *MLL* PTD was seldom included in these studies (Kohlmann *et al.* 2005; Zangrando *et al.* 2009). The *MLL* PTD is exceptional among rearranged *MLL*s because there is no fusion with an external partner gene; instead an internal duplication retains all of the regulatory domains. In this study, a comparison between the gene expression profiles of *MLL* PTD and AML *MLL* WT cases defined a unique gene signature that was then used in conjunction with downstream applications to further our understanding of this disease.

For the first time, a gene signature was defined that separated the *MLL* PTD samples from AML *MLL* PTD WT (Table 5.4., Appendix C). Previous studies were unable to establish a characteristic gene signature, possibly due to smaller cohort sizes analysed, n=20 (Rozovskaia *et al.* 2003) and n=116 (Bullinger *et al.* 2004), compared to 162 in this study. However, one study using a similar number of samples in paediatric AML (n=150) also failed to find discriminating *MLL* PTD genes (Ross *et al.* 2004). Two possible explanations are inherent differences between paediatric and adult AML and differences resulting from PCR sensitivity. The PCR conditions used in the paediatric study produced a 2 fold higher incidence in *MLL* PTD positives than observed in this study, therefore the characteristics of the positive population identified was likely to be different. Nevertheless, molecular heterogeneity in my study was noticeable by the broad PCA clustering of the *MLL* PTD group

in the combined cohort. This loss of unity may be due to the limitations of the Batch Remove programme or to the heterogeneity of the seven samples analysed. In the PCA, samples from each of the GeneChips did not cluster together, appearing inside and outside of the AML *MLL* WT ellipsoid. This indicated the scattered grouping was most likely due to sample heterogeneity and not heterogeneity resulting from differences between the two GeneChips. Larger sample size may help to determine whether this reflected the stages of differentiation and cell types represented by FAB or specific subsets within the *MLL* PTD population.

Analysis of the gene signature demonstrated a difference in the genes regulated in *MLL* PTD AMLs and AMLs with *MLL* translocations. Noticeably absent from the *MLL* PTD significant gene list were genes reported to be up-regulated in *MLL* translocated samples including *MEIS1*, *PBX3*, and members of the *HOX* A cluster (*HOXA4*, *HOXA5*, *HOXA7*, *HOXA9* and *HOXA10*) (Kohlmann *et al.* 2005), with the only exception being the expression of *HOXA10*. This supported previous results observed in a small-scale study done using the GeneChip that was the predecessor to HG-U133A and another in paediatric AML where little to no similarity existed between the gene *MLL* PTD and *MLL* translocations (Rozovskaia *et al.* 2003)(Ross *et al.* 2004). Alternatively, redundancy of function within the *HOX* genes may enable *HOXA10* expression to replace the functions of the other *HOX* genes making this difference inconsequential. The expression of *HOXA10* restored *HOX* function in early haematopoietic progenitors with reduced *HOXA*, B and C gene expression due to the absence of *MLL* (Ernst *et al.* 2004b). Absence of the *HOX* co-factors *MEIS1* and PBX3 make this unlikely. The lack of *HOX* family expression is in contrast to reported upregulation of *HOXA7*, *HOXA9* and *HOXA10* in the BM of *MLL*$^{PTD/WT}$ mice (Dorrance *et al.* 2006). However, the *MLL*$^{PTD/WT}$ mice did not develop leukaemia, so this may represent a pre-leukaemic stage that would require a second mutation for progression and may not be representative of the final gene signature of the disease. The retention of the carboxy terminal end of the protein is the sole consistent difference between *MLL* PTD and the other rearranged

*MLL*s, and this may be a factor causing the difference in expression patterns. The partner genes in translocated *MLL*s replace two important regions in the *MLL* gene, the region containing the transactivating potential of the PHD and the SET domain, which directs histone methylation. The differential regulation of genes supports the hypothesis that the leukaemogeneis of *MLL* PTD operates via a different set of pathways to the other rearranged *MLL*s.

The seven genes identified for validation were reproducibly selected in the significant gene lists of the three sample cohorts illustrated by Venn diagram. Some of the possible reasons for the small number of probe sets (10) identified between the HG-U133A and HG-U133 Plus 2 groups were the different number of samples and the differences between the expression levels of the positives in each cohort. As expected, a higher number of probe sets were found in the intersection between the combined cohort group and the HG-U133A and HG-U133 Plus 2 GeneChips with 32 common probe sets found for each.

Comparison of *MLL* PTD gene signatures in this study identified 7 genes that showed reproducible statistical significance in multiple cohorts. The majority of these genes including CAVI, *CAV2*, *CDKN2C*, *ENPP2*, *TPD52* and *TWIST1*, demonstrate an association with cancer and drug resistance. *CAVI* has been implicated in a number of different types of cancer and may interact with the drug resistance protein MDR1 in diagnostic AML samples (Hatanaka *et al.* 1998; Mercier *et al.* 2009). Expression levels of *CAVI* have been positively correlated with the expression level of the multidrug resistance-1 (MDR-1) gene in diagnostic AML samples. The products for the genes are localised to the plasma membrane and the correlation of gene expression suggests the proteins may interact (Pang, Au, & Kwong 2004). In contrast, this study did not observe significant overexpression of the MDR-1 gene in any of the cohorts studied, so the possibility of this potential interaction in the context of *MLL* PTD is unlikely. *TWIST1* and *ENPP2* are associated with drug resistance in breast cancer, and *TWIST1* is also linked to prometastatic behaviour (Li *et al.* 2009; Vidot *et al.* 2010; Watson *et*

al. 2007). Over expression of *TPD52* has been identified in breast, prostate and ovarian cancers and its use as a potential tumour antigen target in a breast cancer vaccine for mice resulted in enhanced survival (Balleine *et al.* 2000; Byrne *et al.* 2005; Mirshahidi *et al.* 2009; Ummanni *et al.* 2008). The up-regulation of *CDKN2C*, a regulator of cell cycle progression, was not validated by RQ-PCR correlation, so may not be a downstream target of *MLL* PTD. Nevertheless, *CDKN2C* is a downstream target for the *MLL* gene (Milne *et al.* 2005). Gene ontology analysis revealed regulatory functions in cellular processes, biological regulation and the response to stimuli, when the genes from all significant lists were combined. Further exploration of the roles of these genes in a leukaemic setting, ideally primary patient cells, may determine if they are functionally important to the pathogenesis of *MLL* PTD.

Over expression of *WT-1*, a purported marker for AML, was observed in the pathway of interacting proteins (Section 5.3.7). Over-expression of *WT-1* occurs in a high percentage of AML cases (Garg *et al.* 2003b; Menssen *et al.* 1995; Sugiyama 1998) and was also noted in Chapter 3 of this study. Significant over expression of *WT-1* in *MLL* PTD in comparison to the already high level observed in AML has not been seen previously. To determine whether this association is instrumental in the leukaemogenesis of *MLL* PTD would require a greater number of samples and functional studies of the two genes in an *in vitro* or *in vivo* leukaemia model.

An objective of this study was to evaluate the potential of significant genes for use as markers of MRD. RQ-PCR assays compared the ΔCp of *MLL* PTDs, CN AMLs and healthy donors for all GOIs. *CYFIP1* demonstrated a significant up-regulation in all *MLL* PTD samples compared to either CN AML or healthy donors. Expression of the *CYFIP1* gene was assessed for use as an MRD marker along with 8 *MLL* PTD positive samples that exhibited clear separation from the healthy donors in other genes (*CAV2* and *ENPP2*). *CYFIP1* exhibited a low dynamic range that offered no improvement on existing methods, *CAV2* also had a low dynamic range as well as equivocal results and *ENPP2* usefulness appeared to be

limited to a small number of patients. Despite correlation with clinical data, none of the genes were identified as candidate genes for use as MRD markers.

The lack of significant changes in the other genes was due to the similar level of gene expression between the three groups, including *MLL* PTD and the healthy donors. In a similar finding, half of the 21 genes from the paediatric ALL and AML *MLL* translocation GEP were expressed at the same level as normal $CD34^+$ haemopoietic progenitor cells (Ross *et al.* 2004). Examination of the microarray gene expression levels showed the variance between the *MLL* PTD and the AML *MLL* PTD WT patients appeared similar to that observed using RQ-PCR analysis (Section 5.3.13.a). However, there are differences between the two experiments that may account for the disparity; i.e. the use of different samples overall as well as comparison with only the cytogenetically normal subset of AML in the RQ-PCR study.

A variety of methods were used to test the effectiveness of class prediction with the samples from the HG-U133A cohort, HG-U133 Plus2 and combined cohort. All data sets failed to accurately predict *MLL* PTD samples using a programme designed to find the most appropriate prediction model. This was not entirely unexpected, as class prediction also failed in a paediatric AML study where the gene signature was established from a combined group of *MLL* PTD and *MLL* translocation samples (Ross *et al.* 2004). The most likely cause of failure was the small number of positive samples in the population and their heterogeneity of expression.

Potential target areas for therapeutic intervention can be identified by CMAP analysis. A recent publication used CMAP to predict the effectiveness of heat shock protein 90 inhibitors against NOTCH1 mediated T cell leukaemogenesis (Sanda *et al.* 2010). CMAP analysis of the combined cohort identified three potential therapeutic agents that may be beneficial to *MLL* PTD AML patients: tanespimycin, vorinostat and trichostatin A. Tanespimycin is cytotoxic to primary AML cells positive for *FLT3* ITD and has been tested in phase I clinical

trials. If the co-occurrence of *FLT3* ITD with *MLL* PTD influenced the GEP of *MLL* PTD it could explain why tanespimycin is identified in these samples. However, in a study investigating the cytotoxicity of tanespimycin to primary AML cells expressing mutant *FLT3*, a sub-population of *FLT3* WT primary AML samples sensitive to tanespimycin were identified (Al *et al.* 2008). This subset may represent *MLL* PTD AMLs that responded to tanespimycin. Vorinostat and trichostatin A are HDAC inhibitors that have been identified as antineoplastic agents for potential use with AML (Hauswald *et al.* 2009). Phase II clinical trials have been undertaken in older patients with relapsed or refractory AML to determine the efficacy of vorinostat in combination with gemtuzumab ozogamicin, and azacitidine. Activity of HDAC inhibitors against *MLL* PTD leukaemogenesis is supported by functional studies as described below. The amino terminal of the *MLL* gene is essential to the leukaemogenesis of chimeric *MLL*s in mouse models (Lavau *et al.* 1997; Milne *et al.* 2002) and the repression domain necessary for the transformation of bone marrow *in vitro* (Slany, Lavau, & Cleary 1998). HDAC1 mediated repression via the *MLL* repression domain was reduced by the addition of trichostatin A (Xia *et al.* 2003). A more recent report demonstrated hypermethylation of the tumour suppressor gene, SLC5A8 (solute carrier family 5 (iodide transporter) member 5), in *MLL* PTD AML cell lines and subsequent RNA and protein suppression (Whitman *et al.* 2008a). The suppressed protein is a membrane transporter of HDAC inhibitors. Restoration of the SLC5A8 activity via the demethylating agent decitabine followed by treatment with the HDACi valproic acid resulted in increased apoptosis in the *MLL* PTD cell lines (Whitman *et al.* 2008a). Additionally, the HDACi compound increased the susceptibility of *MLL* PTD cells to apoptosis in combination with hypomethylating agents (Whitman *et al.* 2005). In summary, there is clinical and functional evidence to support for the use of the compounds identified by CMAP analysis in the treatment of *MLL* PTD leukaemia. This evidence suggests that HDAC inhibitors (vorinostat and trichostatin A) offer a particularly likely therapeutic avenue.

5.5　Conclusion

The hypothesis that *MLL* PTD has a gene signature unique from other AMLs and can be used to add understanding to the aetiology of the disease was supported by the results of this study. The genes identified on the significant gene list were distinct from those significant genes previously associated with all *MLL* translocations, suggesting that *MLL* PTD leukaemogenesis occurs via different pathways than other forms of *MLL*. While the small sample size and lack of homogeneity within the gene expression profile of the *MLL* PTD samples most likely prevented the class prediction algorithm from accurate classification, the significant gene list derived from the same set of samples identified three potential therapeutic agents that may be beneficial to AML *MLL* PTD patients.

The most promising genes investigated for potential as markers of minimal residual disease showed positive correlation with patient disease level, but ultimately fail for use as suitable therapeutic markers because of the lack of sensitivity or a limited number of target patients.

It would be useful to explore further the potential therapeutic agents identified by means of the *MLL* PTD specific gene signature as a possible improvement on the current treatments for the patients with poor prognosis. The best avenue for further study would be to identify *MLL* PTD patients through association with a large-scale AML trial. This would provide primary patient material from a large number of *MLL* PTD patients which could be used to scale up this study and provide further validation of its findings.

6 General Discussion and Future Directions

6.1   Summary and Relevance of Data

Abnormalities of the *MLL* gene such as recurrent translocations, gene amplification, and PTD aberrations have been reported in AML patients (Poppe *et al.* 2004). The *MLL* PTD is preferentially co-associated in AML patients with trisomy 11 or normal cytogenetics (Caligiuri *et al.* 1994; Caligiuri *et al.* 1998). Use of the duplication as a prognostic marker has been reported either as an indicator of poor outcome or equivocal, although the general consensus is one of poor prognosis (Dohner *et al.* 2002b; Schnittger *et al.* 2000; Steudel *et al.* 2003). On this background, the data presented in Chapter 3 have clearly demonstrated that the duplication is a significant independent prognostic marker capable of identifying a subset of *de novo* AML patients with an increased risk of relapse; the prognostic significance of *MLL* PTD is maintained even after adjustments were made for standard prognostic markers and molecular genetic markers. The *MLL* PTD provided a diagnostic and prognostic indicator for around 5% of patients in the CN-AML subset. Co-occurrence with complementary mutations such as *FLT3* ITD and TKD but not with *NPM1*, provided support for the theory of complementing class I and class II mutations (Introduction, Section 1.2.5) (Gilliland 2002). In addition, the importance of the *MLL* PTD assay was demonstrated since co-operation with complementary mutations such as *FLT3* TKD may have a negative impact on the prognosis. A recent publication confirmed that the occurrence of the *MLL* PTD with *FLT3* TKD negatively influences patient prognosis (Bacher *et al.* 2008).

The increased risk of relapse in *MLL* PTD patients highlighted the need for a strategy to identify mutation positive patients; therefore the potential of RQ-PCR MRD monitoring was investigated (Chapter 4). MRD monitoring of fusion transcripts in *CBF* leukaemias and APL are prognostically useful and have been adopted in clinical practice (Section 1.3). With the high risk of relapse observed in *MLL* PTD patients, a sensitive, accurate and reproducible method of monitoring the patients' therapeutic response would allow for a change in

treatment strategy if a sub-optimal response was detected. This study demonstrated that the duplication was stable upon disease relapse and therefore was an appropriate target for MRD detection. *MLL* PTD positive AML patients who subsequently relapsed had low levels of *MLL* PTD transcript detected while in haematological remission. This persistent residual disease may explain the high level of patient relapse found in this study (Chapter 3). In addition, the proportion of *MLL* PTD patients exhibiting clonal evolution at relapse was greater than published accounts. One explanation is that the *MLL* PTD positive cells may represent a mutator phenotype that have increased genomic instability and are consequently associated with an increased risk of relapse (Alcalay *et al.* 2003). An alternative theory would be that clonal evolution is due to intra-clonal genetic diversity already present in the leukaemic stem cell population at undetectable levels, and the selective pressure of chemotherapy determines the subclone that repopulates the patient BM after treatment (Greaves 2010). The results of the RQ-PCR assay may therefore be beneficial to *MLL* PTD patients who are at an increased risk of relapse, the majority of whom lack a useful marker.

The effect of *MLL* PTD expression on transcription was investigated to gain an insight into potential dysregulated target genes and their downstream signalling pathways through identification of a gene signature. A gene signature unique to the duplication was identified and confirmed using RQ-PCR analysis to validate six genes. CMAP identified three potential chemotherapy compounds that could antagonise the effects of *MLL* PTD expression. Tanespimycin (17AAG) is a *FLT3* ITD inhibitor that functions by inhibiting the protective chaperone activity of the HSP90 leading to degradation of oncogenic proteins in tumour cells (Al *et al.* 2008). 17AAG treatment of a cell line expressing an *MLL-AF4* fusion protein and *FLT3* ITD caused cell cycle inhibition and apoptosis, which may indicate a potential effectiveness in *MLL* mutated AMLs (Yao *et al.* 2003). The other two compounds, trichostatin A and vorinostat, are both HDAC inhibitors which alter transcription through the alteration of chromatin structure by histone acetylation. The *MLL* PTD is associated with

increased histone acetylation in the promoter regions of Hoxa7 and HoxA9 genes and an increase in HSC proliferation in murine models (Dorrance *et al.* 2006). Furthermore, enhanced cell death has been reported to result from the treatment of *MLL* PTD cell lines with a demethylating agent in combination with the HDAC inhibitor, valproate (Appendix) (Whitman *et al.* 2008a). Trichostatin A reduces the activity of the *MLL* repression domain mediated by HDAC interaction (Xia *et al.* 2003). These results support the further investigation of the candidate chemotherapy agents identified. However, my related study on class prediction using gene expression profiling was unsuccessful. Potential reasons for this were the low frequency of *MLL* PTD positive samples in AML patients and the molecular heterogeneity of these samples. This heterogeneity may be due to co-operation between *MLL* PTD and other molecular genetic mutations that were not screened in this study or are not yet discovered.

6.2 Future Work

The above conclusions suggest a number of avenues for further study.

- Investigation of the candidate therapeutic compounds using functional studies in *MLL* PTD cell lines would be the next obvious step forward in this study. Similar studies using other HDAC inhibitors have shown a beneficial impact.

- Investigation into the effects of 17-AAG in primary AML cells found that *FLT3* ITD positive patients were sensitive to the drug but also identified a group of responsive patients that were *FLT3* ITD negative (Al *et al.* 2008). These could be *MLL* PTD positive patients who would potentially benefit from administration of 17-AAG. Functional studies could show that 17AAG has a cytotoxic effect on *MLL* PTD cells in addition to *FLT3* ITD cells. Primary AML samples with *MLL* PTD status ascertained samples could be tested as well as *MLL* PTD positive cell lines.

- Another direction for future work is investigation of the function of the validated genes from the GEP analysis. Confirmation of regulation by *MLL*

PTD could be assessed in a model system in which the effect of *MLL* PTD transfection on candidate target gene expression could be assessed by RQ-PCR. Functional analysis of *MLL* PTD-regulated genes could also be attempted by modulating their expression in an *in vitro* haematopoietic model system through ectopic expression or RNAi technology. If the gene plays a direct role in the leukaemogenesis of *MLL* PTD, the change in its level of expression would be expected to change the characteristics of the cell (e.g. proliferation, survival, and differentiation).

6.3 Broader significance of the results of the thesis

The cumulative results of this study provide a "proof of principle" that the *MLL* PTD is an adverse risk prognostic marker; this finding provides the rationale for larger scale studies. The inclusion of a standardised *MLL* PTD RQ-PCR assay in a large clinical trial would be constructive in furthering understanding of the duplication's prognostic significance. Standardisation of the RQ-PCR would eliminate inconsistencies in the reported frequency of *MLL* PTD, which may be the result of false positives. Prospective identification of a cohort of *MLL* PTD positive patients would provide follow up samples for longitudinal study of the patients' levels of MRD. In addition, the diagnostic characterisation of the patients' *MLL* PTD status could be used to stratify this subset of patients into an adverse risk group to receive specialised treatment. In a similar small-scale study, autologous PBSCT in CR1 enabled 50% (12/24) *MLL* PTD patients to achieve long-term survival, however, 59% still relapsed within 18 months. The outcome of the patients appeared to be influenced by the presence of other molecular markers (Whitman *et al.* 2007). An increase in the number of *MLL* PTD positive samples analysed would increase the confidence of the clinical outcome studies. These would need to be coupled with the identification of other known molecular genetic mutations so that the cumulative prognostic impact could be observed. Furthermore, a greater number of positives might allow GEP analysis on subgroups within the *MLL* PTD

cohort to be analysed, thereby reducing the overall heterogeneity. The ability to analyse a group of *MLL* PTD positive samples that is homogeneous in terms of FAB or WHO classification might improve the likelihood of identifying further gene pathways involved in *MLL* PTD AML by removing the confounding differential expression resulting from analysis of different cell types. However, the difficulties associated with the low incidence of the mutation would remain and any results would have to be assessed to determine if they applied to the MLL PTD cohort as a whole. Greater understanding of *MLL* PTD mediated epigenetic and post-transcriptional dysregulation might also provide insight into the pathobiology of this subgroup. Heterogeneity may be due in part to the effects of global dysregulation within a particular cellular context (Appendix). Likewise, a greater knowledge of patient molecular markers (e.g. *CEBPA*, c-KIT) that could co-occur with *MLL* PTD may prove helpful in stratifying the observed heterogeneity.

The findings presented in this thesis confirm and expand upon the knowledge of the *MLL* PTD in AML. The complexity of the pathogenesis of *MLL* PTD leukaemia dictates the need for more research, and the results of this study provide further important steps towards improving the outcome of this adverse risk group of patients.

7 References

7.1 References

Akashi, K., Traver, D., Miyamoto, T., & Weissman, I. L. 2000, "A clonogenic common myeloid progenitor that gives rise to all myeloid lineages", *Nature*, vol. 404, no. 6774, pp. 193-197.

Al, S. L., Walsby, E., Gilkes, A., Tonks, A., Walsh, V., Mills, K., Burnett, A., & Rowntree, C. 2008, "Heat shock protein 90 inhibition is cytotoxic to primary AML cells expressing mutant *FLT3* and results in altered downstream signalling", *Br J Haematol*, vol. 141, no. 4, pp. 483-493.

Al-Mawali, A., Gillis, D., Hissaria, P., & Lewis, I. 2008, "Incidence, sensitivity, and specificity of leukemia-associated phenotypes in acute myeloid leukemia using specific five-color multiparameter flow cytometry", *Am.J.Clin.Pathol.*, vol. 129, no. 6, pp. 934-945.

Al-Mawali, A., Gillis, D., & Lewis, I. 2009, "The role of multiparameter flow cytometry for detection of minimal residual disease in acute myeloid leukemia", *Am.J.Clin.Pathol.*, vol. 131, no. 1, pp. 16-26.

Alcalay, M., Meani, N., Gelmetti, V., Fantozzi, A., Fagioli, M., Orleth, A., Riganelli, D., Sebastiani, C., Cappelli, E., Casciari, C., Sciurpi, M. T., Mariano, A. R., Minardi, S. P., Luzi, L., Muller, H., Di Fiore, P. P., Frosina, G., & Pelicci, P. G. 2003, "Acute myeloid leukemia fusion proteins deregulate genes involved in stem cell maintenance and DNA repair", *J Clin Invest*, vol. 112, no. 11, pp. 1751-1761.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. 1990, "Basic local alignment search tool", *J.Mol.Biol.*, vol. 215, no. 3, pp. 403-410.

Antonchuk, J., Sauvageau, G., & Humphries, R. K. 2001, "*HOX*B4 overexpression mediates very rapid stem cell regeneration and competitive hematopoietic repopulation", *Exp.Hematol.*, vol. 29, no. 9, pp. 1125-1134.

Arai, F., Hirao, A., Ohmura, M., Sato, H., Matsuoka, S., Takubo, K., Ito, K., Koh, G. Y., & Suda, T. 2004, "Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche", *Cell*, vol. 118, no. 2, pp. 149-161.

Argiropoulos, B. & Humphries, R. K. 2007, "Hox genes in hematopoiesis and leukemogenesis", *Oncogene*, vol. 26, no. 47, pp. 6766-6776.

Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R., & Korsmeyer, S. J. 2002, "MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia", *Nat.Genet.*, vol. 30, no. 1, pp. 41-47.

Ayton, P. M. & Cleary, M. L. 2001, "Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins", *Oncogene*, vol. 20, no. 40, pp. 5695-5707.

Ayton, P. M. & Cleary, M. L. 2003, "Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9", *Genes Dev.*, vol. 17, no. 18, pp. 2298-2307.

Bacher, U., Kern, W., Schnittger, S., Hiddemann, W., Haferlach, T., & Schoch, C. 2005, "Population-based age-specific incidences of cytogenetic subgroups of acute myeloid leukemia", *Haematologica*, vol. 90, no. 11, pp. 1502-1510.

Bacher, U., Haferlach, T., Kern, W., Haferlach, C., & Schnittger, S. 2007a, "A comparative study of molecular mutations in 381 patients with myelodysplastic syndrome and in 4130 patients with acute myeloid leukemia", *Haematologica*, vol. 92, no. 6, pp. 744-752.

Bacher, U., Haferlach, C., Kern, W., Haferlach, T., & Schnittger, S. 2008, "Prognostic relevance of FLT3-TKD mutations in AML: the combination matters--an analysis of 3082 patients", *Blood*, vol. 111, no. 5, pp. 2527-2537.

Baldus, C. D. & Bullinger, L. 2008, "Gene expression with prognostic implications in cytogenetically normal acute myeloid leukemia", *Semin Oncol*, vol. 35, no. 4, pp. 356-364.

Baldus, C. D., Thiede, C., Soucek, S., Bloomfield, C. D., Thiel, E., & Ehninger, G. 2006, "BAALC expression and FLT3 internal tandem duplication mutations in acute myeloid leukemia patients with normal cytogenetics: prognostic implications", *J.Clin.Oncol.*, vol. 24, no. 5, pp. 790-797.

Balleine, R. L., Fejzo, M. S., Sathasivam, P., Basset, P., Clarke, C. L., & Byrne, J. A. 2000, "The hD52 (*TPD52*) gene is a candidate target gene for events resulting in increased 8q21 copy number in human breast carcinoma", *Genes Chromosomes.Cancer*, vol. 29, no. 1, pp. 48-57.

Barragan, E., Pajuelo, J. C., Ballester, S., Fuster, O., Cervera, J., Moscardo, F., Senent, L., Such, E., Sanz, M. A., & Bolufer, P. 2008, "Minimal residual disease detection in acute myeloid leukemia by mutant nucleophosmin (*NPM1*): comparison with *WT-1* gene expression", *Clin Chim.Acta*, vol. 395, no. 1-2, pp. 120-123.

Basecke, J., Cepek, L., Mannhalter, C., Krauter, J., Hildenhagen, S., Brittinger, G., Trumper, L., & Griesinger, F. 2002, "Transcription of *AML1/ETO* in bone marrow and cord blood of individuals without acute myelogenous leukemia", *Blood*, vol. 100, no. 6, pp. 2267-2268.

Basecke, J., Podleschny, M., Clemens, R., Schnittger, S., Viereck, V., Trumper, L., & Griesinger, F. 2006, "Lifelong persistence of AML associated MLL partial tandem duplications (MLL-PTD) in healthy adults", *Leuk.Res.*, vol. 30, no. 9, pp. 1091-1096.

Basecke, J., Whelan, J. T., Griesinger, F., & Bertrand, F. E. 2006c, "The MLL partial tandem duplication in acute myeloid leukaemia", *Br.J.Haematol.*, vol. 135, no. 4, pp. 438-449.

Beillard, E., Pallisgaard, N., Van, D. V., V, Bi, W., Dee, R., van der, S. E., Delabesse, E., Macintyre, E., Gottardi, E., Saglio, G., Watzinger, F., Lion, T., van Dongen, J. J., Hokland, P., & Gabert, J. 2003, "Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using 'real-time' quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) - a Europe against cancer program", *Leukemia*, vol. 17, no. 12, pp. 2474-2486.

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R., & Sultan, C. 1976, "Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group", *Br J Haematol*, vol. 33, no. 4, pp. 451-458.

Beran, M., Luthra, R., Kantarjian, H., & Estey, E. 2004, "*FLT3* mutation and response to intensive chemotherapy in young adult and elderly patients with normal karyotype", *Leuk.Res.*, vol. 28, no. 6, pp. 547-550.

Bereshchenko, O., Mancini, E., Moore, S., Bilbao, D., Mansson, R., Luc, S., Grover, A., Jacobsen, S. E., Bryder, D., & Nerlov, C. 2009, "Hematopoietic stem cell expansion precedes the generation of committed myeloid leukemia-initiating cells in C/EBPalpha mutant AML", *Cancer Cell*, vol. 16, no. 5, pp. 390-400.

Bernard, O. A., Romana, S. P., Schichman, S. A., Mauchauffe, M., Jonveaux, P., & Berger, R. 1995, "Partial duplication of HRX in acute leukemia with trisomy 11", *Leukemia*, vol. 9, no. 9, pp. 1487-1490.

Bertrand, F. E., Spengeman, J. D., Shah, N., & LeBien, T. W. 2003, "B-cell development in the presence of the MLL/AF4 oncoprotein proceeds in the absence of *HOX* A7 and *HOX* A9 expression", *Leukemia*, vol. 17, no. 12, pp. 2454-2459.

Blalock, W. L., Weinstein-Oppenheimer, C., Chang, F., Hoyle, P. E., Wang, X. Y., Algate, P. A., Franklin, R. A., Oberhaus, S. M., Steelman, L. S., & McCubrey, J. A. 1999, "Signal transduction, cell cycle regulatory, and anti-apoptotic pathways regulated by IL-3 in hematopoietic cells: possible sites for intervention with anti-neoplastic drugs", *Leukemia*, vol. 13, no. 8, pp. 1109-1166.

Bloomfield, C. D., Lawrence, D., Byrd, J. C., Carroll, A., Pettenati, M. J., Tantravahi, R., Patil, S. R., Davey, F. R., Berg, D. T., Schiffer, C. A., Arthur, D. C., & Mayer, R. J. 1998, "Frequency of prolonged remission duration after high-dose cytarabine intensification in acute myeloid leukemia varies by cytogenetic subtype", *Cancer Res.*, vol. 58, no. 18, pp. 4173-4179.

Boissel, N., Renneville, A., Biggio, V., Philippe, N., Thomas, X., Cayuela, J. M., Terre, C., Tigaud, I., Castaigne, S., Raffoux, E., de, B. S., Fenaux, P., Dombret, H., & Preudhomme, C. 2005, "Prevalence, clinical profile, and prognosis of NPM mutations in AML with normal karyotype", *Blood*, vol. 106, no. 10, pp. 3618-3620.

Bonnet, D. & Dick, J. E. 1997, "Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell", *Nat.Med.*, vol. 3, no. 7, pp. 730-737.

Bose, S., Deininger, M., Gora-Tybor, J., Goldman, J. M., & Melo, J. V. 1998, "The presence of typical and atypical *BCR-ABL* fusion genes in leukocytes of normal individuals: biologic significance and implications for the assessment of minimal residual disease", *Blood*, vol. 92, no. 9, pp. 3362-3367.

Brun, A. C., Bjornsson, J. M., Magnusson, M., Larsson, N., Leveen, P., Ehinger, M., Nilsson, E., & Karlsson, S. 2004, "Hoxb4-deficient mice undergo normal hematopoietic development but exhibit a mild proliferation defect in hematopoietic stem cells", *Blood*, vol. 103, no. 11, pp. 4126-4133.

Bullinger, L., Dohner, K., Bair, E., Frohling, S., Schlenk, R. F., Tibshirani, R., Dohner, H., & Pollack, J. R. 2004, "Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia", *N.Engl.J.Med.*, vol. 350, no. 16, pp. 1605-1616.

Bullinger, L., Rucker, F. G., Kurz, S., Du, J., Scholl, C., Sander, S., Corbacioglu, A., Lottaz, C., Krauter, J., Frohling, S., Ganser, A., Schlenk, R. F., Dohner, K., Pollack, J. R., & Dohner,

H. 2007, "Gene-expression profiling identifies distinct subclasses of core binding factor acute myeloid leukemia", *Blood*, vol. 110, no. 4, pp. 1291-1300.

Buonamici, S., Ottaviani, E., Testoni, N., Montefusco, V., Visani, G., Bonifazi, F., Amabile, M., Terragna, C., Ruggeri, D., Piccaluga, P. P., Isidori, A., Malagola, M., Baccarani, M., Tura, S., & Martinelli, G. 2002, "Real-time quantitation of minimal residual disease in inv(16)-positive acute myeloid leukemia may indicate risk for clinical relapse and may identify patients in a curable state", *Blood*, vol. 99, no. 2, pp. 443-449.

Burnett, A. K., Goldstone, A. H., Stevens, R. M., Hann, I. M., Rees, J. K., Gray, R. G., & Wheatley, K. 1998, "Randomised comparison of addition of autologous bone-marrow transplantation to intensive chemotherapy for acute myeloid leukaemia in first remission: results of MRC AML 10 trial. UK Medical Research Council Adult and Children's Leukaemia Working Parties", *Lancet*, vol. 351, no. 9104, pp. 700-708.

Butler, L. H., Slany, R., Cui, X., Cleary, M. L., & Mason, D. Y. 1997, "The HRX proto-oncogene product is widely expressed in human tissues and localizes to nuclear structures", *Blood*, vol. 89, no. 9, pp. 3361-3370.

Byrd, J. C., Mrozek, K., Dodge, R. K., Carroll, A. J., Edwards, C. G., Arthur, D. C., Pettenati, M. J., Patil, S. R., Rao, K. W., Watson, M. S., Koduru, P. R., Moore, J. O., Stone, R. M., Mayer, R. J., Feldman, E. J., Davey, F. R., Schiffer, C. A., Larson, R. A., & Bloomfield, C. D. 2002, "Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461)", *Blood*, vol. 100, no. 13, pp. 4325-4336.

Byrne, J. A., Balleine, R. L., Schoenberg, F. M., Mercieca, J., Chiew, Y. E., Livnat, Y., St, H. L., Peters, G. B., Byth, K., Karlan, B. Y., Slamon, D. J., Harnett, P., & Defazio, A. 2005, "Tumor protein D52 (*TPD52*) is overexpressed and a gene amplification target in ovarian cancer", *Int.J.Cancer*, vol. 117, no. 6, pp. 1049-1054.

Caligiuri, M. A., Schichman, S. A., Strout, M. P., Mrozek, K., Baer, M. R., Frankel, S. R., Barcos, M., Herzig, G. P., Croce, C. M., & Bloomfield, C. D. 1994, "Molecular rearrangement of the ALL-1 gene in acute myeloid leukemia without cytogenetic evidence of 11q23 chromosomal translocations", *Cancer Res.*, vol. 54, no. 2, pp. 370-373.

Caligiuri, M. A., Strout, M. P., Lawrence, D., Arthur, D. C., Baer, M. R., Yu, F., Knuutila, S., Mrozek, K., Oberkircher, A. R., Marcucci, G., de la, C. A., Elonen, E., Block, A. W., Rao, P. N., Herzig, G. P., Powell, B. L., Ruutu, T., Schiffer, C. A., & Bloomfield, C. D. 1998, "Rearrangement of ALL1 (MLL) in acute myeloid leukemia with normal cytogenetics", *Cancer Res.*, vol. 58, no. 1, pp. 55-59.

Caligiuri, M. A., Strout, M. P., Oberkircher, A. R., Yu, F., de la, C. A., & Bloomfield, C. D. 1997, "The partial tandem duplication of ALL1 in acute myeloid leukemia with normal cytogenetics or trisomy 11 is restricted to one chromosome", *Proc.Natl.Acad.Sci.U.S.A*, vol. 94, no. 8, pp. 3899-3902.

Caligiuri, M. A., Strout, M. P., Schichman, S. A., Mrozek, K., Arthur, D. C., Herzig, G. P., Baer, M. R., Schiffer, C. A., Heinonen, K., Knuutila, S., Nousiainen, T., Ruutu, T., Block, A. W., Schulman, P., Pedersen-Bjergaard, J., Croce, C. M., & Bloomfield, C. D. 1996, "Partial tandem duplication of ALL1 as a recurrent molecular defect in acute myeloid leukemia with trisomy 11", *Cancer Res.*, vol. 56, no. 6, pp. 1418-1425.

Calvi, L. M., Adams, G. B., Weibrecht, K. W., Weber, J. M., Olson, D. P., Knight, M. C., Martin, R. P., Schipani, E., Divieti, P., Bringhurst, F. R., Milner, L. A., Kronenberg, H. M., & Scadden, D. T. 2003, "Osteoblastic cells regulate the haematopoietic stem cell niche", *Nature*, vol. 425, no. 6960, pp. 841-846.

Campana, D. & Coustan-Smith, E. 1999, "Detection of minimal residual disease in acute leukemia by flow cytometry", *Cytometry*, vol. 38, no. 4, pp. 139-152.

Campana, D. & Pui, C. H. 1995, "Detection of minimal residual disease in acute leukemia: methodologic advances and clinical significance", *Blood*, vol. 85, no. 6, pp. 1416-1434.

Cantor, A. B., Iwasaki, H., Arinobu, Y., Moran, T. B., Shigematsu, H., Sullivan, M. R., Akashi, K., & Orkin, S. H. 2008, "Antagonism of *FOG*-1 and GATA factors in fate choice for the mast cell lineage", *The Journal of Experimental Medicine*, vol. 205, no. 3, pp. 611-624.

Carpenter, B., Hill, K. J., Charalambous, M., Wagner, K. J., Lahiri, D., James, D. I., Andersen, J. S., Schumacher, V., Royer-Pokora, B., Mann, M., Ward, A., & Roberts, S. G. 2004, "*BASP1* is a transcriptional cosuppressor for the Wilms' tumor suppressor protein *WT-1*", *Mol.Cell Biol.*, vol. 24, no. 2, pp. 537-549.

Cassileth, P. A., Harrington, D. P., Appelbaum, F. R., Lazarus, H. M., Rowe, J. M., Paietta, E., Willman, C., Hurd, D. D., Bennett, J. M., Blume, K. G., Head, D. R., & Wiernik, P. H. 1998, "Chemotherapy compared with autologous or allogeneic bone marrow transplantation in the management of acute myeloid leukemia in first remission", *N.Engl.J.Med.*, vol. 339, no. 23, pp. 1649-1656.

Chen, H., Ray-Gallet, D., Zhang, P., Hetherington, C. J., Gonzalez, D. A., Zhang, D. E., Moreau-Gachelin, F., & Tenen, D. G. 1995, "*PU.1* (Spi-1) autoregulates its expression in myeloid cells", *Oncogene*, vol. 11, no. 8, pp. 1549-1560.

Cheson, B. D., Bennett, J. M., Kopecky, K. J., Buchner, T., Willman, C. L., Estey, E. H., Schiffer, C. A., Doehner, H., Tallman, M. S., Lister, T. A., Lo-Coco, F., Willemze, R., Biondi, A., Hiddemann, W., Larson, R. A., Lowenberg, B., Sanz, M. A., Head, D. R., Ohno, R., & Bloomfield, C. D. 2003, "Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia", *J.Clin.Oncol.*, vol. 21, no. 24, pp. 4642-4649.

Chillon, M. C., Fernandez, C., Garcia-Sanz, R., Balanzategui, A., Ramos, F., Fernandez-Calvo, J., Gonzalez, M., & Miguel, J. F. 2004, "*FLT3*-activating mutations are associated with poor prognostic features in AML at diagnosis but they are not an independent prognostic factor", *Hematol.J.*, vol. 5, no. 3, pp. 239-246.

Chou, Q., Russell, M., Birch, D. E., Raymond, J., & Bloch, W. 1992, "Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications", *Nucleic Acids Res*, vol. 20, no. 7, pp. 1717-1723.

Chou, W. C., Tang, J. L., Wu, S. J., Tsay, W., Yao, M., Huang, S. Y., Huang, K. C., Chen, C. Y., Huang, C. F., & Tien, H. F. 2007, "Clinical implications of minimal residual disease monitoring by quantitative polymerase chain reaction in acute myeloid leukemia patients bearing nucleophosmin (*NPM1*) mutations", *Leukemia*, vol. 21, no. 5, pp. 998-1004.

Chou, W. C., Tang, J. L., Lin, L. I., Yao, M., Tsay, W., Chen, C. Y., Wu, S. J., Huang, C. F., Chiou, R. J., Tseng, M. H., Lin, D. T., Lin, K. H., Chen, Y. C., & Tien, H. F. 2006, "Nucleophosmin Mutations in De novo Acute Myeloid Leukemia: The Age-Dependent Incidences and the Stability during Disease Evolution", *Cancer Research*, vol. 66, no. 6, pp. 3310-3316.

Christiansen, D. H. & Pedersen-Bjergaard, J. 2001, "Internal tandem duplications of the *FLT3* and MLL genes are mainly observed in atypical cases of therapy-related acute myeloid leukemia with a normal karyotype and are unrelated to type of previous therapy", *Leukemia*, vol. 15, no. 12, pp. 1848-1851.

Cilloni, D., Renneville, A., Hermitte, F., Hills, R. K., Daly, S., Jovanovic, J. V., Gottardi, E., Fava, M., Schnittger, S., Weiss, T., Izzo, B., Nomdedeu, J., van der, H. A., van der Reijden, B. A., Jansen, J. H., Van, D. V., V, Ommen, H., Preudhomme, C., Saglio, G., & Grimwade, D. 2009, "Real-time quantitative polymerase chain reaction detection of minimal residual disease by standardized *WT-1* assay to enhance risk stratification in acute myeloid leukemia: a European LeukemiaNet study", *Journal of Clinical Oncology*, vol. 27, no. 31, pp. 5195-5201.

Cimino, G., Moir, D. T., Canaani, O., Williams, K., Crist, W. M., Katzav, S., Cannizzaro, L., Lange, B., Nowell, P. C., Croce, C. M., & . 1991, "Cloning of ALL-1, the locus involved in leukemias with the t(4;11)(q21;q23), t(9;11)(p22;q23), and t(11;19)(q23;p13) chromosome translocations", *Cancer Res.*, vol. 51, no. 24, pp. 6712-6714.

Cleary, M. L. 1999, "A new angle on a pervasive oncogene", *Nat.Genet.*, vol. 23, no. 2, pp. 134-135.

Cline, M. J. 1994, "The molecular basis of leukemia", *N.Engl.J.Med.*, vol. 330, no. 5, pp. 328-336.

Cornelissen, J. J., van Putten, W. L., Verdonck, L. F., Theobald, M., Jacky, E., Daenen, S. M., van Marwijk, K. M., Wijermans, P., Schouten, H., Huijgens, P. C., van der, L. H., Fey, M., Ferrant, A., Maertens, J., Gratwohl, A., & Lowenberg, B. 2007, "Results of a HOVON/SAKK donor versus no-donor analysis of myeloablative HLA-identical sibling stem cell transplantation in first remission acute myeloid leukemia in young and middle-aged adults: benefits for whom?", *Blood*, vol. 109, no. 9, pp. 3658-3666.

Costello, R., Sainty, D., Blaise, D., Gastaut, J. A., Gabert, J., Poirel, H., Buzyn-Veil, A., & Macintyre, E. 1997, "Prognosis value of residual disease monitoring by polymerase chain reaction in patients with *CBF* beta/*MYH11*-positive acute myeloblastic leukemia", *Blood*, vol. 89, no. 6, pp. 2222-2223.

Cozzio, A., Passegue, E., Ayton, P. M., Karsunky, H., Cleary, M. L., & Weissman, I. L. 2003, "Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors", *Genes Dev.*, vol. 17, no. 24, pp. 3029-3035.

Dash, A. & Gilliland, D. G. 2001, "Molecular genetics of acute myeloid leukaemia", *Best.Pract.Res.Clin.Haematol.*, vol. 14, no. 1, pp. 49-64.

De, M. R., Testa, U., Luchetti, L., Zeuner, A., Stassi, G., Pelosi, E., Riccioni, R., Felli, N., Samoggia, P., & Peschle, C. 1999, "Apoptotic role of Fas/Fas ligand system in the regulation of erythropoiesis", *Blood*, vol. 93, no. 3, pp. 796-803.

Debernardi, S., Lillington, D. M., Chaplin, T., Tomlinson, S., Amess, J., Rohatiner, A., Lister, T. A., & Young, B. D. 2003, "Genome-wide analysis of acute myeloid leukemia with normal karyotype reveals a unique pattern of homeobox gene expression distinct from those with translocation-mediated fusion events", Genes Chromosomes.Cancer, vol. 37, no. 2, pp. 149-158.

Dennis, G., Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., & Lempicki, R. A. 2003, "DAVID: Database for Annotation, Visualization, and Integrated Discovery", Genome Biol., vol. 4, no. 5, p. 3.

DePinho, R. A. 2000, "The age of cancer", Nature, vol. 408, no. 6809, pp. 248-254.

Dimartino, J. F. & Cleary, M. L. 1999, "Mll rearrangements in haematological malignancies: lessons from clinical and biological studies", Br.J.Haematol., vol. 106, no. 3, pp. 614-626.

Diverio, D., Rossi, V., Avvisati, G., De, S. S., Pistilli, A., Pane, F., Saglio, G., Martinelli, G., Petti, M. C., Santoro, A., Pelicci, P. G., Mandelli, F., Biondi, A., & Lo, C. F. 1998, "Early detection of relapse by prospective reverse transcriptase-polymerase chain reaction analysis of the PML/RARalpha fusion gene in patients with acute promyelocytic leukemia enrolled in the GIMEMA-AIEOP multicenter "AIDA" trial. GIMEMA-AIEOP Multicenter "AIDA" Trial", Blood, vol. 92, no. 3, pp. 784-789.

Dobson, C. L., Warren, A. J., Pannell, R., Forster, A., & Rabbitts, T. H. 2000, "Tumorigenesis in mice with a fusion of the leukaemia oncogene Mll and the bacterial lacZ gene", EMBO J., vol. 19, no. 5, pp. 843-851.

Dohner, K., Tobis, K., Ulrich, R., Frohling, S., Benner, A., Schlenk, R. F., & Dohner, H. 2002, "Prognostic significance of partial tandem duplications of the MLL gene in adult patients 16 to 60 years old with acute myeloid leukemia and normal cytogenetics: a study of the Acute Myeloid Leukemia Study Group Ulm", Journal of Clinical Oncology, vol. 20, no. 15, pp. 3254-3261.

Dohner, K., Schlenk, R. F., Habdank, M., Scholl, C., Rucker, F. G., Corbacioglu, A., Bullinger, L., Frohling, S., & Dohner, H. 2005, "Mutant nucleophosmin (NPM1) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations", Blood, vol. 106, no. 12, pp. 3740-3746.

Dolken, G. 2001, "Detection of minimal residual disease", Adv.Cancer Res., vol. 82, pp. 133-185.

Dorrance, A. M., Liu, S., Yuan, W., Becknell, B., Arnoczky, K. J., Guimond, M., Strout, M. P., Feng, L., Nakamura, T., Yu, L., Rush, L. J., Weinstein, M., Leone, G., Wu, L., Ferketich, A., Whitman, S. P., Marcucci, G., & Caligiuri, M. A. 2006, "Mll partial tandem duplication induces aberrant Hox expression in vivo via specific epigenetic alterations", J.Clin.Invest, vol. 116, no. 10, pp. 2707-2716.

Downing, J. R. & Look, A. T. 1996, "MLL fusion genes in the 11q23 acute leukemias", Cancer Treat.Res., vol. 84, pp. 73-92.

Duprez, E., Benoit, G., Flexor, M., Lillehaug, J. R., & Lanotte, M. 2000, "A mutated PML/RARA found in the retinoid maturation resistant NB4 subclone, NB4-R2, blocks RARA and wild-type PML/RARA transcriptional activities", Leukemia, vol. 14, no. 2, pp. 255-261.

Emig, M., Saussele, S., Wittor, H., Weisser, A., Reiter, A., Willer, A., Berger, U., Hehlmann, R., Cross, N. C., & Hochhaus, A. 1999, "Accurate and rapid analysis of residual disease in patients with CML using specific fluorescent hybridization probes for real time quantitative RT-PCR", *Leukemia*, vol. 13, no. 11, pp. 1825-1832.

Ernst, P., Fisher, J. K., Avery, W., Wade, S., Foy, D., & Korsmeyer, S. J. 2004a, "Definitive hematopoiesis requires the mixed-lineage leukemia gene", *Dev.Cell*, vol. 6, no. 3, pp. 437-443.

Ernst, P., Mabon, M., Davidson, A. J., Zon, L. I., & Korsmeyer, S. J. 2004b, "An Mll-dependent Hox program drives hematopoietic progenitor expansion", *Curr Biol.*, vol. 14, no. 22, pp. 2063-2069.

Esteve, J., Escoda, L., Martin, G., Rubio, V., az-Mediavilla, J., Gonzalez, M., Rivas, C., Alvarez, C., Gonzalez San Miguel, J. D., Brunet, S., Tomas, J. F., Tormo, M., Sayas, M. J., Sanchez, G. P., Colomer, D., Bolufer, P., & Sanz, M. A. 2007, "Outcome of patients with acute promyelocytic leukemia failing to front-line treatment with all-trans retinoic acid and anthracycline-based chemotherapy (PETHEMA protocols LPA96 and LPA99): benefit of an early intervention", *Leukemia*, vol. 21, no. 3, pp. 446-452.

Estey, E., deLima, M., Strom, S., Pierce, S., Freireich, E. J., & Keating, M. J. 1997, "Long-term follow-up of patients with newly diagnosed acute myeloid leukemia treated at the University of Texas M.D. Anderson Cancer Center", *Cancer*, vol. 80, no. 11 Suppl, pp. 2176-2180.

Estey, E. & Dohner, H. 2006, "Acute myeloid leukaemia", *Lancet*, vol. 368, no. 9550, pp. 1894-1907.

Estey, E., Keating, M. J., Pierce, S., & Stass, S. 1995, "Change in karyotype between diagnosis and first relapse in acute myelogenous leukemia", *Leukemia*, vol. 9, no. 6, pp. 972-976.

Falini, B., Bolli, N., Shan, J., Martelli, M. P., Liso, A., Pucciarini, A., Bigerna, B., Pasqualucci, L., Mannucci, R., Rosati, R., Gorello, P., Diverio, D., Roti, G., Tiacci, E., Cazzaniga, G., Biondi, A., Schnittger, S., Haferlach, T., Hiddemann, W., Martelli, M. F., Gu, W., Mecucci, C., & Nicoletti, I. 2006, "Both carboxy-terminus NES motif and mutated tryptophan(s) are crucial for aberrant nuclear export of nucleophosmin leukemic mutants in NPMc+ AML", *Blood*, vol. 107, no. 11, pp. 4514-4523.

Falini, B., Mecucci, C., Tiacci, E., Alcalay, M., Rosati, R., Pasqualucci, L., La, S. R., Diverio, D., Colombo, E., Santucci, A., Bigerna, B., Pacini, R., Pucciarini, A., Liso, A., Vignetti, M., Fazi, P., Meani, N., Pettirossi, V., Saglio, G., Mandelli, F., Lo-Coco, F., Pelicci, P. G., & Martelli, M. F. 2005, "Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype", *N.Engl.J.Med.*, vol. 352, no. 3, pp. 254-266.

Fandy, T. E., Carraway, H., & Gore, S. D. 2007, "DNA demethylating agents and histone deacetylase inhibitors in hematologic malignancies", *Cancer J.*, vol. 13, no. 1, pp. 40-48.

Farag, S. S., Ruppert, A. S., Mrozek, K., Mayer, R. J., Stone, R. M., Carroll, A. J., Powell, B. L., Moore, J. O., Pettenati, M. J., Koduru, P. R., Stamberg, J., Baer, M. R., Block, A. W., Vardiman, J. W., Kolitz, J. E., Schiffer, C. A., Larson, R. A., & Bloomfield, C. D. 2005, "Outcome of induction and postremission therapy in younger adults with acute myeloid

leukemia with normal karyotype: a cancer and leukemia group B study", *J.Clin.Oncol.*, vol. 23, no. 3, pp. 482-493.

Fialkow, P. J., Singer, J. W., Adamson, J. W., Vaidya, K., Dow, L. W., Ochs, J., & Moohr, J. W. 1981, "Acute nonlymphocytic leukemia: heterogeneity of stem cell origin", *Blood*, vol. 57, no. 6, pp. 1068-1073.

Ford, C. E., Hamerton, J. L., Barnes, D. W., & Loutit, J. F. 1956, "Cytological identification of radiation-chimaeras", *Nature*, vol. 177, no. 4506, pp. 452-454.

Freeman, S. D., Jovanovic, J. V., & Grimwade, D. 2008, "Development of minimal residual disease-directed therapy in acute myeloid leukemia", *Semin.Oncol*, vol. 35, no. 4, pp. 388-400.

Frohling, S., Schlenk, R. F., Breitruck, J., Benner, A., Kreitmeier, S., Tobis, K., Dohner, H., & Dohner, K. 2002, "Prognostic significance of activating *FLT3* mutations in younger adults (16 to 60 years) with acute myeloid leukemia and normal cytogenetics: a study of the AML Study Group Ulm", *Blood*, vol. 100, no. 13, pp. 4372-4380.

Gabert, J., Beillard, E., Van, D. V., V, Bi, W., Grimwade, D., Pallisgaard, N., Barbany, G., Cazzaniga, G., Cayuela, J. M., Cave, H., Pane, F., Aerts, J. L., De Micheli, D., Thirion, X., Pradel, V., Gonzalez, M., Viehmann, S., Malec, M., Saglio, G., & van Dongen, J. J. 2003, "Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia - a Europe Against Cancer program", *Leukemia*, vol. 17, no. 12, pp. 2318-2357.

Gale, R. E., Green, C., Allen, C., Mead, A. J., Burnett, A. K., Hills, R. K., & Linch, D. C. 2008, "The impact of *FLT3* internal tandem duplication mutant level, number, size, and interaction with *NPM1* mutations in a large cohort of young adult patients with acute myeloid leukemia", *Blood*, vol. 111, no. 5, pp. 2776-2784.

Gale, R. E., Hills, R., Kottaridis, P. D., Srirangan, S., Wheatley, K., Burnett, A. K., & Linch, D. C. 2005, "No evidence that FLT3 status should be considered as an indicator for transplantation in acute myeloid leukemia (AML): an analysis of 1135 patients, excluding acute promyelocytic leukemia, from the UK MRC AML10 and 12 trials", *Blood*, vol. 106, no. 10, pp. 3658-3665.

Gallagher, R. E., Yeap, B. Y., Bi, W., Livak, K. J., Beaubier, N., Rao, S., Bloomfield, C. D., Appelbaum, F. R., Tallman, M. S., Slack, J. L., & Willman, C. L. 2003, "Quantitative real-time RT-PCR analysis of PML-RAR alpha mRNA in acute promyelocytic leukemia: assessment of prognostic significance in adult patients from intergroup protocol 0129", *Blood*, vol. 101, no. 7, pp. 2521-2528.

Garg, M., Moore, H., Tobal, K., & Liu Yin, J. A. 2003, "Prognostic significance of quantitative analysis of *WT-1* gene transcripts by competitive reverse transcription polymerase chain reaction in acute leukaemia", *Br J Haematol*, vol. 123, no. 1, pp. 49-59.

Garson, O. M., Hagemeijer, A., Sakurai, M., Reeves, B. R., Swansbury, G. J., Williams, G. J., Alimena, G., Arthur, D. C., Berger, R., de la, C. A., & . 1989, "Cytogenetic studies of 103 patients with acute myelogenous leukemia in relapse", *Cancer Genet.Cytogenet.*, vol. 40, no. 2, pp. 187-202.

Gelmetti, V., Zhang, J., Fanelli, M., Minucci, S., Pelicci, P. G., & Lazar, M. A. 1998, "Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner *ETO*", *Mol.Cell Biol.*, vol. 18, no. 12, pp. 7185-7191.

Gilliland, D. G. 2002, "Molecular genetics of human leukemias: new insights into therapy", *Semin.Hematol.*, vol. 39, no. 4 Suppl 3, pp. 6-11.

Gilliland, D. G. 2003, "Proteolytic processing in development and leukemogenesis", *Cell*, vol. 115, no. 3, pp. 248-250.

Gilliland, D. G. & Griffin, J. D. 2002, "The roles of *FLT3* in hematopoiesis and leukemia", *Blood*, vol. 100, no. 5, pp. 1532-1542.

Goldstone, A. H., Burnett, A. K., Wheatley, K., Smith, A. G., Hutchinson, R. M., & Clark, R. E. 2001, "Attempts to improve treatment outcomes in acute myeloid leukemia (AML) in older patients: the results of the United Kingdom Medical Research Council AML11 trial", *Blood*, vol. 98, no. 5, pp. 1302-1311.

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., & Lander, E. S. 1999, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring", *Science*, vol. 286, no. 5439, pp. 531-537.

Gorello, P., Cazzaniga, G., Alberti, F., Dell'Oro, M. G., Gottardi, E., Specchia, G., Roti, G., Rosati, R., Martelli, M. F., Diverio, D., Lo, C. F., Biondi, A., Saglio, G., Mecucci, C., & Falini, B. 2006, "Quantitative assessment of minimal residual disease in acute myeloid leukemia carrying nucleophosmin (*NPM1*) gene mutations", *Leukemia*, vol. 20, no. 6, pp. 1103-1108.

Greaves, M. 2010, "Cancer stem cells: back to Darwin?", *Semin.Cancer Biol.*, vol. 20, no. 2, pp. 65-70.

Greene, M. E., Mundschau, G., Wechsler, J., McDevitt, M., Gamis, A., Karp, J., Gurbuxani, S., Arceci, R., & Crispino, J. D. 2003, "Mutations in GATA1 in both transient myeloproliferative disorder and acute megakaryoblastic leukemia of Down syndrome", *Blood Cells Mol.Dis.*, vol. 31, no. 3, pp. 351-356.

Grignani, F., De, M. S., Nervi, C., Tomassoni, L., Gelmetti, V., Cioce, M., Fanelli, M., Ruthardt, M., Ferrara, F. F., Zamir, I., Seiser, C., Grignani, F., Lazar, M. A., Minucci, S., & Pelicci, P. G. 1998, "Fusion proteins of the retinoic acid receptor-alpha recruit histone deacetylase in promyelocytic leukaemia", *Nature*, vol. 391, no. 6669, pp. 815-818.

Grimwade, D. 2001, "The clinical significance of cytogenetic abnormalities in acute myeloid leukaemia", *Best.Pract.Res.Clin.Haematol.*, vol. 14, no. 3, pp. 497-529.

Grimwade, D. & Hills, R. K. 2009, "Independent prognostic factors for AML outcome", *Hematology.Am Soc.Hematol Educ.Program.* pp. 385-395.

Grimwade, D., Hills, R. K., Moorman, A. V., Walker, H., Chatters, S., Goldstone, A. H., Wheatley, K., Harrison, C. J., & Burnett, A. K. 2010, "Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials", *Blood*, vol. 116, no. 3, pp. 354-365.

Grimwade, D. & Lo, C. F. 2002, "Acute promyelocytic leukemia: a model for the role of molecular diagnosis and residual disease monitoring in directing treatment approach in acute myeloid leukemia", *Leukemia*, vol. 16, no. 10, pp. 1959-1973.

Grimwade, D., Walker, H., Oliver, F., Wheatley, K., Harrison, C., Harrison, G., Rees, J., Hann, I., Stevens, R., Burnett, A., & Goldstone, A. 1998, "The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties", *Blood*, vol. 92, no. 7, pp. 2322-2333.

Grimwade, D., Jovanovic, J. V., Hills, R. K., Nugent, E. A., Patel, Y., Flora, R., Diverio, D., Jones, K., Aslett, H., Batson, E., Rennie, K., Angell, R., Clark, R. E., Solomon, E., Lo-Coco, F., Wheatley, K., & Burnett, A. K. 2009, "Prospective minimal residual disease monitoring to predict relapse of acute promyelocytic leukemia and to direct pre-emptive arsenic trioxide therapy", *J.Clin.Oncol.*, vol. 27, no. 22, pp. 3650-3658.

Grundler, R., Miething, C., Thiede, C., Peschel, C., & Duyster, J. 2005, "*FLT3*-ITD and tyrosine kinase domain mutants induce 2 distinct phenotypes in a murine bone marrow transplantation model", *Blood*, vol. 105, no. 12, pp. 4792-4799.

Guan, Y., Gerhard, B., & Hogge, D. E. 2003, "Detection, isolation, and stimulation of quiescent primitive leukemic progenitor cells from patients with acute myeloid leukemia (AML)", *Blood*, vol. 101, no. 8, pp. 3142-3149.

Gudgin.E. 2009, "Acute Myeloid Leukemia biology and leukemia stem cells", *European Haematology Association* pp. 17-23.

Guinn, B. A., Gilkes, A. F., Woodward, E., Westwood, N. B., Mufti, G. J., Linch, D., Burnett, A. K., & Mills, K. I. 2005, "Microarray analysis of tumour antigen expression in presentation acute myeloid leukaemia", *Biochem.Biophys.Res Commun.*, vol. 333, no. 3, pp. 703-713.

Gulley, M. L., Shea, T. C., & Fedoriw, Y. 2010, "Genetic tests to evaluate prognosis and predict therapeutic response in acute myeloid leukemia", *J.Mol.Diagn.*, vol. 12, no. 1, pp. 3-16.

Guzman, M. L., Neering, S. J., Upchurch, D., Grimes, B., Howard, D. S., Rizzieri, D. A., Luger, S. M., & Jordan, C. T. 2001, "Nuclear factor-kappaB is constitutively activated in primitive human acute myelogenous leukemia cells", *Blood*, vol. 98, no. 8, pp. 2301-2307.

Haferlach, T., Kohlmann, A., Schnittger, S., Dugas, M., Hiddemann, W., Kern, W., & Schoch, C. 2005, "Global approach to the diagnosis of leukemia using gene expression profiling", *Blood*, vol. 106, no. 4, pp. 1189-1198.

Hanahan, D. & Weinberg, R. A. 2000, "The hallmarks of cancer", *Cell*, vol. 100, no. 1, pp. 57-70.

Harper, D. P. & Aplan, P. D. 2008, "Chromosomal Rearrangements Leading to MLL Gene Fusions: Clinical and Biological Aspects", *Cancer Research*, vol. 68, no. 24, pp. 10024-10027.

Hassan, H. T. & Zander, A. 1996, "Stem cell factor as a survival and growth factor in human normal and malignant hematopoiesis", *Acta Haematol*, vol. 95, no. 3-4, pp. 257-262.

Hatanaka, M., Maeda, T., Ikemoto, T., Mori, H., Seya, T., & Shimizu, A. 1998, "Expression of caveolin-1 in human T cell leukemia cell lines", *Biochem.Biophys.Res Commun.*, vol. 253, no. 2, pp. 382-387.

Hauswald, S., Duque-Afonso, J., Wagner, M. M., Schertl, F. M., Lubbert, M., Peschel, C., Keller, U., & Licht, T. 2009, "Histone deacetylase inhibitors induce a very broad, pleiotropic anticancer drug resistance phenotype in acute myeloid leukemia cells by modulation of multiple ABC transporter genes", *Clin.Cancer Res.*, vol. 15, no. 11, pp. 3705-3715.

Heath, V., Suh, H. C., Holman, M., Renn, K., Gooya, J. M., Parkin, S., Klarmann, K. D., Ortiz, M., Johnson, P., & Keller, J. 2004, "C/EBPalpha deficiency results in hyperproliferation of hematopoietic progenitor cells and disrupts macrophage development in vitro and in vivo", *Blood*, vol. 104, no. 6, pp. 1639-1647.

Heinonen, K., Mrozek, K., Lawrence, D., Arthur, D. C., Pettenati, M. J., Stamberg, J., Qumsiyeh, M. B., Verma, R. S., MacCallum, J., Schiffer, C. A., & Bloomfield, C. D. 1998, "Clinical characteristics of patients with de novo acute myeloid leukaemia and isolated trisomy 11: a Cancer and Leukemia Group B study", *Br.J.Haematol.*, vol. 101, no. 3, pp. 513-520.

Heissig, B., Hattori, K., Dias, S., Friedrich, M., Ferris, B., Hackett, N. R., Crystal, R. G., Besmer, P., Lyden, D., Moore, M. A., Werb, Z., & Rafii, S. 2002, "Recruitment of stem and progenitor cells from the bone marrow niche requires MMP-9 mediated release of kit-ligand", *Cell*, vol. 109, no. 5, pp. 625-637.

Hibbert, L. & Johnston, J. A. 2001, "Cytokine signalling and disease", *Expert.Opin.Ther.Targets.*, vol. 5, no. 6, pp. 641-653.

Hokland, P. & Pallisgaard, N. 2000, "Integration of molecular methods for detection of balanced translocations in the diagnosis and follow-up of patients with leukemia", *Semin.Hematol.*, vol. 37, no. 4, pp. 358-367.

Hope, K. J., Jin, L., & Dick, J. E. 2004, "Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity", *Nat.Immunol.*, vol. 5, no. 7, pp. 738-743.

Hsieh, J. J., Cheng, E. H., & Korsmeyer, S. J. 2003, "Taspase1: a threonine aspartase required for cleavage of MLL and proper *HOX* gene expression", *Cell*, vol. 115, no. 3, pp. 293-303.

Hsieh, J. J., Ernst, P., Erdjument-Bromage, H., Tempst, P., & Korsmeyer, S. J. 2003, "Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization", *Mol Cell Biol.*, vol. 23, no. 1, pp. 186-194.

Huang, d. W., Sherman, B. T., & Lempicki, R. A. 2009, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources", *Nat.Protoc.*, vol. 4, no. 1, pp. 44-57.

Hughes, T. P., Kaeda, J., Branford, S., Rudzki, Z., Hochhaus, A., Hensley, M. L., Gathmann, I., Bolton, A. E., van Hoomissen, I. C., Goldman, J. M., & Radich, J. P. 2003, "Frequency of major molecular responses to imatinib or interferon alfa plus cytarabine in newly diagnosed chronic myeloid leukemia", *N.Engl.J.Med.*, vol. 349, no. 15, pp. 1423-1432.

Hughes, T., Deininger, M., Hochhaus, A., Branford, S., Radich, J., Kaeda, J., Baccarani, M., Cortes, J., Cross, N. C., Druker, B. J., Gabert, J., Grimwade, D., Hehlmann, R., Kamel-Reid, S., Lipton, J. H., Longtine, J., Martinelli, G., Saglio, G., Soverini, S., Stock, W., & Goldman, J. M. 2006, "Monitoring CML patients responding to treatment with tyrosine kinase inhibitors: review and recommendations for harmonizing current methodology for detecting BCR-ABL transcripts and kinase domain mutations and for expressing results", *Blood*, vol. 108, no. 1, pp. 28-37.

Huntly, B. J., Shigematsu, H., Deguchi, K., Lee, B. H., Mizuno, S., Duclos, N., Rowan, R., Amaral, S., Curley, D., Williams, I. R., Akashi, K., & Gilliland, D. G. 2004, "MOZ-TIF2, but not *BCR-ABL*, confers properties of leukemic stem cells to committed murine hematopoietic progenitors", *Cancer Cell*, vol. 6, no. 6, pp. 587-596.

Ihle, J. N. & Kerr, I. M. 1995, "Jaks and Stats in signaling by the cytokine receptor superfamily", *Trends Genet.*, vol. 11, no. 2, pp. 69-74.

Ishikawa, Y., Kiyoi, H., Tsujimura, A., Miyawaki, S., Miyazaki, Y., Kuriyama, K., Tomonaga, M., & Naoe, T. 2009, "Comprehensive analysis of cooperative gene mutations between class I and class II in de novo acute myeloid leukemia", *Eur.J.Haematol.*, vol. 83, no. 2, pp. 90-98.

Iwasaki, H., Mizuno, S., Wells, R. A., Cantor, A. B., Watanabe, S., & Akashi, K. 2003, "*GATA-1* converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages", *Immunity.*, vol. 19, no. 3, pp. 451-462.

Jabbour, E. J., Estey, E., & Kantarjian, H. M. 2006, "Adult acute myeloid leukemia", *Mayo Clin.Proc.*, vol. 81, no. 2, pp. 247-260.

Jaeger, U. & Kainz, B. 2003, "Monitoring minimal residual disease in AML: the right time for real time", *Ann.Hematol*, vol. 82, no. 3, pp. 139-147.

Johansen, L. M., Iwama, A., Lodie, T. A., Sasaki, K., Felsher, D. W., Golub, T. R., & Tenen, D. G. 2001, "c-Myc is a critical target for c/EBPalpha in granulopoiesis", *Mol.Cell Biol.*, vol. 21, no. 11, pp. 3789-3806.

Jude, C. D., Climer, L., Xu, D., Artinger, E., Fisher, J. K., & Ernst, P. 2007, "Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors", *Cell Stem Cell*, vol. 1, no. 3, pp. 324-337.

Jurlander, J., Caligiuri, M. A., Ruutu, T., Baer, M. R., Strout, M. P., Oberkircher, A. R., Hoffmann, L., Ball, E. D., Frei-Lahr, D. A., Christiansen, N. P., Block, A. M., Knuutila, S., Herzig, G. P., & Bloomfield, C. D. 1996, "Persistence of the *AML1/ETO* fusion transcript in patients treated with allogeneic bone marrow transplantation for t(8;21) leukemia", *Blood*, vol. 88, no. 6, pp. 2183-2191.

Kagoshima, H., Shigesada, K., Satake, M., Ito, Y., Miyoshi, H., Ohki, M., Pepling, M., & Gergen, P. 1993, "The Runt domain identifies a new family of heteromeric transcriptional regulators", *Trends Genet.*, vol. 9, no. 10, pp. 338-341.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., & Hirakawa, M. 2010, "KEGG for representation and analysis of molecular networks involving diseases and drugs", *Nucleic Acids Res.*, vol. 38, no. Database issue, p. D355-D360.

Karp, C. L., Grupe, A., Schadt, E., Ewart, S. L., Keane-Moore, M., Cuomo, P. J., Kohl, J., Wahl, L., Kuperman, D., Germer, S., Aud, D., Peltz, G., & Wills-Karp, M. 2000, "Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma", *Nat.Immunol.*, vol. 1, no. 3, pp. 221-226.

Kern, W., Haferlach, C., Haferlach, T., & Schnittger, S. 2008, "Monitoring of minimal residual disease in acute myeloid leukemia", *Cancer*, vol. 112, no. 1, pp. 4-16.

Kern, W., Haferlach, T., Schnittger, S., Ludwig, W. D., Hiddemann, W., & Schoch, C. 2002, "Karyotype instability between diagnosis and relapse in 117 patients with acute myeloid leukemia: implications for resistance against therapy", *Leukemia*, vol. 16, no. 10, pp. 2084-2091.

Kern, W. & Schnittger, S. 2003, "Monitoring of acute myeloid leukemia by flow cytometry", *Curr.Oncol.Rep.*, vol. 5, no. 5, pp. 405-412.

Kern, W., Voskova, D., Schoch, C., Schnittger, S., Hiddemann, W., & Haferlach, T. 2004, "Prognostic impact of early response to induction therapy as assessed by multiparameter flow cytometry in acute myeloid leukemia", *Haematologica*, vol. 89, no. 5, pp. 528-540.

Kiel, M. J., Radice, G. L., & Morrison, S. J. 2007, "Lack of evidence that hematopoietic stem cells depend on N-cadherin-mediated adhesion to osteoblasts for their maintenance", *Cell Stem Cell*, vol. 1, no. 2, pp. 204-217.

Kiel, M. J., Yilmaz, O. H., Iwashita, T., Yilmaz, O. H., Terhorst, C., & Morrison, S. J. 2005, "SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells", *Cell*, vol. 121, no. 7, pp. 1109-1121.

Kirstetter, P., Schuster, M. B., Bereshchenko, O., Moore, S., Dvinge, H., Kurz, E., Theilgaard-Monch, K., Mansson, R., Pedersen, T. A., Pabst, T., Schrock, E., Porse, B. T., Jacobsen, S. E., Bertone, P., Tenen, D. G., & Nerlov, C. 2008, "Modeling of C/EBPalpha mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells", *Cancer Cell*, vol. 13, no. 4, pp. 299-310.

Kiyoi, H., Naoe, T., Nakano, Y., Yokota, S., Minami, S., Miyawaki, S., Asou, N., Kuriyama, K., Jinnai, I., Shimazaki, C., Akiyama, H., Saito, K., Oh, H., Motoji, T., Omoto, E., Saito, H., Ohno, R., & Ueda, R. 1999, "Prognostic implication of *FLT3* and N-RAS gene mutations in acute myeloid leukemia", *Blood*, vol. 93, no. 9, pp. 3074-3080.

Knapper, S. 2007, "*FLT3* inhibition in acute myeloid leukaemia", *Br J Haematol*, vol. 138, no. 6, pp. 687-699.

Knapper, S., Burnett, A. K., Littlewood, T., Kell, W. J., Agrawal, S., Chopra, R., Clark, R., Levis, M. J., & Small, D. 2006, "A phase 2 trial of the *FLT3* inhibitor lestaurtinib (CEP701) as first-line treatment for older patients with acute myeloid leukemia not considered fit for intensive chemotherapy", *Blood*, vol. 108, no. 10, pp. 3262-3270.

Knudson, A. G., Jr. 1971, "Mutation and cancer: statistical study of retinoblastoma", *Proc.Natl.Acad.Sci.U.S.A*, vol. 68, no. 4, pp. 820-823.

Kohlmann, A., Kipps, T. J., Rassenti, L. Z., Downing, J. R., Shurtleff, S. A., Mills, K. I., Gilkes, A. F., Hofmann, W. K., Basso, G., Dell'orto, M. C., Foa, R., Chiaretti, S., de, V. J., Rauhut, S., Papenhausen, P. R., Hernandez, J. M., Lumbreras, E., Yeoh, A. E., Koay, E. S., Li, R., Liu, W. M., Williams, P. M., Wieczorek, L., & Haferlach, T. 2008, "An international standardization programme towards the application of gene expression profiling in routine leukaemia diagnostics: the Microarray Innovations in LEukemia study prephase", *Br.J.Haematol.*, vol. 142, no. 5, pp. 802-807.

Kohlmann, A., Schoch, C., Dugas, M., Schnittger, S., Hiddemann, W., Kern, W., & Haferlach, T. 2005, "New insights into MLL gene rearranged acute leukemias using gene expression profiling: shared pathways, lineage commitment, and partner genes", *Leukemia*, vol. 19, no. 6, pp. 953-964.

Kollet, O., Dar, A., Shivtiel, S., Kalinkovich, A., Lapid, K., Sztainberg, Y., Tesio, M., Samstein, R. M., Goichberg, P., Spiegel, A., Elson, A., & Lapidot, T. 2006, "Osteoclasts degrade endosteal components and promote mobilization of hematopoietic progenitor cells", *Nat.Med.*, vol. 12, no. 6, pp. 657-664.

Kondo, M., Weissman, I. L., & Akashi, K. 1997, "Identification of clonogenic common lymphoid progenitors in mouse bone marrow", *Cell*, vol. 91, no. 5, pp. 661-672.

Kottaridis, P. D., Gale, R. E., Frew, M. E., Harrison, G., Langabeer, S. E., Belton, A. A., Walker, H., Wheatley, K., Bowen, D. T., Burnett, A. K., Goldstone, A. H., & Linch, D. C. 2001, "The presence of a *FLT3* internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials", *Blood*, vol. 98, no. 6, pp. 1752-1759.

Kottaridis, P. D., Gale, R. E., Langabeer, S. E., Frew, M. E., Bowen, D. T., & Linch, D. C. 2002, "Studies of *FLT3* mutations in paired presentation and relapse samples from patients with acute myeloid leukemia: implications for the role of *FLT3* mutations in leukemogenesis, minimal residual disease detection, and possible therapy with *FLT3* inhibitors", *Blood*, vol. 100, no. 7, pp. 2393-2398.

Kottaridis, P. D., Gale, R. E., & Linch, D. C. 2003, "Flt3 mutations and leukaemia", *Br.J.Haematol.*, vol. 122, no. 4, pp. 523-538.

Kramer, M. F. & Coen, D. M. 2001, "Enzymatic amplification of DNA by PCR: standard procedures and optimization", *Curr.Protoc.Immunol.*, Chapter 10.

Krauter, J., Gorlich, K., Ottmann, O., Lubbert, M., Dohner, H., Heit, W., Kanz, L., Ganser, A., & Heil, G. 2003, "Prognostic value of minimal residual disease quantification by real-time reverse transcriptase polymerase chain reaction in patients with core binding factor leukemias", *Journal of Clinical Oncology*, vol. 21, no. 23, pp. 4413-4422.

Krivtsov, A. V., Twomey, D., Feng, Z., Stubbs, M. C., Wang, Y., Faber, J., Levine, J. E., Wang, J., Hahn, W. C., Gilliland, D. G., Golub, T. R., & Armstrong, S. A. 2006, "Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9", *Nature*, vol. 442, no. 7104, pp. 818-822.

Krumlauf, R. 1994, "Hox genes in vertebrate development", *Cell*, vol. 78, no. 2, pp. 191-201.

Ku, H., Yonemura, Y., Kaushansky, K., & Ogawa, M. 1996, "Thrombopoietin, the ligand for the Mpl receptor, synergizes with steel factor and other early acting cytokines in supporting proliferation of primitive hematopoietic progenitors of mice", *Blood*, vol. 87, no. 11, pp. 4544-4551.

Kumar, A. R., Hudson, W. A., Chen, W., Nishiuchi, R., Yao, Q., & Kersey, J. H. 2004, "Hoxa9 influences the phenotype but not the incidence of Mll-AF9 fusion gene leukemia", *Blood*, vol. 103, no. 5, pp. 1823-1828.

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J. P., Subramanian, A., Ross, K. N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S., & Golub, T. R. 2006, "The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease", *Science*, vol. 313, no. 5795, pp. 1929-1935.

Lange, B. J., Kobrinsky, N., Barnard, D. R., Arthur, D. C., Buckley, J. D., Howells, W. B., Gold, S., Sanders, J., Neudorf, S., Smith, F. O., & Woods, W. G. 1998, "Distinctive demography, biology, and outcome of acute myeloid leukemia and myelodysplastic syndrome in children with Down syndrome: Children's Cancer Group Studies 2861 and 2891", *Blood*, vol. 91, no. 2, pp. 608-615.

Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M. A., & Dick, J. E. 1994, "A cell initiating human acute myeloid leukaemia after transplantation into SCID mice", *Nature*, vol. 367, no. 6464, pp. 645-648.

Lapillonne, H., Renneville, A., Auvrignon, A., Flamant, C., Blaise, A., Perot, C., Lai, J. L., Ballerini, P., Mazingue, F., Fasola, S., Dehee, A., Bellman, F., Adam, M., Labopin, M., Douay, L., Leverger, G., Preudhomme, C., & Landman-Parker, J. 2006, "High *WT-1* expression after induction therapy predicts high risk of relapse and death in pediatric acute myeloid leukemia", *J.Clin.Oncol.*, vol. 24, no. 10, pp. 1507-1515.

Lavau, C., Szilvassy, S. J., Slany, R., & Cleary, M. L. 1997, "Immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced HRX-*ENL*", *EMBO J.*, vol. 16, no. 14, pp. 4226-4237.

Lee, B. H., Tothova, Z., Levine, R. L., Anderson, K., Buza-Vidas, N., Cullen, D. E., McDowell, E. P., Adelsperger, J., Frohling, S., Huntly, B. J., Beran, M., Jacobsen, S. E., & Gilliland, D. G. 2007, "*FLT3* Mutations Confer Enhanced Proliferation and Survival Properties to Multipotent Progenitors in a Murine Model of Chronic Myelomonocytic Leukemia", *Cancer Cell*, vol. 12, no. 4, pp. 367-380.

Leroy, H., de, B. S., Grardel-Duflos, N., Darre, S., Leleu, X., Roumier, C., Morschhauser, F., Lai, J. L., Bauters, F., Fenaux, P., & Preudhomme, C. 2005, "Prognostic value of real-time quantitative PCR (RQ-PCR) in AML with t(8;21)", *Leukemia*, vol. 19, no. 3, pp. 367-372.

Li, Q. Q., Xu, J. D., Wang, W. J., Cao, X. X., Chen, Q., Tang, F., Chen, Z. Q., Liu, X. P., & Xu, Z. D. 2009, "Twist1-mediated adriamycin-induced epithelial-mesenchymal transition relates to multidrug resistance and invasive potential in breast cancer cells", *Clin.Cancer Res.*, vol. 15, no. 8, pp. 2657-2665.

Libura, M., Asnafi, V., Tu, A., Delabesse, E., Tigaud, I., Cymbalista, F., naceur-Griscelli, A., Villarese, P., Solbu, G., Hagemeijer, A., Beldjord, K., Hermine, O., & Macintyre, E. 2003,

"*FLT3* and MLL intragenic abnormalities in AML reflect a common category of genotoxic stress", *Blood*, vol. 102, no. 6, pp. 2198-2204.

Lion, T. 1996a, "Appropriate controls for RT-PCR", *Leukemia*, vol. 10, no. 11, p. 1843.

Lion, T. 1996b, "Control genes in reverse transcriptase-polymerase chain reaction assays", *Leukemia*, vol. 10, no. 9, pp. 1527-1528.

Liu Yin, J. A. 2002b, "Minimal residual disease in acute myeloid leukaemia", *Best.Pract.Res.Clin.Haematol.*, vol. 15, no. 1, pp. 119-135.

Liu Yin, J. A. 2002a, "Minimal residual disease in acute myeloid leukaemia", *Best.Pract.Res.Clin.Haematol.*, vol. 15, no. 1, pp. 119-135.

Liu-Yin, J., Sale, M., Daly, S., Hills, R., Wheatley, K., & Burnett, A. K. 2008, "Predictive Value of Minimal Residual Disease (MRD) Monitoring by RQ-PCR in *WT-1* Positive Patients Entered in the UK MRC AML-15 Trial", *ASH Annual Meeting Abstracts*, vol. 112, no. 11, p. 697.

Lo, C. F., Diverio, D., Avvisati, G., Petti, M. C., Meloni, G., Pogliani, E. M., Biondi, A., Rossi, G., Carlo-Stella, C., Selleri, C., Martino, B., Specchia, G., & Mandelli, F. 1999, "Therapy of molecular relapse in acute promyelocytic leukemia", *Blood*, vol. 94, no. 7, pp. 2225-2229.

Lo, C. F., Diverio, D., Falini, B., Biondi, A., Nervi, C., & Pelicci, P. G. 1999c, "Genetic diagnosis and molecular monitoring in the management of acute promyelocytic leukemia", *Blood*, vol. 94, no. 1, pp. 12-22.

Lo-Coco, F. & Ammatuna, E. 2006, "The biology of acute promyelocytic leukemia and its impact on diagnosis and treatment", *Hematology Am Soc Hematol Educ Program* pp. 156-161.

Look, A. T. 1997, "Oncogenic transcription factors in the human acute leukemias", *Science*, vol. 278, no. 5340, pp. 1059-1064.

Lyman, S. D., Brasel, K., Rousseau, A. M., & Williams, D. E. 1994, "The flt3 ligand: a hematopoietic stem cell factor whose activities are distinct from steel factor", *Stem Cells*, vol. 12 Suppl 1, pp. 99-107.

Lyman, S. D., James, L., Zappone, J., Sleath, P. R., Beckmann, M. P., & Bird, T. 1993, "Characterization of the protein encoded by the flt3 (flk2) receptor-like tyrosine kinase gene", *Oncogene*, vol. 8, no. 4, pp. 815-822.

Magic, Z., Supic, G., & Brankovic-Magic, M. 2009, "Towards targeted epigenetic therapy of cancer", *J.BUON.*, vol. 14 Suppl 1, p. S79-S88.

Mahadevan, D. & List, A. F. 2004, "Targeting the multidrug resistance-1 transporter in AML: molecular regulation and therapeutic strategies", *Blood*, vol. 104, no. 7, pp. 1940-1951.

Marcucci, G., Strout, M. P., Bloomfield, C. D., & Caligiuri, M. A. 1998, "Detection of unique ALL1 (MLL) fusion transcripts in normal human bone marrow and blood: distinct origin of normal versus leukemic ALL1 fusion transcripts", *Cancer Res.*, vol. 58, no. 4, pp. 790-793.

Mardis, E. R. & Wilson, R. K. 2009, "Cancer genome sequencing: a review", *Hum.Mol Genet.*, vol. 18, no. R2, p. R163-R168.

Marks, P., Rifkind, R. A., Richon, V. M., Breslow, R., Miller, T., & Kelly, W. K. 2001, "Histone deacetylases and cancer: causes and therapies", *Nat.Rev.Cancer*, vol. 1, no. 3, pp. 194-202.

Martin, G., Barragan, E., Bolufer, P., Chillon, C., Garcia-Sanz, R., Gomez, T., Brunet, S., Gonzalez, M., & Sanz, M. A. 2000, "Relevance of presenting white blood cell count and kinetics of molecular remission in the prognosis of acute myeloid leukemia with CBFbeta/*MYH11* rearrangement", *Haematologica*, vol. 85, no. 7, pp. 699-703.

Martin, M. E., Milne, T. A., Bloyer, S., Galoian, K., Shen, W., Gibbs, D., Brock, H. W., Slany, R., & Hess, J. L. 2003, "Dimerization of MLL fusion proteins immortalizes hematopoietic cells", *Cancer Cell*, vol. 4, no. 3, pp. 197-207.

McCULLOCH, E. A., Howatson, A. F., Buick, R. N., Minden, M. D., & Izaguirre, C. A. 1979, "Acute myeloblastic leukemia considered as a clonal hemopathy", *Blood Cells*, vol. 5, no. 2, pp. 261-282.

McKenna, H. J., Stocking, K. L., Miller, R. E., Brasel, K., De, S. T., Maraskovsky, E., Maliszewski, C. R., Lynch, D. H., Smith, J., Pulendran, B., Roux, E. R., Teepe, M., Lyman, S. D., & Peschon, J. J. 2000, "Mice lacking flt3 ligand have deficient hematopoiesis affecting hematopoietic progenitor cells, dendritic cells, and natural killer cells", *Blood*, vol. 95, no. 11, pp. 3489-3497.

McMahon, K. A., Hiew, S. Y., Hadjur, S., Veiga-Fernandes, H., Menzel, U., Price, A. J., Kioussis, D., Williams, O., & Brady, H. J. 2007, "Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal", *Cell Stem Cell*, vol. 1, no. 3, pp. 338-345.

Mead, A. J., Linch, D. C., Hills, R. K., Wheatley, K., Burnett, A. K., & Gale, R. E. 2007, "*FLT3* tyrosine kinase domain mutations are biologically distinct from and have a significantly more favorable prognosis than *FLT3* internal tandem duplications in patients with acute myeloid leukemia", *Blood*, vol. 110, no. 4, pp. 1262-1270.

Menssen, H. D., Renkl, H. J., Rodeck, U., Maurer, J., Notter, M., Schwartz, S., Reinhardt, R., & Thiel, E. 1995, "Presence of Wilms' tumor gene (wt1) transcripts and the *WT-1* nuclear protein in the majority of human acute leukemias", *Leukemia*, vol. 9, no. 6, pp. 1060-1067.

Mercier, I., Casimiro, M. C., Zhou, J., Wang, C., Plymire, C., Bryant, K. G., Daumer, K. M., Sotgia, F., Bonuccelli, G., Witkiewicz, A. K., Lin, J., Tran, T. H., Milliman, J., Frank, P. G., Jasmin, J. F., Rui, H., Pestell, R. G., & Lisanti, M. P. 2009, "Genetic Ablation of Caveolin-1 Drives Estrogen-Hypersensitivity and the Development of DCIS-Like Mammary Lesions", *Am J Pathol*, vol. 174, no. 4, pp. 1172-1190.

Metzeler, K. H., Hummel, M., Bloomfield, C. D., Spiekermann, K., Braess, J., Sauerland, M. C., Heinecke, A., Radmacher, M., Marcucci, G., Whitman, S. P., Maharry, K., Paschka, P., Larson, R. A., Berdel, W. E., Buchner, T., Wormann, B., Mansmann, U., Hiddemann, W., Bohlander, S. K., & Buske, C. 2008, "An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia", *Blood*, vol. 112, no. 10, pp. 4193-4201.

Miller, J., Horner, A., Stacy, T., Lowrey, C., Lian, J. B., Stein, G., Nuckolls, G. H., & Speck, N. A. 2002, "The core-binding factor beta subunit is required for bone formation and hematopoietic maturation", *Nat.Genet.*, vol. 32, no. 4, pp. 645-649.

Milne, T. A., Briggs, S. D., Brock, H. W., Martin, M. E., Gibbs, D., Allis, C. D., & Hess, J. L. 2002, "MLL targets SET domain methyltransferase activity to Hox gene promoters", *Mol Cell*, vol. 10, no. 5, pp. 1107-1117.

Milne, T. A., Hughes, C. M., Lloyd, R., Yang, Z., Rozenblatt-Rosen, O., Dou, Y., Schnepp, R. W., Krankel, C., Livolsi, V. A., Gibbs, D., Hua, X., Roeder, R. G., Meyerson, M., & Hess, J. L. 2005, "Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors", *Proc.Natl.Acad.Sci.U.S.A*, vol. 102, no. 3, pp. 749-754.

Mirshahidi, S., Kramer, V. G., Whitney, J. B., Essono, S., Lee, S., Dranoff, G., Anderson, K. S., & Ruprecht, R. M. 2009, "Overlapping synthetic peptides encoding *TPD52* as breast cancer vaccine in mice: prolonged survival", *Vaccine*, vol. 27, no. 12, pp. 1825-1833.

Miyamoto, T., Weissman, I. L., & Akashi, K. 2000, "*AML1/ETO*-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation", *Proc.Natl.Acad.Sci.U.S.A*, vol. 97, no. 13, pp. 7521-7526.

Mrozek, K., Heinonen, K., & Bloomfield, C. D. 2001, "Clinical importance of cytogenetics in acute myeloid leukaemia", *Best.Pract.Res.Clin.Haematol.*, vol. 14, no. 1, pp. 19-47.

Mrozek, K., Heinonen, K., Lawrence, D., Carroll, A. J., Koduru, P. R., Rao, K. W., Strout, M. P., Hutchison, R. E., Moore, J. O., Mayer, R. J., Schiffer, C. A., & Bloomfield, C. D. 1997, "Adult patients with de novo acute myeloid leukemia and t(9; 11)(p22; q23) have a superior outcome to patients with other translocations involving band 11q23: a cancer and leukemia group B study", *Blood*, vol. 90, no. 11, pp. 4532-4538.

Mrozek, K., Marcucci, G., Paschka, P., Whitman, S. P., & Bloomfield, C. D. 2007, "Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: are we ready for a prognostically prioritized molecular classification?", *Blood*, vol. 109, no. 2, pp. 431-448.

Mrozek, K., Radmacher, M. D., Bloomfield, C. D., & Marcucci, G. 2009, "Molecular signatures in acute myeloid leukemia", *Curr.Opin.Hematol.*, vol. 16, no. 2, pp. 64-69.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. 1986, "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction", *Cold Spring Harb.Symp.Quant.Biol.*, vol. 51 Pt 1, pp. 263-273.

Munoz, L., Nomdedeu, J. F., Villamor, N., Guardia, R., Colomer, D., Ribera, J. M., Torres, J. P., Berlanga, J. J., Fernandez, C., Llorente, A., Queipo de Llano, M. P., Sanchez, J. M., Brunet, S., & Sierra, J. 2003, "Acute myeloid leukemia with MLL rearrangements: clinicobiological features, prognostic impact and value of flow cytometry in the detection of residual leukemic cells", *Leukemia*, vol. 17, no. 1, pp. 76-82.

Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C. M., & Canaani, E. 2002, "ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation", *Mol.Cell*, vol. 10, no. 5, pp. 1119-1128.

Nakano, Y., Kiyoi, H., Miyawaki, S., Asou, N., Ohno, R., Saito, H., & Naoe, T. 1999, "Molecular evolution of acute myeloid leukaemia in relapse: unstable N-ras and *FLT3* genes compared with p53 gene", *Br J Haematol*, vol. 104, no. 4, pp. 659-664.

Nakao, M., Yokota, S., Iwai, T., Kaneko, H., Horiike, S., Kashima, K., Sonoda, Y., Fujimoto, T., & Misawa, S. 1996, "Internal tandem duplication of the flt3 gene found in acute myeloid leukemia", *Leukemia*, vol. 10, no. 12, pp. 1911-1918.

Nerlov, C. 2004, "C/EBPalpha mutations in acute myeloid leukaemias", *Nat.Rev.Cancer*, vol. 4, no. 5, pp. 394-400.

Nerlov, C., Querfurth, E., Kulessa, H., & Graf, T. 2000, "*GATA-1* interacts with the myeloid *PU.1* transcription factor and represses *PU.1*-dependent transcription", *Blood*, vol. 95, no. 8, pp. 2543-2551.

Nervi, C., Fazi, F., & Grignani, F. 2008, "Oncoproteins, heterochromatin silencing and microRNAs: a new link for leukemogenesis", *Epigenetics.*, vol. 3, no. 1, pp. 1-4.

Nilson, I., Lochner, K., Siegler, G., Greil, J., Beck, J. D., Fey, G. H., & Marschalek, R. 1996, "Exon/intron structure of the human ALL-1 (MLL) gene involved in translocations to chromosomal region 11q23 and acute leukaemias", *Br.J.Haematol.*, vol. 93, no. 4, pp. 966-972.

Nucifora, G. & Rowley, J. D. 1995, "*AML1* and the 8;21 and 3;21 translocations in acute and chronic myeloid leukemia", *Blood*, vol. 86, no. 1, pp. 1-14.

Odero, M. D., Carlson, K. M., Calasanz, M. J., & Rowley, J. D. 2001, "Further characterization of complex chromosomal rearrangements in myeloid malignancies: spectral karyotyping adds precision in defining abnormalities associated with poor prognosis", *Leukemia*, vol. 15, no. 7, pp. 1133-1136.

Oelschlagel, U., Nowak, R., Schaub, A., Koppel, C., Herbst, R., Mohr, B., Loffler, C., Range, U., Gunther, H., Assmann, M., Siegert, E., Wendt, E., Huhn, R., Brautigam, E., & Ehninger, G. 2000, "Shift of aberrant antigen expression at relapse or at treatment failure in acute leukemia", *Cytometry*, vol. 42, no. 4, pp. 247-253.

Ogawa, M. 1993, "Differentiation and proliferation of hematopoietic stem cells", *Blood*, vol. 81, no. 11, pp. 2844-2853.

Ogawa, M., Matsuzaki, Y., Nishikawa, S., Hayashi, S., Kunisada, T., Sudo, T., Kina, T., Nakauchi, H., & Nishikawa, S. 1991, "Expression and function of c-kit in hemopoietic progenitor cells", *J Exp.Med.*, vol. 174, no. 1, pp. 63-71.

Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V. M., Su, L., Xu, G., & Zhang, Y. 2005, "hDOT1L links histone methylation to leukemogenesis", *Cell*, vol. 121, no. 2, pp. 167-178.

Oken, M. M., Creech, R. H., Tormey, D. C., Horton, J., Davis, T. E., McFadden, E. T., & Carbone, P. P. 1982, "Toxicity and response criteria of the Eastern Cooperative Oncology Group", *Am.J.Clin.Oncol.*, vol. 5, no. 6, pp. 649-655.

Olesen, L. H., Nyvold, C. G., Aggerholm, A., Norgaard, J. M., Guldberg, P., & Hokland, P. 2005a, "Delineation and molecular characterization of acute myeloid leukemia patients with coduplication of *FLT3* and MLL", *Eur.J Haematol*, vol. 75, no. 3, pp. 185-192.

Olsson, I., Bergh, G., Ehinger, M., & Gullberg, U. 1996, "Cell differentiation in acute myeloid leukemia", *Eur.J.Haematol.*, vol. 57, no. 1, pp. 1-16.

Ommen, H. B., Nyvold, C. G., Braendstrup, K., Andersen, B. L., Ommen, I. B., Hasle, H., Hokland, P., & Ostergaard, M. 2008, "Relapse prediction in acute myeloid leukaemia patients in complete remission using *WT-1* as a molecular marker: development of a mathematical model to predict time from molecular to clinical relapse and define optimal sampling intervals", *Br.J.Haematol.*, vol. 141, no. 6, pp. 782-791.

Ono, R., Nakajima, H., Ozaki, K., Kumagai, H., Kawashima, T., Taki, T., Kitamura, T., Hayashi, Y., & Nosaka, T. 2005, "Dimerization of MLL fusion proteins and FLT3 activation synergize to induce multiple-lineage leukemogenesis", *J.Clin.Invest*, vol. 115, no. 4, pp. 919-929.

Orkin, S. H. & Zon, L. I. 2008, "Hematopoiesis: an evolving paradigm for stem cell biology", *Cell*, vol. 132, no. 4, pp. 631-644.

Ostergaard, M., Olesen, L. H., Hasle, H., Kjeldsen, E., & Hokland, P. 2004, "*WT-1* gene expression: an excellent tool for monitoring minimal residual disease in 70% of acute myeloid leukaemia patients - results from a single-centre study", *Br.J.Haematol.*, vol. 125, no. 5, pp. 590-600.

Pang, A., Au, W. Y., & Kwong, Y. L. 2004, "Caveolin-1 gene is coordinately regulated with the multidrug resistance 1 gene in normal and leukemic bone marrow", *Leuk.Res.*, vol. 28, no. 9, pp. 973-977.

Paschka, P., Marcucci, G., Ruppert, A. S., Mrozek, K., Chen, H., Kittles, R. A., Vukosavljevic, T., Perrotti, D., Vardiman, J. W., Carroll, A. J., Kolitz, J. E., Larson, R. A., & Bloomfield, C. D. 2006, "Adverse prognostic significance of KIT mutations in adult acute myeloid leukemia with inv(16) and t(8;21): a Cancer and Leukemia Group B Study", *J.Clin.Oncol.*, vol. 24, no. 24, pp. 3904-3911.

Paschka, P., Marcucci, G., Ruppert, A. S., Mrozek, K., Whitman, S. P., Maharry, K., Langer, C., Baldus, C. D., Powell, B. L., Baer, M. R., Carroll, A. J., Caligiuri, M. A., Kolitz, J. E., Larson, R. A., & Bloomfield, C. D. 2007, "Wilms Tumor 1 (*WT-1*) Gene Mutations Predict Poor Outcome in Adults with Cytogenetically Normal (CN) Acute Myeloid Leukemia (AML): A Cancer and Leukemia Group B (CALGB) Study", *ASH Annual Meeting Abstracts*, vol. 110, no. 11, p. 362.

Passegue, E., Jamieson, C. H., Ailles, L. E., & Weissman, I. L. 2003, "Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics?", *Proc.Natl.Acad.Sci.U.S.A*, vol. 100 Suppl 1, pp. 11842-11849.

Pedersen-Bjergaard, J., Andersen, M. K., Christiansen, D. H., & Nerlov, C. 2002, "Genetic pathways in therapy-related myelodysplasia and acute myeloid leukemia", *Blood*, vol. 99, no. 6, pp. 1909-1912.

Perea, G., Lasa, A., Aventin, A., Domingo, A., Villamor, N., Queipo de Llano, M. P., Llorente, A., Junca, J., Palacios, C., Fernandez, C., Gallart, M., Font, L., Tormo, M., Florensa,

L., Bargay, J., Marti, J. M., Vivancos, P., Torres, P., Berlanga, J. J., Badell, I., Brunet, S., Sierra, J., & Nomdedeu, J. F. 2006, "Prognostic value of minimal residual disease (MRD) in acute myeloid leukemia (AML) with favorable cytogenetics [t(8;21) and inv(16)]", *Leukemia*, vol. 20, no. 1, pp. 87-94.

Pileri, S. A., Ascani, S., Cox, M. C., Campidelli, C., Bacci, F., Piccioli, M., Piccaluga, P. P., Agostinelli, C., Asioli, S., Novero, D., Bisceglia, M., Ponzoni, M., Gentile, A., Rinaldi, P., Franco, V., Vincelli, D., Pileri, A., Jr., Gasbarra, R., Falini, B., Zinzani, P. L., & Baccarani, M. 2007, "Myeloid sarcoma: clinico-pathologic, phenotypic and cytogenetic analysis of 92 adult patients", *Leukemia*, vol. 21, no. 2, pp. 340-350.

Pineault, N., Helgason, C. D., Lawrence, H. J., & Humphries, R. K. 2002, "Differential expression of Hox, Meis1, and Pbx1 genes in primitive cells throughout murine hematopoietic ontogeny", *Exp.Hematol.*, vol. 30, no. 1, pp. 49-57.

Poppe, B., Vandesompele, J., Schoch, C., Lindvall, C., Mrozek, K., Bloomfield, C. D., Beverloo, H. B., Michaux, L., Dastugue, N., Herens, C., Yigit, N., De Paepe, A., Hagemeijer, A., & Speleman, F. 2004, "Expression analyses identify MLL as a prominent target of 11q23 amplification and support an etiologic role for MLL gain of function in myeloid malignancies", *Blood*, vol. 103, no. 1, pp. 229-235.

Porse, B. T., Pedersen, T. A., Xu, X., Lindberg, B., Wewer, U. M., Friis-Hansen, L., & Nerlov, C. 2001, "E2F repression by C/EBPalpha is required for adipogenesis and granulopoiesis in vivo", *Cell*, vol. 107, no. 2, pp. 247-258.

Prasad, R., Yano, T., Sorio, C., Nakamura, T., Rallapalli, R., Gu, Y., Leshkowitz, D., Croce, C. M., & Canaani, E. 1995, "Domains with transcriptional regulatory activity within the ALL1 and AF4 proteins involved in acute leukemia", *Proc.Natl.Acad.Sci.U.S.A*, vol. 92, no. 26, pp. 12160-12164.

Puccetti, E. & Ruthardt, M. 2004, "Acute promyelocytic leukemia: PML/RARalpha and the leukemic stem cell", *Leukemia*, vol. 18, no. 7, pp. 1169-1175.

Pui, C. H., Behm, F. G., Raimondi, S. C., Dodge, R. K., George, S. L., Rivera, G. K., Mirro, J., Jr., Kalwinsky, D. K., Dahl, G. V., & Murphy, S. B. 1989, "Secondary acute myeloid leukemia in children treated for acute lymphoid leukemia", *N.Engl.J.Med.*, vol. 321, no. 3, pp. 136-142.

Quesnel, B., Kantarjian, H., Bjergaard, J. P., Brault, P., Estey, E., Lai, J. L., Tilly, H., Stoppa, A. M., Archimbaud, E., Harousseau, J. L., & . 1993, "Therapy-related acute myeloid leukemia with t(8;21), inv(16), and t(8;16): a report on 25 cases and review of the literature", *J.Clin.Oncol.*, vol. 11, no. 12, pp. 2370-2379.

Rabbitts, T. H. 1991, "Translocations, master genes, and differences between the origins of acute and chronic leukemias", *Cell*, vol. 67, no. 4, pp. 641-644.

Radmacher, M. D., Marcucci, G., Ruppert, A. S., Mrozek, K., Whitman, S. P., Vardiman, J. W., Paschka, P., Vukosavljevic, T., Baldus, C. D., Kolitz, J. E., Caligiuri, M. A., Larson, R. A., & Bloomfield, C. D. 2006, "Independent confirmation of a prognostic gene-expression signature in adult acute myeloid leukemia with a normal karyotype: a Cancer and Leukemia Group B study", *Blood*, vol. 108, no. 5, pp. 1677-1683.

Radomska, H. S., Huettner, C. S., Zhang, P., Cheng, T., Scadden, D. T., & Tenen, D. G. 1998, "CCAAT/enhancer binding protein alpha is a regulatory switch sufficient for induction of granulocytic development from bipotential myeloid progenitors", *Mol.Cell Biol.*, vol. 18, no. 7, pp. 4301-4314.

Raponi, M., Lancet, J. E., Fan, H., Dossey, L., Lee, G., Gojo, I., Feldman, E. J., Gotlib, J., Morris, L. E., Greenberg, P. L., Wright, J. J., Harousseau, J. L., Lowenberg, B., Stone, R. M., De, P. P., Wang, Y., & Karp, J. E. 2008, "A 2-gene classifier for predicting response to the farnesyltransferase inhibitor tipifarnib in acute myeloid leukemia", *Blood*, vol. 111, no. 5, pp. 2589-2596.

Ravandi, F. & Estrov, Z. 2006, "Eradication of Leukemia Stem Cells as a New Goal of Therapy in Leukemia", *Clinical Cancer Research*, vol. 12, no. 2, pp. 340-344.

Reiter, A., Lengfelder, E., & Grimwade, D. 2004, "Pathogenesis, diagnosis and monitoring of residual disease in acute promyelocytic leukaemia", *Acta Haematol.*, vol. 112, no. 1-2, pp. 55-67.

Renneville, A., Roumier, C., Biggio, V., Nibourel, O., Boissel, N., Fenaux, P., & Preudhomme, C. 2008, "Cooperating gene mutations in acute myeloid leukemia: a review of the literature", *Leukemia*, vol. 22, no. 5, pp. 915-931.

Rhoades, K. L., Hetherington, C. J., Rowley, J. D., Hiebert, S. W., Nucifora, G., Tenen, D. G., & Zhang, D. E. 1996, "Synergistic up-regulation of the myeloid-specific promoter for the macrophage colony-stimulating factor receptor by *AML1* and the t(8;21) fusion protein may contribute to leukemogenesis", *Proc.Natl.Acad.Sci.U.S.A*, vol. 93, no. 21, pp. 11895-11900.

Rhoads, D. D., Dixit, A., & Roufa, D. J. 1986, "Primary structure of human ribosomal protein S14 and the gene that encodes it", *Mol Cell Biol.*, vol. 6, no. 8, pp. 2774-2783.

Rizo, A., Vellenga, E., de, H. G., & Schuringa, J. J. 2006, "Signaling pathways in self-renewing hematopoietic and leukemic stem cells: do all stem cells need a niche?", *Hum.Mol Genet.*, vol. 15 Spec No 2, p. R210-R219.

Robak, T. & Wierzbowska, A. 2009, "Current and emerging therapies for acute myeloid leukemia", *Clin Ther.*, vol. 31 Pt 2, pp. 2349-2370.

Rosen, J. M. & Jordan, C. T. 2009, "The increasing complexity of the cancer stem cell paradigm", *Science*, vol. 324, no. 5935, pp. 1670-1673.

Rosenbauer, F. & Tenen, D. G. 2007, "Transcription factors in myeloid development: balancing differentiation with transformation", *Nat.Rev.Immunol.*, vol. 7, no. 2, pp. 105-117.

Ross, M. E., Mahfouz, R., Onciu, M., Liu, H. C., Zhou, X., Song, G., Shurtleff, S. A., Pounds, S., Cheng, C., Ma, J., Ribeiro, R. C., Rubnitz, J. E., Girtman, K., Williams, W. K., Raimondi, S. C., Liang, D. C., Shih, L. Y., Pui, C. H., & Downing, J. R. 2004, "Gene expression profiling of pediatric acute myelogenous leukemia", *Blood*, vol. 104, no. 12, pp. 3679-3687.

Rossi, D. J., Bryder, D., Zahn, J. M., Ahlenius, H., Sonu, R., Wagers, A. J., & Weissman, I. L. 2005, "Cell intrinsic alterations underlie hematopoietic stem cell aging", *Proc.Natl.Acad.Sci.U.S.A*, vol. 102, no. 26, pp. 9194-9199.

Roussel, M. J. & Lanotte, M. 2001, "Maturation sensitive and resistant t(15;17) NB4 cell lines as tools for APL physiopathology: nomenclature of cells and repertory of their known genetic alterations and phenotypes", *Oncogene*, vol. 20, no. 49, pp. 7287-7291.

Rowe, J. M. 2008, "Consolidation therapy: what should be the standard of care?", *Best.Pract.Res.Clin.Haematol.*, vol. 21, no. 1, pp. 53-60.

Rowe, J. M., Neuberg, D., Friedenberg, W., Bennett, J. M., Paietta, E., Makary, A. Z., Liesveld, J. L., Abboud, C. N., Dewald, G., Hayes, F. A., Tallman, M. S., & Wiernik, P. H. 2004, "A phase 3 study of three induction regimens and of priming with GM-CSF in older adults with acute myeloid leukemia: a trial by the Eastern Cooperative Oncology Group", *Blood*, vol. 103, no. 2, pp. 479-485.

Rowley, J. D. & Olney, H. J. 2002, "International workshop on the relationship of prior therapy to balanced chromosome aberrations in therapy-related myelodysplastic syndromes and acute leukemia: overview report", *Genes Chromosomes.Cancer*, vol. 33, no. 4, pp. 331-345.

Rozen, S. & Skaletsky, H. 2000, "Primer3 on the WWW for general users and for biologist programmers," in *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, Krawetz S & Misener S, eds., Humana Press, Totowa, NJ, pp. 365-386.

Rozovskaia, T., Ravid-Amir, O., Tillib, S., Getz, G., Feinstein, E., Agrawal, H., Nagler, A., Rappaport, E. F., Issaeva, I., Matsuo, Y., Kees, U. R., Lapidot, T., Lo, C. F., Foa, R., Mazo, A., Nakamura, T., Croce, C. M., Cimino, G., Domany, E., & Canaani, E. 2003, "Expression profiles of acute lymphoblastic and myeloblastic leukemias with ALL-1 rearrangements", *Proc.Natl.Acad.Sci.U.S.A*, vol. 100, no. 13, pp. 7853-7858.

Rubnitz, J. E., Behm, F. G., & Downing, J. R. 1996, "11q23 rearrangements in acute leukemia", *Leukemia*, vol. 10, no. 1, pp. 74-82.

Rund, D., Krichevsky, S., Bar-Cohen, S., Goldschmidt, N., Kedmi, M., Malik, E., Gural, A., Shafran-Tikva, S., Ben-Neriah, S., & Ben-Yehuda, D. 2005, "Therapy-related leukemia: clinical characteristics and analysis of new molecular risk factors in 96 adult patients", *Leukemia*, vol. 19, no. 11, pp. 1919-1928.

Rusten, L. S., Lyman, S. D., Veiby, O. P., & Jacobsen, S. E. 1996, "The *FLT3* ligand is a direct and potent stimulator of the growth of primitive and committed human CD34+ bone marrow progenitor cells in vitro", *Blood*, vol. 87, no. 4, pp. 1317-1325.

Sachs, L. 1986, "Cell differentiation and malignancy", *Cell Biophys.*, vol. 9, no. 1-2, pp. 225-242.

Sambrook, J. & Russel, D. 2001, "Techniques for Quantitating RNA," in *Molecular Cloning: A laboratory manual*, Third edn, p. 7.65-7.66.

San Miguel, J. F., Vidriales, M. B., Lopez-Berges, C., az-Mediavilla, J., Gutierrez, N., Canizo, C., Ramos, F., Calmuntia, M. J., Perez, J. J., Gonzalez, M., & Orfao, A. 2001, "Early immunophenotypical evaluation of minimal residual disease in acute myeloid leukemia identifies different patient risk groups and may contribute to postinduction treatment stratification", *Blood*, vol. 98, no. 6, pp. 1746-1751.

Sanda, T., Li, X., Gutierrez, A., Ahn, Y., Neuberg, D. S., O'Neil, J., Strack, P. R., Winter, C. G., Winter, S. S., Larson, R. S., von, B. H., & Look, A. T. 2010, "Interconnecting molecular pathways in the pathogenesis and drug sensitivity of T-cell acute lymphoblastic leukemia", *Blood*, vol. 115, no. 9, pp. 1735-1745.

Sandler, D. P. & Collman, G. W. 1987, "Cytogenetic and environmental factors in the etiology of the acute leukemias in adults", *Am.J.Epidemiol.*, vol. 126, no. 6, pp. 1017-1032.

Santamaria, C. M., Chillon, M. C., Garcia-Sanz, R., Perez, C., Caballero, M. D., Ramos, F., de Coca, A. G., Alonso, J. M., Giraldo, P., Bernal, T., Queizan, J. A., Rodriguez, J. N., Fernandez-Abellan, P., Barez, A., Penarrubia, M. J., Balanzategui, A., Vidriales, M. B., Sarasquete, M. E., Alcoceba, M., az-Mediavilla, J., San Miguel, J. F., & Gonzalez, M. 2009, "Molecular stratification model for prognosis in cytogenetically normal acute myeloid leukemia", *Blood*, vol. 114, no. 1, pp. 148-152.

Sauvageau, G., Lansdorp, P. M., Eaves, C. J., Hogge, D. E., Dragowska, W. H., Reid, D. S., Largman, C., Lawrence, H. J., & Humphries, R. K. 1994, "Differential expression of homeobox genes in functionally distinct CD34+ subpopulations of human bone marrow cells", *Proc.Natl.Acad.Sci.U.S.A*, vol. 91, no. 25, pp. 12223-12227.

Sawyers, C. L., Denny, C. T., & Witte, O. N. 1991, "Leukemia and the disruption of normal hematopoiesis", *Cell*, vol. 64, no. 2, pp. 337-350.

Scherer, P. E., Okamoto, T., Chun, M., Nishimoto, I., Lodish, H. F., & Lisanti, M. P. 1996, "Identification, sequence, and expression of caveolin-2 defines a caveolin gene family", *Proc.Natl.Acad.Sci.U.S.A*, vol. 93, no. 1, pp. 131-135.

Schichman, S. A., Caligiuri, M. A., Gu, Y., Strout, M. P., Canaani, E., Bloomfield, C. D., & Croce, C. M. 1994a, "ALL-1 partial duplication in acute leukemia", *Proc.Natl.Acad.Sci.U.S.A*, vol. 91, no. 13, pp. 6236-6239.

Schichman, S. A., Caligiuri, M. A., Strout, M. P., Carter, S. L., Gu, Y., Canaani, E., Bloomfield, C. D., & Croce, C. M. 1994b, "ALL-1 tandem duplication in acute myeloid leukemia with a normal karyotype involves homologous recombination between Alu elements", *Cancer Res.*, vol. 54, no. 16, pp. 4277-4280.

Schlenk, R. F., Dohner, K., Krauter, J., Frohling, S., Corbacioglu, A., Bullinger, L., Habdank, M., Spath, D., Morgan, M., Benner, A., Schlegelberger, B., Heil, G., Ganser, A., & Dohner, H. 2008, "Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia", *N.Engl.J.Med.*, vol. 358, no. 18, pp. 1909-1918.

Schnittger, S., Kinkelin, U., Schoch, C., Heinecke, A., Haase, D., Haferlach, T., Buchner, T., Wormann, B., Hiddemann, W., & Griesinger, F. 2000, "Screening for MLL tandem duplication in 387 unselected patients with AML identify a prognostically unfavorable subset of AML", *Leukemia*, vol. 14, no. 5, pp. 796-804.

Schnittger, S., Kohl, T. M., Haferlach, T., Kern, W., Hiddemann, W., Spiekermann, K., & Schoch, C. 2006, "KIT-D816 mutations in *AML1-ETO*-positive AML are associated with impaired event-free and overall survival", *Blood*, vol. 107, no. 5, pp. 1791-1799.

Schnittger, S., Schoch, C., Dugas, M., Kern, W., Staib, P., Wuchter, C., Loffler, H., Sauerland, C. M., Serve, H., Buchner, T., Haferlach, T., & Hiddemann, W. 2002, "Analysis of *FLT3* length mutations in 1003 patients with acute myeloid leukemia: correlation to cytogenetics, FAB subtype, and prognosis in the AMLCG study and usefulness as a marker for the detection of minimal residual disease", *Blood*, vol. 100, no. 1, pp. 59-66.

Schnittger, S., Schoch, C., Kern, W., Hiddemann, W., & Haferlach, T. 2004, "*FLT3* length mutations as marker for follow-up studies in acute myeloid leukaemia", *Acta Haematol.*, vol. 112, no. 1-2, pp. 68-78.

Schnittger, S., Schoch, C., Kern, W., Mecucci, C., Tschulik, C., Martelli, M. F., Haferlach, T., Hiddemann, W., & Falini, B. 2005a, "Nucleophosmin gene mutations are predictors of favorable prognosis in acute myelogenous leukemia with a normal karyotype", *Blood*, vol. 106, no. 12, pp. 3733-3739.

Schnittger, S., Weisser, M., Schoch, C., Hiddemann, W., Haferlach, T., & Kern, W. 2003, "New score predicting for prognosis in *PML-RAR+*, *AML1-ETO+*, or CBFBMYH11+ acute myeloid leukemia based on quantification of fusion transcripts", *Blood*, vol. 102, no. 8, pp. 2746-2755.

Schnittger, S., Wormann, B., Hiddemann, W., & Griesinger, F. 1998, "Partial tandem duplications of the MLL gene are detectable in peripheral blood and bone marrow of nearly all healthy donors", *Blood*, vol. 92, no. 5, pp. 1728-1734.

Schofield, R. 1978, "The relationship between the spleen colony-forming cell and the haemopoietic stem cell", *Blood Cells*, vol. 4, no. 1-2, pp. 7-25.

Scholl, C., Schlenk, R. F., Eiwen, K., Dohner, H., Frohling, S., & Dohner, K. 2005, "The prognostic value of MLL-AF9 detection in patients with t(9;11)(p22;q23)-positive acute myeloid leukemia", *Haematologica*, vol. 90, no. 12, pp. 1626-1634.

Scott, E. W., Simon, M. C., Anastasi, J., & Singh, H. 1994, "Requirement of transcription factor *PU.1* in the development of multiple hematopoietic lineages", *Science*, vol. 265, no. 5178, pp. 1573-1577.

Secker-Walker, L. M. & Craig, J. M. 1993, "Prognostic implications of breakpoint and lineage heterogeneity in Philadelphia-positive acute lymphoblastic leukemia: a review", *Leukemia*, vol. 7, no. 2, pp. 147-151.

Sheikhha, M. H., Awan, A., Tobal, K., & Liu Yin, J. A. 2003, "Prognostic significance of *FLT3* ITD and D835 mutations in AML patients", *Hematol.J.*, vol. 4, no. 1, pp. 41-46.

Shiah, H. S., Kuo, Y. Y., Tang, J. L., Huang, S. Y., Yao, M., Tsay, W., Chen, Y. C., Wang, C. H., Shen, M. C., Lin, D. T., Lin, K. H., & Tien, H. F. 2002, "Clinical and biological implications of partial tandem duplication of the MLL gene in acute myeloid leukemia without chromosomal abnormalities at 11q23", *Leukemia*, vol. 16, no. 2, pp. 196-202.

Shih, L. Y., Liang, D. C., Fu, J. F., Wu, J. H., Wang, P. N., Lin, T. L., Dunn, P., Kuo, M. C., Tang, T. C., Lin, T. H., & Lai, C. L. 2006a, "Characterization of fusion partner genes in 114 patients with de novo acute myeloid leukemia and MLL rearrangement", *Leukemia*, vol. 20, no. 2, pp. 218-223.

Shih, L. Y., Liang, D. C., Huang, C. F., Wu, J. H., Lin, T. L., Wang, P. N., Dunn, P., Kuo, M. C., & Tang, T. C. 2006b, "AML patients with CEBPalpha mutations mostly retain identical mutant patterns but frequently change in allelic distribution at relapse: a comparative analysis on paired diagnosis and relapse samples", *Leukemia*, vol. 20, no. 4, pp. 604-609.

Shipley, J. L. & Butera, J. N. 2009, "Acute myelogenous leukemia", *Exp.Hematol*, vol. 37, no. 6, pp. 649-658.

Slany, R. K., Lavau, C., & Cleary, M. L. 1998, "The oncogenic capacity of HRX-*ENL* requires the transcriptional transactivation activity of *ENL* and the DNA binding motifs of HRX", *Mol.Cell Biol.*, vol. 18, no. 1, pp. 122-129.

Slovak, M. L., Kopecky, K. J., Cassileth, P. A., Harrington, D. H., Theil, K. S., Mohamed, A., Paietta, E., Willman, C. L., Head, D. R., Rowe, J. M., Forman, S. J., & Appelbaum, F. R. 2000, "Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study", *Blood*, vol. 96, no. 13, pp. 4075-4083.

Small, D. 2006, "*FLT3* Mutations: Biology and Treatment", *Hematology Am Soc Hematol Educ Program* pp. 178-184.

Smith, S. M., Le Beau, M. M., Huo, D., Karrison, T., Sobecks, R. M., Anastasi, J., Vardiman, J. W., Rowley, J. D., & Larson, R. A. 2003, "Clinical-cytogenetic associations in 306 patients with therapy-related myelodysplasia and myeloid leukemia: the University of Chicago series", *Blood*, vol. 102, no. 1, pp. 43-52.

So, C. W. & Cleary, M. L. 2004, "Dimerization: a versatile switch for oncogenesis", *Blood*, vol. 104, no. 4, pp. 919-922.

So, C. W., Karsunky, H., Wong, P., Weissman, I. L., & Cleary, M. L. 2004, "Leukemic transformation of hematopoietic progenitors by MLL-*GAS7* in the absence of Hoxa7 or Hoxa9", *Blood*, vol. 103, no. 8, pp. 3192-3199.

So, C. W., Lin, M., Ayton, P. M., Chen, E. H., & Cleary, M. L. 2003, "Dimerization contributes to oncogenic activation of MLL chimeras in acute leukemias", *Cancer Cell*, vol. 4, no. 2, pp. 99-110.

Somervaille, T. C., Matheny, C. J., Spencer, G. J., Iwasaki, M., Rinn, J. L., Witten, D. M., Chang, H. Y., Shurtleff, S. A., Downing, J. R., & Cleary, M. L. 2009, "Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells2", *Cell Stem Cell*, vol. 4, no. 2, pp. 129-140.

Somervaille, T. C. P. & Cleary, M. L. 2006, "Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia", *Cancer Cell*, vol. 10, no. 4, pp. 257-268.

Stentoft, J., Hokland, P., Ostergaard, M., Hasle, H., & Nyvold, C. G. 2006, "Minimal residual core binding factor AMLs by real time quantitative PCR--initial response to chemotherapy predicts event free survival and close monitoring of peripheral blood unravels the kinetics of relapse", *Leuk.Res.*, vol. 30, no. 4, pp. 389-395.

Steudel, C., Wermke, M., Schaich, M., Schakel, U., Illmer, T., Ehninger, G., & Thiede, C. 2003, "Comparative analysis of MLL partial tandem duplication and *FLT3* internal tandem

duplication mutations in 956 adult patients with acute myeloid leukemia", *Genes Chromosomes.Cancer*, vol. 37, no. 3, pp. 237-251.

Stirewalt, D. L. & Radich, J. P. 2003, "The role of *FLT3* in haematopoietic malignancies", *Nat.Rev.Cancer*, vol. 3, no. 9, pp. 650-665.

Stone, R. M., O'Donnell, M. R., & Sekeres, M. A. 2004, "Acute myeloid leukemia", *Hematology.Am Soc.Hematol Educ.Program.* pp. 98-117.

Strout, M. P., Marcucci, G., Bloomfield, C. D., & Caligiuri, M. A. 1998, "The partial tandem duplication of ALL1 (MLL) is consistently generated by Alu-mediated homologous recombination in acute myeloid leukemia", *Proc.Natl.Acad.Sci.U.S.A*, vol. 95, no. 5, pp. 2390-2395.

Suciu, S., Mandelli, F., de, W. T., Zittoun, R., Gallo, E., Labar, B., De, R. G., Belhabri, A., Giustolisi, R., Delarue, R., Liso, V., Mirto, S., Leone, G., Bourhis, J. H., Fioritoni, G., Jehn, U., Amadori, S., Fazi, P., Hagemeijer, A., & Willemze, R. 2003, "Allogeneic compared with autologous stem cell transplantation in the treatment of patients younger than 46 years with acute myeloid leukemia (AML) in first complete remission (CR1): an intention-to-treat analysis of the EORTC/GIMEMAAML-10 trial", *Blood*, vol. 102, no. 4, pp. 1232-1240.

Suela, J., Alvarez, S., Cifuentes, F., Largo, C., Ferreira, B. I., Blesa, D., Ardanaz, M., Garcia, R., Marquez, J. A., Odero, M. D., Calasanz, M. J., & Cigudosa, J. C. 2007, "DNA profiling analysis of 100 consecutive de novo acute myeloid leukemia cases reveals patterns of genomic instability that affect all cytogenetic risk groups", *Leukemia*, vol. 21, no. 6, pp. 1224-1231.

Sugiyama, H. 1998, "Wilms tumor gene (*WT-1*) as a new marker for the detection of minimal residual disease in leukemia", *Leuk.Lymphoma*, vol. 30, no. 1-2, pp. 55-61.

Suzuki, T., Kiyoi, H., Ozeki, K., Tomita, A., Yamaji, S., Suzuki, R., Kodera, Y., Miyawaki, S., Asou, N., Kuriyama, K., Yagasaki, F., Shimazaki, C., Akiyama, H., Nishimura, M., Motoji, T., Shinagawa, K., Takeshita, A., Ueda, R., Kinoshita, T., Emi, N., & Naoe, T. 2005a, "Clinical characteristics and prognostic implications of *NPM1* mutations in acute myeloid leukemia", *Blood*, vol. 106, no. 8, pp. 2854-2861.

Suzuki, T., Kiyoi, H., Ozeki, K., Tomita, A., Yamaji, S., Suzuki, R., Kodera, Y., Miyawaki, S., Asou, N., Kuriyama, K., Yagasaki, F., Shimazaki, C., Akiyama, H., Nishimura, M., Motoji, T., Shinagawa, K., Takeshita, A., Ueda, R., Kinoshita, T., Emi, N., & Naoe, T. 2005b, "Clinical characteristics and prognostic implications of *NPM1* mutations in acute myeloid leukemia", *Blood*, vol. 106, no. 8, pp. 2854-2861.

Swerdlow, S., Campo, E., Harris, N. L., Jaffe, E. S., Pileri, S. A., Stein, H., Thiele, J., & Vardiman, J. W. E. 2008, *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*, Fourth Edition edn, IARC:Lyon 2008.

Takahashi, A., Satake, M., Yamaguchi-Iwai, Y., Bae, S. C., Lu, J., Maruyama, M., Zhang, Y. W., Oka, H., Arai, N., Arai, K., & . 1995, "Positive and negative regulation of granulocyte-macrophage colony-stimulating factor promoter activity by *AML1*-related transcription factor, PEBP2", *Blood*, vol. 86, no. 2, pp. 607-616.

Takenokuchi, M., Yasuda, C., Takeuchi, K., Nakamachi, Y., Mukai, M., Kondo, S., Kumagai, S., Saigo, K., Murayama, T., Koizumi, T., & Tatsumi, E. 2004, "Quantitative nested reverse

transcriptase PCR vs. real-time PCR for measuring *AML1/ETO* (MTG8) transcripts", *Clin Lab Haematol*, vol. 26, no. 2, pp. 107-114.

Tamaki, H., Ogawa, H., Inoue, K., Soma, T., Yamagami, T., Miyake, S., Oka, Y., Oji, Y., Tatekawa, T., Tsuboi, A., Tagawa, S., Kitani, T., Aozasa, K., Kishimoto, T., Sugiyama, H., Miwa, H., & Kita, K. 1996, "Increased expression of the Wilms tumor gene (*WT-1*) at relapse in acute leukemia", *Blood*, vol. 88, no. 11, pp. 4396-4398.

Taussig, D. C., Miraki-Moud, F., njos-Afonso, F., Pearce, D. J., Allen, K., Ridler, C., Lillington, D., Oakervee, H., Cavenagh, J., Agrawal, S. G., Lister, T. A., Gribben, J. G., & Bonnet, D. 2008, "Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells", *Blood*, vol. 112, no. 3, pp. 568-575.

Testa, J. R., Mintz, U., Rowley, J. D., Vardiman, J. W., & Golomb, H. M. 1979, "Evolution of karyotypes in acute nonlymphocytic leukemia", *Cancer Res.*, vol. 39, no. 9, pp. 3619-3627.

Thiede, C., Steudel, C., Mohr, B., Schaich, M., Schakel, U., Platzbecker, U., Wermke, M., Bornhauser, M., Ritter, M., Neubauer, A., Ehninger, G., & Illmer, T. 2002, "Analysis of *FLT3*-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis", *Blood*, vol. 99, no. 12, pp. 4326-4335.

Thiede, C., Koch, S., Creutzig, E., Steudel, C., Illmer, T., Schaich, M., & Ehninger, G. 2006, "Prevalence and prognostic impact of *NPM1* mutations in 1485 adult patients with acute myeloid leukemia (AML)", *Blood* pp. 2005-2008.

Thorsteinsdottir, U., Mamo, A., Kroon, E., Jerome, L., Bijl, J., Lawrence, H. J., Humphries, K., & Sauvageau, G. 2002, "Overexpression of the myeloid leukemia-associated Hoxa9 gene in bone marrow cells induces stem cell expansion", *Blood*, vol. 99, no. 1, pp. 121-129.

Till, J. E. & McCulloch, E. A. 1961, "A direct measurement of the radiation sensitivity of normal mouse bone marrow cells", *Radiat.Res*, vol. 14, pp. 213-222.

Tkachuk, D. C., Kohler, S., & Cleary, M. L. 1992, "Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias", *Cell*, vol. 71, no. 4, pp. 691-700.

Tobal, K. & Yin, J. A. 1996a, "Monitoring of minimal residual disease by quantitative reverse transcriptase-polymerase chain reaction for *AML1*-MTG8 transcripts in AML-M2 with t(8; 21)", *Blood*, vol. 88, no. 10, pp. 3704-3709.

Ummanni, R., Teller, S., Junker, H., Zimmermann, U., Venz, S., Scharf, C., Giebel, J., & Walther, R. 2008, "Altered expression of tumor protein D52 regulates apoptosis and migration of prostate cancer cells", *FEBS J.*, vol. 275, no. 22, pp. 5703-5713.

Valk, P. J., Verhaak, R. G., Beijen, M. A., Erpelinck, C. A., Barjesteh van Waalwijk van Doorn-Khosrovani, Boer, J. M., Beverloo, H. B., Moorhouse, M. J., van der Spek, P. J., Lowenberg, B., & Delwel, R. 2004, "Prognostically useful gene-expression profiles in acute myeloid leukemia", *N.Engl.J.Med.*, vol. 350, no. 16, pp. 1617-1628.

van Dongen, J. J., Macintyre, E. A., Gabert, J. A., Delabesse, E., Rossi, V., Saglio, G., Gottardi, E., Rambaldi, A., Dotti, G., Griesinger, F., Parreira, A., Gameiro, P., Diaz, M. G., Malec, M., Langerak, A. W., San Miguel, J. F., & Biondi, A. 1999, "Standardized RT-PCR

analysis of fusion gene transcripts from chromosome aberrations in acute leukemia for detection of minimal residual disease. Report of the BIOMED-1 Concerted Action: investigation of minimal residual disease in acute leukemia", *Leukemia*, vol. 13, no. 12, pp. 1901-1928.

Van, D. V., V, Hochhaus, A., Cazzaniga, G., Szczepanski, T., Gabert, J., & van Dongen, J. J. 2003a, "Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects", *Leukemia*, vol. 17, no. 6, pp. 1013-1034.

Van, D. V., V, Hochhaus, A., Cazzaniga, G., Szczepanski, T., Gabert, J., & van Dongen, J. J. 2003b, "Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects", *Leukemia*, vol. 17, no. 6, pp. 1013-1034.

Vardiman, J. W., Harris, N. L., & Brunning, R. D. 2002, "The World Health Organization (WHO) classification of the myeloid neoplasms", *Blood*, vol. 100, no. 7, pp. 2292-2302.

Vardiman, J. W., Thiele, J., Arber, D. A., Brunning, R. D., Borowitz, M. J., Porwit, A., Harris, N. L., Le Beau, M. M., Hellstrom-Lindberg, E., Tefferi, A., & Bloomfield, C. D. 2009, "The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes", *Blood*, vol. 114, no. 5, pp. 937-951.

Verhaak, R. G., Goudswaard, C. S., van, P. W., Bijl, M. A., Sanders, M. A., Hugens, W., Uitterlinden, A. G., Erpelinck, C. A., Delwel, R., Lowenberg, B., & Valk, P. J. 2005a, "Mutations in nucleophosmin (*NPM1*) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance", *Blood*, vol. 106, no. 12, pp. 3747-3754.

Verhaak, R. G., Goudswaard, C. S., van, P. W., Bijl, M. A., Sanders, M. A., Hugens, W., Uitterlinden, A. G., Erpelinck, C. A., Delwel, R., Lowenberg, B., & Valk, P. J. 2005, "Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance", *Blood*, vol. 106, no. 12, pp. 3747-3754.

Verhaak, R. G. W., Wouters, B. J., Erpelinck, C. A. J., Abbas, S., Beverloo, H. B., Lugthart, S., Lowenberg, B., Delwel, R., & Valk, P. J. M. 2009, "Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling", *Haematologica*, vol. 94, no. 1, pp. 131-134.

Vidot, S., Witham, J., Agarwal, R., Greenhough, S., Bamrah, H. S., Tigyi, G. J., Kaye, S. B., & Richardson, A. 2010, "Autotaxin delays apoptosis induced by carboplatin in ovarian cancer cells", *Cell Signal.*, vol. 22, no. 6, pp. 926-935.

Virtaneva, K., Wright, F. A., Tanner, S. M., Yuan, B., Lemon, W. J., Caligiuri, M. A., Bloomfield, C. D., de la, C. A., & Krahe, R. 2001, "Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics", *Proc.Natl.Acad.Sci.U.S.A*, vol. 98, no. 3, pp. 1124-1129.

Vitoux, D., Nasr, R., & de, T. H. 2007, "Acute promyelocytic leukemia: new issues on pathogenesis and treatment response", *Int.J.Biochem.Cell Biol.*, vol. 39, no. 6, pp. 1063-1070.

Voskova, D., Schoch, C., Schnittger, S., Hiddemann, W., Haferlach, T., & Kern, W. 2004, "Stability of leukemia-associated aberrant immunophenotypes in patients with acute myeloid leukemia between diagnosis and relapse: comparison with cytomorphologic, cytogenetic, and molecular genetic findings", *Cytometry B Clin Cytom.*, vol. 62, no. 1, pp. 25-38.

Walsby, E., Walsh, V., Pepper, C., Burnett, A., & Mills, K. 2008, "Effects of the aurora kinase inhibitors AZD1152-HQPA and ZM447439 on growth arrest and polyploidy in acute myeloid leukemia cell lines and primary blasts", *Haematologica*, vol. 93, no. 5, pp. 662-669.

Wang, S. A., Jabbar, K., Lu, G., Chen, S. S., Galili, N., Vega, F., Jones, D., Raza, A., Kantarjian, H., Garcia-Manero, G., McDonnell, T. J., & Medeiros, L. J. 2010, "Trisomy 11 in myelodysplastic syndromes defines a unique group of disease with aggressive clinicopathologic features", *Leukemia*, vol. 24, no. 4, pp. 740-747.

Watson, M. A., Ylagan, L. R., Trinkaus, K. M., Gillanders, W. E., Naughton, M. J., Weilbaecher, K. N., Fleming, T. P., & Aft, R. L. 2007, "Isolation and molecular profiling of bone marrow micrometastases identifies *TWIST1* as a marker of early tumor relapse in breast cancer patients", *Clin.Cancer Res.*, vol. 13, no. 17, pp. 5001-5009.

Weisser, M., Haferlach, C., Hiddemann, W., & Schnittger, S. 2007, "The quality of molecular response to chemotherapy is predictive for the outcome of *AML1-ETO*-positive AML and is independent of pretreatment risk factors", *Leukemia*, vol. 21, no. 6, pp. 1177-1182.

Weisser, M., Kern, W., Rauhut, S., Schoch, C., Hiddemann, W., Haferlach, T., & Schnittger, S. 2005a, "Prognostic impact of RT-PCR-based quantification of *WT-1* gene expression during MRD monitoring of acute myeloid leukemia", *Leukemia*, vol. 19, no. 8, pp. 1416-1423.

Weisser, M., Kern, W., Schoch, C., Hiddemann, W., Haferlach, T., & Schnittger, S. 2005b, "Risk assessment by monitoring expression levels of partial tandem duplications in the MLL gene in acute myeloid leukemia during therapy", *Haematologica*, vol. 90, no. 7, pp. 881-889.

Wheatley, K., Brookes, C. L., Howman, A. J., Goldstone, A. H., Milligan, D. W., Prentice, A. G., Moorman, A. V., & Burnett, A. K. 2009, "Prognostic factor analysis of the survival of elderly patients with AML in the MRC AML11 and LRF AML14 trials", *Br.J.Haematol.*, vol. 145, no. 5, pp. 598-605.

Whitman, S. P., Archer, K. J., Feng, L., Baldus, C., Becknell, B., Carlson, B. D., Carroll, A. J., Mrozek, K., Vardiman, J. W., George, S. L., Kolitz, J. E., Larson, R. A., Bloomfield, C. D., & Caligiuri, M. A. 2001, "Absence of the wild-type allele predicts poor prognosis in adult de novo acute myeloid leukemia with normal cytogenetics and the internal tandem duplication of *FLT3*: a cancer and leukemia group B study", *Cancer Res.*, vol. 61, no. 19, pp. 7233-7239.

Whitman, S. P., Liu, S., Vukosavljevic, T., Rush, L. J., Yu, L., Liu, C., Klisovic, M. I., Maharry, K., Guimond, M., Strout, M. P., Becknell, B., Dorrance, A., Klisovic, R. B., Plass, C., Bloomfield, C. D., Marcucci, G., & Caligiuri, M. A. 2005, "The MLL partial tandem duplication: evidence for recessive gain-of-function in acute myeloid leukemia identifies a novel patient subgroup for molecular-targeted therapy", *Blood*, vol. 106, no. 1, pp. 345-352.

Whitman, S. P., Hackanson, B., Liyanarachchi, S., Liu, S., Rush, L. J., Maharry, K., Margeson, D., Davuluri, R., Wen, J., Witte, T., Yu, L., Liu, C., Bloomfield, C. D., Marcucci, G., Plass, C., & Caligiuri, M. A. 2008, "DNA hypermethylation and epigenetic silencing of

the tumor suppressor gene, SLC5A8, in acute myeloid leukemia with the MLL partial tandem duplication", *Blood*, vol. 112, no. 5, pp. 2013-2016.

Whitman, S. P., Ruppert, A. S., Marcucci, G., Mrozek, K., Paschka, P., Langer, C., Baldus, C. D., Wen, J., Vukosavljevic, T., Powell, B. L., Carroll, A. J., Kolitz, J. E., Larson, R. A., Caligiuri, M. A., & Bloomfield, C. D. 2007, "Long-term disease-free survivors with cytogenetically normal acute myeloid leukemia and MLL partial tandem duplication: a Cancer and Leukemia Group B study", *Blood*, vol. 109, no. 12, pp. 5164-5167.

Whitman, S. P., Ruppert, A. S., Radmacher, M. D., Mrozek, K., Paschka, P., Langer, C., Baldus, C. D., Wen, J., Racke, F., Powell, B. L., Kolitz, J. E., Larson, R. A., Caligiuri, M. A., Marcucci, G., & Bloomfield, C. D. 2008b, "*FLT3* D835/I836 mutations are associated with poor disease-free survival and a distinct gene-expression signature among younger adults with de novo cytogenetically normal acute myeloid leukemia lacking *FLT3* internal tandem duplications", *Blood*, vol. 111, no. 3, pp. 1552-1559.

Wouters, B. J., Jorda, M. A., Keeshan, K., Louwers, I., Erpelinck-Verschueren, C. A., Tielemans, D., Langerak, A. W., He, Y., Yashiro-Ohtani, Y., Zhang, P., Hetherington, C. J., Verhaak, R. G., Valk, P. J., Lowenberg, B., Tenen, D. G., Pear, W. S., & Delwel, R. 2007, "Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced *CEBPA* and mutations in NOTCH1", *Blood*, vol. 110, no. 10, pp. 3706-3714.

Wouters, B. J., Lowenberg, B., & Delwel, R. 2009a, "A decade of genome-wide gene expression profiling in acute myeloid leukemia: flashback and prospects", *Blood*, vol. 113, no. 2, pp. 291-298.

Wouters, B. J., Lowenberg, B., & Delwel, R. 2009b, "A decade of genome-wide gene expression profiling in acute myeloid leukemia: flashback and prospects", *Blood*, vol. 113, no. 2, pp. 291-298.

Xia, Z. B., Anderson, M., Diaz, M. O., & Zeleznik, L. 2003, "MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein", *Proc.Natl.Acad.Sci.U.S.A*, vol. 100, no. 14, pp. 8342-8347.

Yamada, T., Abe, M., Higashi, T., Yamamoto, H., Kihara-Negishi, F., Sakurai, T., Shirai, T., & Oikawa, T. 2001, "Lineage switch induced by overexpression of Ets family transcription factor *PU.1* in murine erythroleukemia cells", *Blood*, vol. 97, no. 8, pp. 2300-2307.

Yanada, M., Matsuo, K., Suzuki, T., Kiyoi, H., & Naoe, T. 2005, "Prognostic significance of *FLT3* internal tandem duplication and tyrosine kinase domain mutations for acute myeloid leukemia: a meta-analysis", *Leukemia*, vol. 19, no. 8, pp. 1345-1349.

Yao, Q., Nishiuchi, R., Li, Q., Kumar, A. R., Hudson, W. A., & Kersey, J. H. 2003, "*FLT3* expressing leukemias are selectively sensitive to inhibitors of the molecular chaperone heat shock protein 90 through destabilization of signal transduction-associated kinases", *Clin.Cancer Res.*, vol. 9, no. 12, pp. 4483-4493.

Yokoyama, A., Kitabayashi, I., Ayton, P. M., Cleary, M. L., & Ohki, M. 2002, "Leukemia proto-oncoprotein MLL is proteolytically processed into 2 fragments with opposite transcriptional properties", *Blood*, vol. 100, no. 10, pp. 3710-3718.

Yokoyama, A., Somervaille, T. C., Smith, K. S., Rozenblatt-Rosen, O., Meyerson, M., & Cleary, M. L. 2005, "The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis", *Cell*, vol. 123, no. 2, pp. 207-218.

Yu, M., Honoki, K., Andersen, J., Paietta, E., Nam, D. K., & Yunis, J. J. 1996, "MLL tandem duplication and multiple splicing in adult acute myeloid leukemia with normal karyotype", *Leukemia*, vol. 10, no. 5, pp. 774-780.

Zangrando, A., Dell'orto, M. C., Te, K. G., & Basso, G. 2009, "MLL rearrangements in pediatric acute lymphoblastic and myeloblastic leukemias: MLL specific and lineage specific signatures", *BMC.Med.Genomics*, vol. 2, p. 36.

Zhang, D. E., Zhang, P., Wang, N. D., Hetherington, C. J., Darlington, G. J., & Tenen, D. G. 1997, "Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice", *Proc.Natl.Acad.Sci.U.S.A*, vol. 94, no. 2, pp. 569-574.

Zhang, J., Niu, C., Ye, L., Huang, H., He, X., Tong, W. G., Ross, J., Haug, J., Johnson, T., Feng, J. Q., Harris, S., Wiedemann, L. M., Mishina, Y., & Li, L. 2003, "Identification of the haematopoietic stem cell niche and control of the niche size", *Nature*, vol. 425, no. 6960, pp. 836-841.

Zhang, P., Iwasaki-Arai, J., Iwasaki, H., Fenyus, M. L., Dayaram, T., Owens, B. M., Shigematsu, H., Levantini, E., Huettner, C. S., Lekstrom-Himes, J. A., Akashi, K., & Tenen, D. G. 2004, "Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha", *Immunity.*, vol. 21, no. 6, pp. 853-863.

Zhang, P., Zhang, X., Iwama, A., Yu, C., Smith, K. A., Mueller, B. U., Narravula, S., Torbett, B. E., Orkin, S. H., & Tenen, D. G. 2000, "*PU.1* inhibits *GATA-1* function and erythroid differentiation by blocking *GATA-1* DNA binding", *Blood*, vol. 96, no. 8, pp. 2641-2648.

Ziemin-van der, P. S., McCabe, N. R., Gill, H. J., Espinosa, R., III, Patel, Y., Harden, A., Rubinelli, P., Smith, S. D., LeBeau, M. M., Rowley, J. D., & . 1991, "Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias", *Proc.Natl.Acad.Sci.U.S.A*, vol. 88, no. 23, pp. 10735-10739.

Zipursky, A., Brown, E. J., Christensen, H., & Doyle, J. 1999, "Transient myeloproliferative disorder (transient leukemia) and hematologic manifestations of Down syndrome", *Clin.Lab Med.*, vol. 19, no. 1, pp. 157-67, vii.