STUDIES ON CEBPA MUTATIONS IN ACUTE MYELOID LEUKAEMIA

Claire Louise Green

UCL

A thesis submitted for the degree of Doctor of Philosophy

2012

DECLARATION

I, Claire Louise Green, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in this thesis.

Signed:

Date:

ABSTRACT

Acute myeloid leukaemia (AML) is a highly heterogeneous disease with regard to clinical outcome, and molecular markers with prognostic impact can be used to stratify patients for risk-adapted therapy. *CEBPA* mutations have been associated with a favourable prognosis, however several questions remained, in particular whether one (*CEBPA*-single) or two (*CEBPA*-double) mutations were necessary for this benefit, and their interaction with other molecular markers. A method of detecting *CEBPA* mutations in patient samples using denaturing HPLC was developed and the *CEBPA* status of 1427 young adult AML patients (median age 43 years, range 15-68 years) determined. Overall, 107 (7%) were *CEBPA*-mutant: 48 (45%) *CEBPA*-single and 59 (55%) *CEBPA*-double. The majority of *CEBPA*-double patients (83%) had an out-of-frame insertion/deletion in the N-terminus and a mutation in the C-terminal DNA-binding/leucine zipper domains (DBD/LZD) that were on different alleles as determined by cloning. By contrast, mutations in *CEBPA*-single cases were distributed across the gene. *CEBPA*-double patients were less likely to have a *FLT3*/ITD (P=.04) and highly unlikely to have an *NPM1* mutation (P<.0001) compared to *CEBPA*-WT/*CEBPA*-single cases. Eight year overall survival (OS) was higher in *CEBPA*-double patients compared to *CEBPA*-WT and *CEBPA*-single cases (54%, 34%, 31%, respectively, P=.004). In multivariate analyses, *CEBPA*-double, but not *CEBPA*-single, was an independent favourable factor for OS (P=.004) and relapse (P=.02). However, this benefit was completely lost in the presence of a *FLT3*/ITD. The mutant level of 101 mutations was determined by fragment analysis and the majority were of a level consistent with a heterozygous mutation present in most cells. The impact of ten atypical *CEBPA* mutations on C/EBPα transactivation activity was explored by a luciferase reporter assay. Only mutations affecting the DBD or LZD functional domains had an impact on transactivation activity. This work provides insight into the biology of *CEBPA* mutations and their use as clinical markers.

ACKNOWLEDGMENTS

I would like to thank my supervisors, Professors Rosemary Gale and David Linch, who have been so supportive and who gave me the opportunity to work towards a doctorate alongside my technician role in the department. I have learnt so much in the six years that I have spent working with them, and am grateful for the time they have so generously spent providing advice and helpful comments throughout the preparation of this thesis. I would also like to thank Kenneth Koo, who worked with me screening patients for *CEBPA* mutations. Without his hard work it would have taken even longer than the eighteen months it took to complete all the mutation analysis. These studies would not have been possible if not for all the patients consenting to enter the trials and allowing their samples to be used for research, the clinicians who entered and managed the patients, those who have managed the trials over the years and those who have diligently collected and stored the samples. Another large thank you is owed to my parents. They have provided love and support, giving financial aid and proof reading chapters, as well as kindly refraining from asking when I would be finished writing! Finally, I would like to thank my lovely fiancé, Rich. He has helped me through this whole process, put up with me during the more stressful moments and was patient throughout all the evenings and weekends spent writing.

TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER 1: INTRODUCTION

CHAPTER 2: MATERIALS AND METHODS

CHAPTER 3: DETECTION OF CEBPA MUTATIONS IN AML

CHAPTER 4: PROGNOSTIC IMPACT OF CEBPA MUTATIONS IN YOUNG ADULT PATIENTS

CHAPTER 5: QUANTIFICATION OF CEBPA MUTANT LEVEL AND CONFIRMATION OF BIALLELIC MUTATIONS

CHAPTER 6: TRANSACTIVATION POTENTIAL OF NON-CLASSICAL *CEBPA* MUTATIONS

LIST OF TABLES

CHAPTER 1: INTRODUCTION

CHAPTER 3: DETECTION OF CEBPA MUTATIONS IN AML

CHAPTER 4: PROGNOSTIC IMPACT OF CEBPA MUTATIONS IN YOUNG ADULT PATIENTS

CHAPTER 5: QUANTIFICATION OF CEBPA MUTANT LEVEL AND CONFIRMATION OF BIALLELIC MUTATIONS

CHAPTER 6: TRANSACTIVATION POTENTIAL OF NON-CLASSICAL *CEBPA* MUTATIONS

APPENDIX

COMMONLY USED ABBREVIATIONS

AA	Amino acid
AK	Abnormal karyotype
Allo-SCT	Allogeneic stem cell transplantation
AML	Acute myeloid leukaemia
APL	Acute promyelocytic leukaemia
ATRA	*All-trans*-retinoic acid
BM	Bone marrow
BMT	Bone marrow transplant
bp	Base pairs
CBF	Core binding factor
cDNA	Complementary DNA
CEBPA	CCAAT-enhancer binding protein α
CI	Confidence intervals
CIR	Cumulative incidence of relapse
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CMV	Cytomegalovirus
CR	Complete remission
dHPLC	Denaturing high-performance liquid chromatography
DNA	Deoxyribonucleic acid
DBD	DNA-binding domain
DFS	Disease-free survival
eBFP2	Enhanced blue fluorescent protein 2
EFS	Event-free survival
FAB	French-American-British
FLT3	Fms-like tyrosine kinase-3
G-CSF	Granulocyte-colony stimulating factor
gDNA	Genomic DNA
GEP	Gene expression profile
GFP	Green fluorescent protein
GMP	Granulocyte/macrophage progenitor
GVHD	Graft-versus-host disease
GVL	Graft-versus-leukaemia
HSC	Haemopoietic stem cell

HLA	Human leukocyte antigen
HR	Hazard ratio
ID	Induction death
ITD	Internal tandem duplication
LIC	Leukaemia initiating cell
LSC	Leukaemic stem cell
LZD	Leucine zipper domain
MDS	Myelodysplasia
MEP	Megakaryocyte/erythroid progenitors
MLL	Myeloid/lymphoid leukaemia
MK	Monosomal karyotype
MM	Mismatch
MRC	Medical Research Council
MRD	Minimal Residual Disease
MSCV	Murine stem cell virus
NGS	Next generation sequencing
NK	Normal karyotype
NPM1	Nucleophosmin
nt	Nucleotide
OR	Odds ratio
OS	Overall survival
PCR	Polymerase chain reaction
RD	Resistant disease
RED	Restriction enzyme digest
RFS	Relapse-free survival
RQ-PCR	Real-time quantitative polymerase chain reaction
RTK	Receptor tyrosine kinase
SCT	Stem cell transplantation
SEM	Standard error of the mean
SNP	Single nucleotide polymorphism
TA	Transactivation activity
TAD	Transactivation domain
TKD	Tyrosine kinase domain
TRM	Treatment-related mortality
WBC	White blood cell count
WHO	World Health Organization
WT	Wild-type

CHAPTER 1: INTRODUCTION

Acute myeloid leukaemia (AML) is an aggressive malignancy characterised by an accumulation of malignant white cells of the myeloid lineage in the bone marrow (BM) and peripheral blood of patients. Whilst the presence of these leukaemic blast cells is the common feature between cases, there is considerable heterogeneity between patients with respect to a number of characteristics, such that AML may more accurately be described as a collection of related malignancies. Variable features include morphology, immunophenotype, karyotype and gene mutations, and this variability is reflected in the significant differences that may be seen in the response to therapy and longer term prognosis between patients.

This thesis describes the investigation of mutations in *CEBPA*, a gene coding for the CCAAT/enhancer binding protein α, in AML and detailed information regarding C/EBPα structure and function and *CEBPA* mutations is given in the introductions to the relevant chapters. This chapter will seek to place the work into context by giving an overview of AML with respect to its presentation, classification and treatment. The heterogeneity of the disease and the factors which are currently known to be important in patient outcome will also be introduced.

1.1 Haemopoiesis

The leukaemic blast cells that accumulate in the BM and peripheral blood of patients are the result of disorder in the normal process of haemopoiesis. Haemopoiesis is the formation of blood cells of different lineages from haemopoietic stem cells (HSCs), and an overview is given in Figure 1.1. This process occurs throughout life and, in adults, the primary sites of haemopoiesis are in the BM of the central skeleton and the proximal ends of the femurs and humeri. HSCs are rare, estimated to be only 1 in 20 $x10^6$ BM cells, pluripotent and have the capacity for self-renewal, providing the homeostatic maintenance of the haemopoietic system (Kondo *et al*, 2003). The current roadmap suggests that HSCs generate the multipotent common myeloid progenitor (CMP) and common lymphoid progenitor (CLP) cells, which can differentiate into cells of the myeloid and lymphoid lineages, respectively (Orkin & Zon, 2008). The CMP produces megakaryocyte/erythroid progenitors (MEPs), which then give rise to erythrocytes and megakaryocytes, and granulocyte/macrophage

[Copyright protected image removed]

Figure 1.1 Overview of haemopoiesis, figure from Orkin and Zon (2008) Haemopoiesis is the process of formation of mature blood cells by differentiation from the pluripotent long-term haemopoietic stem cells (LT-HSC) and short-term HSCs (ST-HSC), to multipotent progenitors, committed precursors and ultimately terminally differentiated cells. This process requires a number of different transcriptions factors and the stages at which haemopoiesis is blocked in the absence of a given transcription factor, as determined through conventional gene knockouts, are indicated by red bars. Abbreviations: CMP, common myeloid progenitor; CLP, common lymphoid progenitor; MEP, megakaryocyte/erythroid progenitor; GMP, granulocyte/macrophage progenitor; RBCs, red blood cells.

progenitors (GMPs), which can differentiate through more lineage-restricted committed precursor cells into terminally differentiated cells, such as neutrophils and macrophages. As cells become more differentiated, they lose self-renewal capacity and pluripotency which are characteristics of the HSC. This process relies upon the closely regulated activity of specific transcription factors and haemopoietic growth factors.

1.2 AML

1.2.1 Leukaemic stem cells and leukaemogenesis

The leukaemic stem cell (LSC) generating leukaemic blast cells in AML is analogous to the hierarchical arrangement of HSCs, committed progenitors and terminally differentiated blood cells in normal haemopoiesis (Bonnet & Dick, 1997). Leukaemia initiating cells (LICs), the putative LSCs, were first described through xenotransplantation assays, where the immature $CD34^{+}CD38^{-}$ fraction of human AML tumour cells, when transplanted into severe combined immunodeficient (SCID) mice, was uniquely able to initiate leukaemic engraftment and produce large numbers of mature blast cells (Lapidot *et al*, 1994). More recent data has found that both the $CD34^{+}$ and $CD34^{-}$ fractions of primary AML samples may contain cells able to initiate leukaemia in mice (Taussig *et al*, 2010). LSCs are cells in the tumour that have stem cell-like properties, in particular self-renewal and repopulation capacity, with potential for these characteristics varying between LSCs (Dick, 2008). LSCs may also be quiescent, and it is hypothesised that their dormancy may confer drug resistance to conventional chemotherapy agents, allowing survival of the LSC, which can then cause disease relapse in patients who enter remission (Dick, 2008). The normal haemopoietic cell from which the LSC is derived, the cell of origin, is the subject of some discussion (Dick, 2008). The cell of origin may be a multipotent HSC which has undergone leukaemic transformation, as suggested by the immature phenotype in common between LICs and HSCs. Alternatively, a more committed myeloid progenitor cell may have acquired some stem-cell properties, in particular self-renewal capacity, during leukaemogenesis, with the maturation state of the leukaemic blasts a reflection of the maturation state of the cell of origin (Krivtsov *et al*, 2006;Goardon *et al*, 2011).

The transformation of normal haemopoietic cells to leukaemic cells involves the acquisition of multiple genetic alterations, such as chromosomal translocations or gene mutations, over a period of time, in a multistep process (Kelly & Gilliland, 2002). The advent of next generation sequencing technologies has enabled the sequencing of individual cancer

genomes, revealing large numbers (1,000 to 100,000) of somatically acquired genetic alterations, which may be divided into driver and passenger mutations (Stratton *et al*, 2009). Driver mutations are defined as conferring a selection advantage on the cells that acquire them and are generally in cancer-associated genes. Passenger mutations do not confer a selection advantage, but may have been acquired in a cell before it acquired a driver mutation. It is hypothesised that most cancers have more than one driver mutation and that their frequency varies between cancer types, with an estimate of between 5 and 20 in an individual cancer (Stratton *et al*, 2009). These genetic "hits" disrupt normal cellular processes and their accumulation in a clone results in a cancerous phenotype, identified by the "hallmarks of cancer". These may include the ability to sustain proliferative signalling, evade growth suppressors, resist apoptosis and induce replicative immortality (Hanahan & Weinberg, 2011). In AML, for example, an increase in cell proliferation and/or survival may be caused by activating mutations in the growth factor receptor fms-like tyrosine kinase 3 (FLT3) or signal transduction pathway components NRAS and KRAS. Another characteristic feature of AML is a block in the normal differentiation programme and subsequent apoptosis. This is caused by "hits" such as the fusion proteins produced by the chromosomal translocations t(8;21) and t(15;17), and through mutations in haemopoietic transcription factors, such as C/EBPα and RUNX1 (Kelly & Gilliland, 2002). The growth advantage the leukaemic blast cells have over normal cells cause out-growth of the leukaemic clone and accumulation of leukaemic blasts in the peripheral blood and BM.

1.2.2 Presentation and aetiology

The result of the accumulation of immature blast cells is BM failure. This leads to a broad range of symptoms at presentation, for instance, fatigue and breathlessness caused by anaemia, infections, particularly of the chest, mouth and skin, caused by neutropenia and bleeding caused by thrombocytopenia. Infiltration of organs such as the liver, spleen and lymph nodes may also occur.

The incidence of AML is approximately 2-3 cases per 100,000 people per year, although this incidence rises with age to about 12-15 cases per 100,000 people per year in patients in their seventh and eighth decades (Burnett *et al*, 2011). In the majority of cases there is no known direct cause, although there is an association with irradiation, smoking and certain conditions which pre-dispose to AML, such as Down's syndrome. The most commonly identified causes are progression from a pre-existing myeloproliferative or myelodysplastic disease (secondary AML) or as a result of prior chemotherapy for a different malignancy (therapy-related AML).

1.3 Classification of AML

Historically, AML has been classified using the French-American-British (FAB) sub-types (Bennett *et al*, 1976;Bennett *et al*, 1985). These are primarily based on the morphology and cytochemistry of the leukaemic cells, with at least 30% blasts in the BM required. The different sub-types are given in Table 1.1. It has also long been known that non-random chromosomal abnormalities can be present in tumour cells, occurring in approximately 60% of patients (Grimwade *et al*, 1998), and including either structural gains or loss of chromosomal material as well as balanced translocations. Certain chromosomal abnormalities are associated with particular FAB types, such as t(8;21)(q22;q22) with M2, t(15;17)(q22;q12) with M3 and inv(16)(q23) with M4Eo.

In recent times, however, more detailed genetic information has been obtained, and this, in combination with cytogenetic, morphological, and immunophenotypic information, was used by the World Health Organization (WHO) to develop a classification system, first reported in 2001 (Jaffe *et al*, 2001), which was updated recently (Table 1.2) (Swerdlow *et al*, 2008). In this scheme, acute leukaemia is defined as at least 20% blasts in the BM, although this limit is not applied if one of the recurrent genetic abnormalities (Table 1.2) is detected. There are 6 main groups within this classification: AML with myelodysplasia-related changes, myeloid neoplasms related to therapy, those related to Down's syndrome, AML not otherwise specified, and AML with recurrent genetic abnormalities. The latter group largely comprises those with specific cytogenetic abnormalities, such as t(8;21) or t(15;17), however, there are also two provisional entries for patients with mutations in the nucleophosmin (*NPM1)* or *CEBPA* genes. While most patients can be classified by a specific genetic abnormality, around a third of cases are placed, by default, into the group of "AML, not otherwise specified". However, an ever increasing number of gene mutations are being detected in patients at diagnosis which will, likely, aid in the classification of these cases in the future, leading to amendments to the current scheme.

Table 1.1 French-American-British (FAB) classification of AML

FAB subtype	Name
M0	Undifferentiated acute myeloblastic leukaemia
M1	Acute myeloblastic leukaemia with minimal maturation
M2	Acute myeloblastic leukaemia with maturation
M3	Acute promyelocytic leukaemia
M4	Acute myelomonocytic leukaemia
M4Eo	Acute myelomonocytic leukaemia with eosinophilia
M5	Acute monoblastic leukaemia
M6	Acute erythroid leukaemia
M7	Acute megakaryoblastic leukaemia

Table 1.2 WHO classification of AML*

Acute myeloid leukaemia with recurrent genetic abnormalities	AML with t(8;21)(q22;q22); *RUNX1-RUNX1T1*
	AML with inv(16)(p13.1;q22) 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)(q21;q26.2) 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 leukaemia with myelodysplasia-related changes	
Therapy-related myeloid neoplasms (t-AML)	
Acute myeloid leukaemia, not otherwise specified	AML with minimal differentiation
	AML without maturation
	AML with maturation
	Acute myelomonocytic leukaemia
	Acute monoblastic/monocytic leukaemia
	Acute erythroid leukaemia
	Acute megakaryoblastic leukaemia
	Acute basophilic leukaemia
	Acute panmyelosis with myelofibrosis
Myeloid sarcoma	
Myeloid proliferations related to Down's syndrome	Transient abnormal myelopoiesis
	Myeloid leukaemia associated with Down's syndrome

*Adapted from Vardiman *et al* (2009)

1.4 Factors associated with patient outcome in AML

The prognosis of patients with AML is highly heterogeneous, and there are a number of features that are known to be associated with particular clinical outcomes. There is considerable interest in understanding factors that may be used to stratify patients by risk and thereby develop risk-adapted protocols for personalised therapy. For example, given the toxicity of conventional chemotherapy it may be beneficial to reduce treatment intensity for patients with a favourable prognosis. By contrast, patients where a poor outcome is predicted could receive more intensive or experimental therapy. Some of the factors known to impact on patient outcome are introduced below, with a particular focus on those important for young adult AML patients, as this group of patients is the focus of the work in this thesis.

1.4.1 Clinical features

The highest incidence of AML is in patients greater than 60 years of age and higher age is an independent adverse factor for outcome, with very poor long-term survival rates for older patients of less than 10% (Appelbaum *et al*, 2006;Juliusson *et al*, 2009). When patients are stratified by age at diagnosis, from infants to the elderly, rates of remission are lower and overall survival decreases with increasing age (Smith *et al*, 2011). Other factors at diagnosis associated with an adverse impact on patient outcome include higher WHO performance status scores, secondary or therapy-related rather than *de novo* AML, and higher white blood counts (WBC) (Wheatley *et al*, 1999;Dohner *et al*, 2010). In addition to characteristics at diagnosis, the response to therapy and the course of the disease are of major importance for patient outcome. For instance, the length of remission is a prognostic factor, with early disease relapse within 6 months of achieving remission associated with poorer patient outcome (Dohner *et al*, 2010), and a higher percentage of BM blasts after induction therapy also negatively impacts upon prognosis (Wheatley *et al*, 1999). Response to therapy and the maintenance of remission may also be monitored through assessment of minimal residual disease (MRD). MRD positivity after induction or consolidation therapy is associated with higher relapse rates and poorer patient survival (Grimwade *et al*, 2010b), moreover, molecular relapse can be detected prior to morphological relapse (Schnittger *et al*, 2009). Techniques for MRD detection include flow cytometry, used to identify cells with particular immunophenotypes, and real-time quantitative polymerase chain reaction (RQ-PCR) assays, which can be used to detect leukaemia-specific targets, such as fusion genes or particular mutations (Grimwade *et al*, 2010b).

1.4.2 Cytogenetics

Non-random chromosomal changes are present in the leukaemic cells of approximately 60% of AML patients, as defined by analysis of metaphase spreads from BM aspirates, and karyotype at diagnosis has long been known to be an important determinant of patient outcome. Analysis of 1,612 patients aged up to 55 years treated in the UK Medical Research Council (MRC) AML 10 trial enabled classification of patients by karyotype into one of three prognostic risk groups: favourable, intermediate or adverse (Table 1.3) (Grimwade *et al*, 1998). The MRC classification was recently refined through analysis of 5,876 young adult patients aged 16-59 years treated in three successive MRC trials (Table 1.3) (Grimwade *et al*, 2010a). In general, there is a great deal of concordance between the MRC classification and those devised by other study groups (Grimwade & Hills, 2009).

The favourable-risk group accounts for approximately one quarter of young adult patients, and is defined by the presence of the t(15;17), t(8;21) or inv(16)/t(16;16) chromosomal translocations. The latter two abnormalities cause the transcription of the fusion proteins RUNX1-RUNX1T1 and CBFB-MYH11, respectively, and are known as core-binding factor (CBF) leukaemia abnormalities. Patients with favourable-risk cytogenetics have a significantly better long-term overall survival (OS) compared to patients with a normal karyotype (NK), in addition to lower rates of relapse in t(8;21) and t(15;17) cases (Grimwade *et al*, 2010a). There is some discussion regarding the prognostic impact of additional abnormalities in the CBF leukaemias, for instance the HOVON/SAKK classification excludes inv(16) cases with additional unfavourable abnormalities from the favourable-risk group (Cornelissen *et al*, 2007). However, no adverse effect was observed for additional abnormalities in CBF leukaemias in the very large cohort analysed by the MRC (Grimwade *et al*, 2010a).

The adverse-risk group comprises approximately 10% of young adult patients, and there are a number of different karyotypes that are associated with poorer patient outcome (Table 1.3). These include monosomies of chromosomes 5 or 7, and deletions of the long arm of chromosome 5, which are designated as myelodysplasia-related (MDS-related) changes in the 2008 WHO classification (Table 1.2). In young adults, these karyotypes are associated with very poor rates of remission (less than 60%), high rates of relapse, and a very

Table 1.3 Cytogenetic risk group classification of young adult AML

Risk group	Proportion of patients[1]	Original MRC[1]	Refined MRC[2]
Favourable	12%	t(15;17)	t(15;17)
	8%	t(8;21)	t(8;21)
	3%	inv(16) or t(16;16)	inv(16) or t(16;16)
Intermediate	42%	Normal karyotype	Normal karyotype
	25%	Other non-complex	Other non-complex
Adverse	10%	abn(3q)	abn(3q) excluding t(3;5)
			inv(3) or t(3;3)
		-5 or del(5q)	add(5q), del(5q) or -5
		-7	-7, add(7q) or del(7q)
			t(6;11)
			t(10;11)
			t(9;22)
			t(11q23) excluding t(9;11) and t(11;19)
			-17 or abn(17p)
		Complex (≥ 5 unrelated abn)	Complex (≥ 4 unrelated abn)
		Excluding those with favourable changes	Excluding those with favourable changes

[1]Grimwade *et al*, 1998. [2]Grimwade *et al*, 2010a. Abbreviations: MRC, Medical Research Council; abn, abnormality; del, deletion; add, addition; t, translocation.

poor long-term overall survival of less than 15% at either 4 or 10 years (Grimwade *et al*, 2010a). The presence of certain translocation partners involved in 11q23 abnormalities is also associated with poor patient outcome. This locus is commonly involved in reciprocal translocations in AML, normally targeting the myeloid/lymphoid leukaemia (*MLL*) gene, and patients with either t(6;11) (Krauter *et al*, 2009) or t(10;11) (Grimwade *et al*, 2010a) have been reported to have a survival of less than 15% at 10 years, although these abnormalities were not individually identified in the original MRC classification. In addition to these specific abnormalities, a complex karyotype has also been found to be associated with poor patient outcome. Whilst the original MRC classification defined this as 5 or more unrelated abnormalities (Grimwade *et al*, 1998), this was recently updated to 4 or more unrelated abnormalities, as this was found to be the most useful cut-off (Grimwade *et al*, 2010a). However, a number of other classification systems use a definition of 3 or more unrelated abnormalities (Grimwade & Hills, 2009). A further group of patients with a very poor clinical outcome are those with a monosomal karyotype (MK), with a 4-year OS of 4% (Breems *et al*, 2008). An MK is defined as the presence of either two or more autosomal monosomies, or a single autosomal monosomy in combination with at least one structural abnormality. Whilst an MK is not recognised separately within the MRC classification, nearly all patients satisfying the criteria for an MK were found to have karyotypes placing them in the adverse-risk group (Grimwade *et al*, 2010a).

Patients with an NK or karyotypic abnormalities that are neither adverse- nor favourable-risk are generally assigned, by default, to the intermediate-risk group. Typical abnormalities in intermediate-risk karyotype patients include trisomies of chromosomes 8 or 21, and deletions of 13q. Approximately 40% of AML patients have an NK, so that this single karyotype comprises the largest number of patients in the intermediate-risk cytogenetics group. Despite the large number of patients classified as intermediate-risk, karyotype at diagnosis currently remains a key way to stratify patients according to risk.

In addition to the large number of different cytogenetic abnormalities found in AML patients, there is also variation in diagnostic karyotypes according to patient age. In older patients, it has been observed that adverse-risk karyotypes are more common and favourable-risk abnormalities less common than in younger adult patients (Grimwade *et al*, 2001). The distribution of chromosomal abnormalities is also different in children with AML, compared to adults. For instance, whilst the incidence of an NK is lower in paediatric patients at approximately 25%, the incidence of certain abnormalities is higher, such as translocations affecting the *MLL* locus (Harrison *et al*, 2010). It has also been shown that the prognostic impact of cytogenetics may be affected by patient age, for instance it has been

reported that a complex karyotype does not have a significant effect on outcome in paediatric patients (Harrison *et al*, 2010).

1.4.3 Gene mutations

There are a large number of genes that have been found to be recurrently mutated in the leukaemic cells of AML patients, which, in combination with the variety of chromosomal abnormalities, is the basis of the vast heterogeneity seen in this disease. Kelly and Gilliland proposed a two-hit model hypothesis, which divides mutations and other genetic alterations into two classes, I and II (2002). Class I mutations are those which confer a proliferative and/or survival advantage to the leukaemic cells. These include mutations in genes considered to be classical oncogenes, for instance activating mutations in genes for growth factor receptors, such as the receptor tyrosine kinases (RTKs) *FLT3* (fms-like tyrosine kinase 3) and *c-KIT* (proto-oncogene c-KIT), and in components of signal transduction pathways important for proliferation, such as the *RAS* family members neuroblastoma-*RAS* (*NRAS*) and Kirsten-*RAS* (*KRAS*). Class II mutations impair normal differentiation and typical class II mutations include loss-of-function mutations in the genes encoding transcription factors important for normal haemopoietic differentiation, such as *RUNX1* and *CEBPA.* Kelly and Gilliland hypothesised that class I mutations collaborate with class II mutations to cause an AML phenotype. Indeed, mouse models of leukaemogenesis have been able to show cooperation between different genetic lesions, such as the combination of *CEBPA* and *FLT3* mutations (Reckzeh *et al*, 2012).

However, a two-hit model is probably an over-simplification for most leukaemias in humans, with evidence that tumourigenesis in human cells may require perturbation of six pathways, compared to just two in murine cells (Rangarajan *et al*, 2004). The advent of next generation sequencing (NGS) technology has allowed DNA sequencing of a cancer genome or exome, which has revealed a much broader spectrum of genes affected by mutations. The sequencing of an AML genome was first published in 2008 (Ley *et al*, 2008), and subsequent studies of other AML genomes or exomes have, over a very short period of time, facilitated the detection of recurrent mutations in genes that would not necessarily have been predicted to have a role in leukaemogenesis, for instance, mutations in the metabolic enzyme isocitrate dehydrogenase 1 (*IDH1*) in 2009 (Mardis *et al*, 2009), and the DNA methyltransferase *DNMT3A* in 2010 (Ley *et al*, 2010). In addition, these studies have given an indication of the number of acquired mutations in known coding regions or splice sites in an AML genome, which varies from case to case but was approximately 21 per patient in one study (Ding *et al*, 2012).

The association of the presence of specific gene mutations with patient outcome has long been an area of intensive research. Whilst for some molecular markers a general consensus has been reached regarding their impact on patient prognosis, for others, either a lack of impact has been reported, or the impact remains under investigation, with conflicting data reported. Further complexity is added due to the frequent presence of multiple mutations in the same patient, with the particular combination of mutations often important for patient outcome. *FLT3* and *NPM1* are the two genes most commonly utilised as molecular markers in clinical practice, particularly in the context of an intermediate-risk karyotype, due to their frequency, impact on patient outcome and utility for patient management. These are highly relevant markers for the work in this thesis, and are introduced below in sections 1.4.3.1 and 1.4.3.2, respectively. *CEBPA* mutation status may be investigated during the diagnostic work-up of a new AML case, and *CEBPA* mutations and their impact on patient prognosis are explored in detail in chapters 3 and 4. Some of the other genes that are commonly mutated in AML are briefly introduced in section 1.4.3.3, along with a summary of the impact they are reported to have, if any, on the outcome of patients. However, this is not an exhaustive list, particularly given the rate at which novel gene mutations are being detected.

1.4.3.1 *FLT3*

FLT3 is a class III tyrosine kinase receptor, which is normally expressed on the surface of haemopoietic progenitor cells and plays an important role in cell proliferation, differentiation and survival (Small, 2006). There are two classes of acquired mutations described in *FLT3*, internal tandem duplications (ITD, also known as length mutations) in the juxtamembrane domain, and tyrosine kinase domain (TKD) mutations.

FLT3/ITDs were first described by Nakao *et al* in 1996, and are present in approximately 25% of young adult AML patients, rising to approximately a third of those with an NK (Small, 2006). They are in-frame insertions of between 3 and more than 400 base pairs (bp), and cause constitutive activation of the receptor through disruption of the negative regulatory activity of the juxtamembrane domain. *FLT3*/ITD mutations are associated with poor patient prognosis, including higher rates of relapse and lower OS (Small, 2006). In addition, there is evidence that loss of the normal *FLT3* allele, caused by uniparental disomy (Raghavan *et al*, 2005;Griffiths *et al*, 2005), is associated with a particularly poor outcome (Gale *et al*, 2008).

FLT3/TKD mutations are generally missense mutations or small in-frame size changes, primarily affecting residues Asp835 and Ile836. These are less common than *FLT3*/ITD mutations, occurring in approximately 7% of patients (Abu-Duhier *et al*, 2001;Yamamoto *et al*, 2001), but are also known to cause aberrant receptor activation through constitutive tyrosine phosphorylation. However, the outcome of *FLT3*/TKD-mutant patients remains controversial, with either no significant impact of a *FLT3*/TKD mutation on OS (Yamamoto *et al*, 2001;Frohling *et al*, 2002), an adverse impact (Thiede *et al*, 2002), a favourable impact (Mead *et al*, 2007), or an adverse impact in certain subgroups (Bacher *et al*, 2008) all variously reported.

1.4.3.2 *NPM1*

NPM1 is a nuclear histone chaperone which shuttles between the nucleus and cytoplasm, and has a number of roles including in ribosome biogenesis, cell survival and centrosome duplication (Colombo *et al*, 2011). First detected in 2005, *NPM1* mutations are currently the most commonly identified gene mutations in AML, found in approximately 28% to 41% of all adult patients (Verhaak *et al*, 2005;Thiede *et al*, 2006;Gale *et al*, 2008), with the incidence rising to between 35% and 62% in cases with an NK (Falini *et al*, 2005;Schnittger *et al*, 2005;Dohner *et al*, 2005;Thiede *et al*, 2006;Gale *et al*, 2008). Mutations are heterozygous and are generally 4 bp insertions in exon 12 of *NPM1*, with the resultant frame-shift causing loss of the nuclear localisation signal and aberrant localisation of the protein to the cytoplasm. The latter is detectable by immunohistochemistry (Falini *et al*, 2005). The exact role of mutant *NPM1* in leukaemogenesis is still under investigation, however, loss of normal *NPM1* has been associated with genomic instability and accelerated oncogenesis in mice models (Grisendi *et al*, 2005).

NPM1 mutations are considered a favourable factor for patient prognosis (Schnittger *et al*, 2005;Dohner *et al*, 2005;Verhaak *et al*, 2005;Thiede *et al*, 2006;Gale *et al*, 2008). In addition, there is a correlation between the presence of *NPM1* mutations and *FLT3*/ITDs, with some debate over the interaction between the two markers. Whilst it is widely accepted that the presence of an *NPM1* mutation without a *FLT3*/ITD is a favourable genotype and a *FLT3*/ITD without an *NPM1* mutation is unfavourable, the outcome of patients either lacking or positive for both markers remains controversial. Data from our own large cohort of young adult AML patients found no evidence for an interaction between the two mutations, with a benefit seen for an *NPM1* mutation in both *FLT3*/ITD-positive and negative cases, and an overall intermediate prognosis for those with either both mutations or neither (Gale *et al*, 2008). In other reported cohorts, the favourable effect of an *NPM1*

mutation was not seen in the presence of a *FLT3*/ITD (Schnittger *et al*, 2005;Dohner *et al*, 2005;Verhaak *et al*, 2005;Thiede *et al*, 2006).

1.4.3.3 Other recurrently mutated genes in AML

Apart from *FLT3*, *NPM1* and *CEBPA* mutations there are a number of other genes recurrently mutated in AML, and a selection of these are summarised in Table 1.4. Abnormalities in several of these genes are generally associated with an adverse impact on patient prognosis, such as mutations of *DNMT3A*, *WT1*, *TET2* or *RUNX1*, and partial tandem duplications of the *MLL* gene (Table 1.4). However, not all gene mutations have been associated with patient outcome, for instance *NRAS* and *KRAS* mutations, where no impact on prognosis has been observed (Bowen *et al*, 2005). In addition, there remains considerable debate regarding the clinical impact of certain mutations, such as *IDH1* and *IDH2*, with different outcomes and interacting mutations reported by different study groups (Table 1.4).

Whilst many of the recurrent mutations, such as *DNMT3A*, *IDH1*, *IDH2* and *WT1* mutations, are associated with intermediate-risk karyotypes, in particular an NK, some are more common in those with particular chromosomal abnormalities. For instance, activating mutations in the *c-KIT* receptor tyrosine kinase are frequent in patients with the favourable-risk CBF leukaemias, and may have an adverse effect on prognosis, although there is evidence that the impact may depend upon the type of *c-KIT* mutation and the karyotype, inv(16) or t(8;21) (Table 1.4). By contrast, loss-of-function mutations in the tumour suppressor gene *TP53* are associated with adverse-risk karyotypes. In one study, 60% of patients with complex karyotypes had a *TP53* mutation, and these were found to predict a particularly dismal patient outcome (Rucker *et al*, 2012). In another study, there were no significant differences in outcome between *TP53* mutant or WT

Table 1.4 Summary of selected recurrent gene mutations detected in AML

Gene	Function	Commonly detected mutations	Approximate incidence	Impact on outcome	References
DNMT3A	DNA methyltransferase 3A: Catalyses the *de novo* methylation of CpG dinucleotides	Missense Nonsense Frame-shift Splice-site	18% - 23% 29% - 34% (NK)	Adverse	Ley *et al*, 2010 Thol *et al*, 2011 Yan *et al*, 2011 Marcucci *et al*, 2012 Renneville *et al*, 2012 Ribeiro *et al*, 2012
IDH1	Isocitrate dehydrogenase 1: Cytosolic metabolic enzyme catalysing conversion of isocitrate to α-ketoglutarate	Missense at Arg132	4% - 11%	Unclear: • Favourable in *FLT3*/ITD-positive and adverse in *FLT3*/ITD-negative cases • No impact • Adverse when analysed with *IDH2* mutant cases in NK/*NPM1*-mutant/*FLT3*-ITD-negative patients	Mardis *et al*, 2009 Abbas *et al*, 2010 Boissel *et al*, 2010 Chou *et al*, 2010 Green *et al*, 2010 Ho *et al*, 2010 Paschka *et al*, 2010 Wagner *et al*, 2010
IDH2	Isocitrate dehydrogenase 2: Mitochondrial metabolic enzyme catalysing conversion of isocitrate to α-ketoglutarate	Missense at Arg140 or Arg172	9% - 12%	Unclear: • R140 favourable and R172 adverse • Any mutant favourable • No impact	Abbas *et al*, 2010 Marcucci *et al*, 2010 Paschka *et al*, 2010 Thol *et al*, 2010 Chou *et al*, 2011b Green *et al*, 2011

Table 1.4 Continued

Gene	Function	Commonly detected mutations	Approximate incidence	Impact on outcome	References
TET2	Tet oncogene family member 2: Catalyses conversion of 5-methylcytosine to 5-hydroxymethylcytosine	Missense Nonsense Frame-shift	8% - 13% 23% (NK)	Unclear: • Adverse • No impact	Chou *et al*, 2011a Metzeler *et al*, 2011 Gaidzik *et al*, 2012
WT1	Wilms' tumour 1: Transcription factor expressed in immature haemopoietic cells	Missense Nonsense Frame-shift	10% - 13% (NK)	Unclear: • Adverse • No impact	Paschka *et al*, 2008 Virappane *et al*, 2008 Gaidzik *et al*, 2009
RUNX1	Runt-related transcription factor 1: Haemopoietic transcription factor frequently involved in chromosomal translocations	Missense Nonsense Frame-shift	6% - 13%	Adverse	Tang *et al*, 2009 Gaidzik *et al*, 2011 Schnittger *et al*, 2011b
MLL	Myeloid/lymphoid leukaemia gene: Frequently involved in chromosomal translocations	Partial tandem duplications (PTD)	5% - 8%	Adverse	Schnittger *et al*, 2000 Dohner *et al*, 2002 Steudel *et al*, 2003
NRAS	Neuroblastoma-*RAS*: GTPase component of signal transduction pathways	Missense at Gly12, Gly13 or Gly61	10% - 11%	No impact	Bowen *et al*, 2005 Bacher *et al*, 2006

Table 1.4 Continued

Gene	Function	Commonly detected mutations	Approximate incidence	Impact on outcome	References
KRAS	Kirsten-*RAS*: GTPase component of signal transduction pathways	Missense at Gly12, Gly13 or Gly61	5%	No impact	Bowen *et al*, 2005
c-KIT	Proto-oncogene *c-KIT*: Class III receptor tyrosine kinase, receptor for stem cell factor	Missense (activation loop or transmembrane domain) In-frame indels (extra-cellular domain) In-frame insertions/tandem duplications (juxtamembrane domain)	17% - 46% (CBF)	Adverse Evidence impact affected by: • Type of mutation • Karyotype, t(8;21) or inv(16)	Care *et al*, 2003 Boissel *et al*, 2006 Cairoli *et al*, 2006 Paschka *et al*, 2006 Allen *et al*, 2011
TP53	Encodes the tumour suppressor p53, a regulator of the cell cycle	Missense Nonsense Frame-shifts In-frame indels	53% - 60% (Complex karyotype)	Adverse	Bowen *et al*, 2009 Rucker *et al*, 2012

Abbreviations: NK, normal karyotype; CBF, core binding factor leukaemias; *NPM1*, nucleophosmin; *FLT3*/ITD, internal tandem duplications of the fms-like tyrosine kinase 3.

complex karyotype patients, however, 5 year overall survival was extremely poor for all patients at 0% and 2%, respectively (Bowen *et al*, 2009).

In addition to associations of mutations with particular karyotypes, mutations in different genes may show positive or negative correlations with one another. For instance, *NPM1* mutations are commonly detected in the same patients as *FLT3*/ITDs (see section 1.4.3.2) and they also frequently co-exist with either *IDH1* or *IDH2* mutations, with an *NPM1* mutation detected in 65% of *IDH1*-mutant and 61% of *IDH2*-mutant patients in our own cohort (Green *et al*, 2010;Green *et al*, 2011). By contrast, *IDH1*, *IDH2* and *TET2* mutations are generally mutually exclusive in patients, thought to be due to overlapping roles in leukaemogenesis in that they may all disrupt the same epigenetic pathway (Figueroa *et al*, 2010). The particular combination of mutations is also often important, for instance clinical outcome is different between patients with an *NPM1* mutation with or without a *FLT3*/ITD (see section 1.4.3.2). *FLT3*/ITD status also determined the impact of an *IDH1* mutation on outcome in our own study. In the whole cohort there was no difference in prognosis between *IDH1*-mutant and *IDH1*-WT cases. When patients were stratified by their *FLT3*/ITD status there was a beneficial impact of a mutation in *FLT3*/ITD-positive cases, but a negative impact in *FLT3*/ITD-negative cases (Green *et al*, 2010).

Another observation that can be made from these studies is that different mutations within the same gene may differ in their impact on patient outcome. This is the case for *FLT3*/ITD and *FLT3*/TKD mutations, with a poor patient outcome consistently reported for *FLT3*/ITD-positive patients and a more controversial impact on prognosis seen for *FLT3*/TKD mutations (see section 1.4.3.1), which is likely a reflection of the different impact that the mutations have on signalling (Small, 2006). Whilst the prognostic impact of *IDH2* mutations remains the subject of debate (Table 4.1), there are several reports indicating that the outcome of patients with *IDH2*-R140 mutations is very different from those with *IDH2*-R172 mutations. In our own study of patients enrolled on the UK MRC AML 10 and AML 12 trials, *IDH2* mutations were associated with improved clinical outcome, however when patients were stratified by the type of *IDH2* mutation it was found that only those with *IDH2*-R140 mutations had a good outcome, comparable to those with favourable-risk cytogenetics. Patients with mutations affecting *IDH2*-R172 had a very poor long-term outcome (Green *et al*, 2011). Evidence from other study groups also shows an *IDH2*-R172 mutation confers an extremely poor prognosis (Boissel *et al*, 2010;Marcucci *et al*, 2010). The functional basis for this difference is not yet clear, however, as both mutations have been reported to cause loss of the normal enzymatic function catalysing the conversion of

isocitrate to α-ketoglutarate and gain of a neomorphic function in converting α-ketoglutarate to 2-hydroxyglutarate (Ward *et al*, 2010).

NGS of AML genomes or exomes has enabled the detection of mutations in numerous other genes in tumour cells, many of which remain to be validated as leukaemogenic. The evolution of a tumour from diagnosis to relapse may also be explored through NGS, with intra-tumour heterogeneity indicated by the identification of subclones within the founding clone that may predominate at relapse (Ding *et al*, 2012). The complexity caused by the variety of mutations and the way in which they may interact with other markers means that determining the prognostic value of gene mutations remains a challenge. This is particularly true if a marker is only present in a small proportion of patients, and large cohorts of patients are therefore required for meaningful subgroup analyses.

1.4.4 Other molecular markers

In addition to gene mutations, there are a number of other molecular markers that have been associated with patient prognosis. These may be genome-wide patterns of gene expression, microRNA expression and methylation, or expression of individual genes or microRNAs (Marcucci *et al*, 2011). Of these, gene expression has been the most heavily investigated and genome-wide gene expression profiles (GEPs), generated by microarray-based techniques, have been used to identify particular patterns of gene expression in cohorts of AML samples. These signatures can be characteristic and predictive of sub-types of AML that are already defined by a prognostically relevant molecular marker, for instance inv(16) or *CEBPA* mutations, although patients with markers such as *MLL* rearrangements or *FLT3* mutations are not accurately predicted (Valk *et al*, 2004;Mrozek *et al*, 2009). The GEP may also provide additional prognostic information to that from known molecular markers, for instance a prognostic score base on the GEP of a particular probe set has been developed that is predictive of patient outcome in NK AML (Metzeler *et al*, 2008). Altered expression of several genes, such as *BAALC*, *MN1*, *ERG* and *EVI1*, has also been linked to prognosis in AML (Marcucci *et al*, 2011). However, at present genome-wide expression profiling remains a tool for research, rather than for clinical practice, largely due to difficulties in standardisation of assays and establishment of validated cut-offs.

1.5 Treatment of AML

1.5.1 Chemotherapy

The backbone of treatment in AML is intensive chemotherapy with drugs such as the nucleoside analogue cytarabine, in combination with an anthracycline such as daunorubicin, and this has been the standard of care for the last 40 years (Burnett *et al*, 2011). However, these drugs have limited specificity for leukaemic cells over normal BM cells, and their toxic effects cause periods of severe marrow failure and pancytopenia. For this reason, a number of patients are not considered fit for intensive therapy, in particular older patients (≥ 60 years) in whom comorbidities and resistant disease are more frequent, and a more palliative approach may be adopted. Where intensive chemotherapy is considered appropriate, the aim is to induce a complete remission (CR), usually defined as fewer than 5% blasts in the BM with recovery of blood cell counts. CR is generally achieved after induction therapy in between 70% and 80% of younger patients (< 60 years), which decreases to between 40% and 65% of older patients (Burnett *et al*, 2011). Once in CR, induction therapy is then consolidated with further courses of chemotherapy, with the hope of eliminating the disease. Throughout treatment, supportive care for BM failure is essential, for instance through fungal and antibiotic prophylaxis or transfusion support, as required. Whilst remission is achieved in most cases, the majority will relapse within 2 to 3 years of presentation, with relapse more likely in older patients: 50% to 55% of younger patients compared to around 85% of older patients within this time frame (Burnett *et al*, 2011).

1.5.2 Stem cell transplantation

Haemopoietic stem cell transplantation (SCT) is an alternative option to further chemotherapy. SCT from a human leucocyte antigen (HLA)-matched sibling donor (sibling allograft) has long been part of standard care and can be effective at preventing relapse (Cornelissen *et al*, 2007), due in part to a potent graft-versus-leukaemia (GVL) effect. However, this is an aggressive therapy associated with significant treatment-related mortality (TRM), due to both the intensive treatment required to ablate the recipients' BM, as well as the effects of the donated cells. Complications of allografts (allo-SCT) include graft-versus-host disease (GVHD), graft failure and severe infections, in particular cytomegalovirus (CMV)-associated interstitial pneumonitis. The risk of these complications increases with age and, therefore, this treatment is normally reserved for younger patients, with some evidence that the benefit of an allo-SCT is restricted to patients less than 35 years of age (Cornelissen *et al*, 2007).

The choice of whether to deploy an allo-SCT in an individual patient is a balance between the relapse risk of the disease if treated with chemotherapy alone versus the risk of the procedure itself, with the benefit for a transplant seen when relapse risk at 4 years is greater than approximately 35% (Cornelissen *et al*, 2007). An assessment of the risk of transplant in an individual case can be quantified in a risk score (Gratwohl *et al*, 2009), which takes into account factors that impact on the likelihood of a successful transplant. These include the age of the donor and recipient, donor-recipient sex combination, disease stage, time interval from diagnosis to transplant, donor type (HLA-identical sibling or unrelated) and CMV status of the donor and recipient (Gratwohl *et al*, 2009;Dohner *et al*, 2010). The relapse risk of the disease is reliant on a number of factors, which are discussed in detail in section 1.4. The relapse risk of the disease if treated with chemotherapy alone is also important in determining the timing of the transplant, in particular whether it should be performed in the first or second remission. For instance, if it is considered likely that a second remission can be achieved if a patient relapses, a rationale exists for delaying transplant to the second remission.

Reduced-intensity conditioning (RIC) transplants, which involve immunosuppression rather than ablation of the host BM, are also being explored. This is due to the limited applicability of traditional transplant regimens, in particular for older patients. Lacking the intensive conditioning of a standard allo-SCT, RIC transplants rely upon the GVL effect to maintain remission. There is also some evidence that this is a feasible approach in older patients, or those with comorbidities, with full chimeric engraftment observed (Hegenbart *et al*, 2006;Herr *et al*, 2007).

An alternative to allo-SCT is an autologous SCT, especially in cases where an HLA-matched donor is not available. Autologous SCT is associated with a relative lack of toxicity, and complications such as GVHD are avoided. However, this approach lacks the benefit of a GVL effect, and the possibility remains that leukaemic cells contaminating the graft will contribute to relapse. Whilst autologous SCT as a post-induction therapy can be at least as effective as intensive chemotherapy, there remains some debate as to which groups of patients may benefit, and at which stage of treatment transplant should be deployed (Breems & Lowenberg, 2007).

1.5.3 Targeted therapies

Greater understanding of the molecular basis of AML has identified newer targets for therapy. One of these is aberrant signalling through receptor tyrosine kinases (RTKs) in leukaemic cells, caused by mutations in genes such as *FLT3* and *c-KIT*. This has led to the development of tyrosine kinase inhibitors, primarily as inhibitors of FLT3, although most also inhibit other RTKs (Fathi & Levis, 2011). Many of these RTK inhibitors are already being tested in combination with chemotherapy as part of multi-centre phase III clinical trials, for instance midostaurin in a Cancer and Leukemia Group B (CALGB) study and lestaurtinib (CEP-701) in the UK National Cancer Research Institute (NCRI) AML 15 and AML 17 trials. These agents seem to be well tolerated, although the patient benefit remains to be fully determined (Fathi & Levis, 2011). Other small molecule inhibitors under early investigation include the aminopeptidase inhibitor Tosedostat, the mammalian target of rapamycin (mTOR) inhibitor voreloxin and demethylating agents, such as azacytidine and decitabine (Burnett *et al*, 2011).

1.5.4 Treatment of acute promyelocytic leukaemia

Acute promyelocytic leukaemia (APL) is a special sub-type of AML characterised by the translocation t(15;17)(q22;q12), and patients with APL are treated differently to those with other types of AML. In this translocation, part of the promyelocyte leukaemia gene (*PML*) on chromosome 15 is fused to the retinoic acid receptor α gene (*RARA*) on chromosome 17, leading to translation of the PML-RARα fusion protein. This fusion protein is a transcriptional repressor of both RARα and non-RARα target genes and disrupts PML nuclear bodies, causing proliferation and the inhibition of terminal differentiation (de The & Chen, 2010). It has been recognised for nearly 30 years that the presence of t(15;17) predicts for sensitivity to treatment with *all-trans*- retinoic acid (ATRA) (de The & Chen, 2010). Together, ATRA and arsenic trioxide induce differentiation and apoptosis of APL cells through transcriptional de-repression and degradation of the PML-RARα fusion protein (de The & Chen, 2010). Current therapy, combining treatment with ATRA, arsenic trioxide and an anthracycline, such as idarubicin, can cure up to 90% of patients, with relapse rates as low as 5% (Grimwade *et al*, 2009;de The & Chen, 2010).

1.6 Combining prognostic factors and risk-adapted therapy

As detailed above, there are many factors that can impact upon patient outcome, and while karyotype at diagnosis has provided the most effective way of stratifying cases, the intermediate-risk group remains the largest and most heterogeneous. Efforts have been made to combine different prognostic markers so that clinical, cytogenetic and other molecular data can be utilised to refine stratification of patients into risk-groups beyond those defined by cytogenetics alone, with the aim of helping therapy decisions. A risk score has been developed by analysis of data from MRC trial patients which combines patient and disease characteristics (age, sex and *de novo* or secondary AML), cytogenetics, and response to induction therapy, to delineate three patient risk groups (Burnett *et al*, 2006). An alternative risk score for NK AML has been developed using age, WBC, and specific molecular markers: mutations in *FLT3*/ITD, *NPM1* and *CEBPA*, a *WT1* polymorphism and expression of *BAALC*, *ERG*, *MN1* and *WT1* (Damm *et al*, 2011). Whilst this risk score enabled stratification of patients into three groups with different clinical outcomes, whether it is necessary to investigate such a large panel of molecular markers in all patients with an NK remains to be determined, especially given the cost implications. The European LeukemiaNet currently recommends a molecular risk reporting system which combines cytogenetic data with the mutation status of *FLT3*/ITD, *NPM1* and *CEBPA* (Dohner *et al*, 2010). In this classification, the favourable-risk genetic group includes the CBF leukaemias and NK patients with either mutated *CEBPA* or mutated *NPM1* without a *FLT3*/ITD. The large intermediate-risk group is further divided into two groups, with the intermediate-I group containing NK patients with all other *NPM1* and *FLT3*/ITD genotypes and the intermediate-II group made up of those with a t(9;11)(p22:q23) translocation and all other cytogenetic abnormalities not classified as favourable- or adverse-risk. There is evidence that this is a useful system for younger adult AML patients (Rollig *et al*, 2011), however newer markers, such as *DNMT3A* mutations, remain to be incorporated into such classification schemes.

One area where validated risk stratification schemes may be used is in making transplant decisions. Specific genetic markers can also be useful in deciding whether to perform a transplant in first remission or not. In those patients with favourable-risk cytogenetics, an allo-SCT has been shown to provide no benefit (Cornelissen *et al*, 2007). Conversely, there is evidence that a transplant from an HLA-matched sibling donor in first remission is beneficial for those with intermediate or adverse-risk cytogenetics, particularly younger patients, with the benefit for a transplant seen when relapse risk at 4 years is greater than approximately 35% (Cornelissen *et al*, 2007). Gene mutations with prognostic value have

also been investigated in the context of transplant decisions, particularly within the heterogeneous intermediate-risk cytogenetics group. For instance, in NK patients it has been reported that those with the favourable genotype of mutant *NPM1*, without an additional *FLT3*/ITD, gain no benefit from a transplant (Schlenk *et al*, 2008). By contrast, those with either a *FLT3*/ITD or those without an *NPM1* or *CEBPA* mutation had an improved relapse-free survival in this study if an HLA-matched related donor was available. However, the use of transplant in patients with *FLT3*/ITD mutations remains the subject of debate (Gale *et al*, 2005;Meshinchi *et al*, 2006;Bornhauser *et al*, 2007).

1.7 Aims of this thesis

This thesis presents data investigating *CEBPA* mutations in younger adult AML with respect to their incidence and types of mutations (chapter 3), the characteristics and clinical outcome of *CEBPA*-mutated patients, and the factors which modify this impact (chapter 4). The mutations detected are further explored in terms of their mutant level and allelic distribution (chapter 5), and the impact of non-classical mutations on C/EBPα protein function (chapter 6).

CHAPTER 2: MATERIALS AND METHODS

Specific methods are discussed in the relevant results chapter.

2.1 Molecular Biology

2.1.1 Reagents

Acetonitrile (VWR International Ltd., Lutterworth, UK)

Agar (Sigma-Aldrich Company Ltd., Poole, UK)

Agarose (Bioline, London, UK)

Betaine (Sigma-Aldrich Company Ltd., Poole, UK)

BIOTAQ DNA polymerase (Bioline, London, UK)

BIOTAQ DNA polymerase buffer and magnesium chloride (Bioline, London, UK)

Boric acid (VWR International Ltd., Lutterworth, UK)

Bromophenol blue (Merck, Frankfurt, Germany)

Carbenicillin (Melford Laboratories Ltd., Ipswich, UK)

Dimethylformamide (Sigma-Aldrich Company Ltd., Poole, UK)

Dimethyl sulfoxide (DMSO) (Sigma-Aldrich Company Ltd., Poole, UK)

dNTPs (Bioline, London, UK)

Ethylenediamine tetraacetic acid disodium salt (EDTA) (VWR International Ltd., Lutterworth, UK)

Ethidium bromide (Sigma-Aldrich Company Ltd., Poole, UK)

Glycerol (VWR International Ltd., Lutterworth, UK)

Luria-Bertani (LB) broth capsules (MP Biomedicals, London, UK)

One Shot Max Efficiency DH5α-T1 competent *E. coli* (Invitrogen Life Technologies, Paisley, UK)

Phusion Hot Start High-Fidelity DNA Polymerase (New England Biolabs UK Ltd., Hitchin, UK)

Phusion HF buffer (New England Biolabs UK Ltd., Hitchin, UK)

Primers, unlabelled (Integrated DNA Technologies, Leuven, Belgium)

QIAprep Spin Miniprep Kit and HiSpeed Plasmid Midi Kit (QIAGEN, Crawley, UK)

QIAquick Gel Extraction kit (QIAGEN, Crawley, UK)

QIAquick PCR Purification kit (QIAGEN, Crawley, UK)

Restriction enzymes and buffers (New England Biolabs UK Ltd., Hitchin, UK)

Super optimal broth with catabolite repression (SOC) (Invitrogen Life Technologies, Paisley, UK)

T4 DNA ligase (New England Biolabs UK Ltd., Hitchin, UK)

T4 DNA ligase buffer (New England Biolabs UK Ltd., Hitchin, UK)

TOPO cloning kit (Invitrogen Life Technologies, Paisley, UK)

Tri-ethylene ammonium acetate (TEAA) (Transgenomic Ltd., Glasgow, UK)

Tris base (Tris(hydroxymethyl)aminomethane) (VWR International Ltd., Lutterworth, UK)

WellRED oligos (Sigma-Aldrich Company Ltd., Poole, UK)

X-Gal (Invitrogen Life Technologies, Paisley, UK)

2.1.2 Polymerase chain reaction (PCR)

The PCR was used to amplify specific regions of a DNA template (genomic, complementary or plasmid) by cycling between temperatures that caused denaturation of the template DNA followed by annealing of sequence-specific forward and reverse oligonucleotide primers, which defined the region to be amplified, and finally an extension step of DNA polymerisation. A master mix of all required reagents, sufficient for all the samples in each PCR, was made and divided into aliquots before addition of the DNA template. This ensured consistency in the reaction mix between samples. A negative control, where water was added instead of DNA template, was included in all PCRs to ensure no contamination was present. There were two different DNA polymerases used during this work: BIOTAQ DNA polymerase (Bioline, London, UK) and Phusion Hot Start High-Fidelity DNA Polymerase (New England Biolabs UK Ltd., Hitchin, UK).

The standard reaction mix for a 20µl PCR using BIOTAQ DNA Polymerase contained 2µl 10x NH_4 reaction buffer (670nM Tris-HCl, 160mM $(NH_4)_2SO_4$, 100mM KCl, 0.1% stabiliser), 1mM $MgCl_2$, 200µM of each dNTP, 0.5µM each of the forward and reverse

primers, 0.5U BIOTAQ DNA Polymerase and 10-100ng DNA template. The annealing temperature used was specific to the primers and this information is given in the relevant chapters. The standard cycling conditions were 35 cycles of denaturation at 95°C for 30 seconds, annealing of primers for 30 seconds and extension at 72°C for 30 seconds, followed by a final extension incubation of 5 minutes at 72°C.

The standard reaction mix for a 20µl PCR using Phusion Hot Start High-Fidelity DNA Polymerase contained 4µl 5x Phusion HF buffer (containing 7.5 mM $MgCl_2$, providing 1.5mM $MgCl_2$ in the final reaction), 200µM of each dNTP, 0.5µM each of the forward and reverse primers, 0.2U Phusion Hot Start High-Fidelity DNA Polymerase and 10-100ng DNA template. The standard cycling conditions used were an initial denaturation step of 98°C for 2 minutes, then 35 cycles of denaturation at 98°C for 42 seconds, annealing of primers for 42 seconds and extension at 72°C for 42 seconds, followed by a final extension incubation of 5 minutes at 72°C. As indicated in the specific chapters, PCR additives such as betaine and DMSO were also added to PCR reaction mixes to enable amplification of difficult templates.

2.1.3 Agarose gel electrophoresis

Buffers

10x TBE: 108.9g Tris base, 55.7g boric acid and 7.4g EDTA in 1 litre of ddH_2O

5x Loading buffer: 30% glycerol and 0.025% bromophenol blue in 1x TBE

Agarose gel electrophoresis was used to both detect PCR products and to separate different fragments of DNA by size, for instance after restriction enzyme digestion. An appropriate amount of agarose (between 1% and 3% weight/volume, depending on the purpose) was dissolved in 35ml of 1xTBE by heating in a microwave oven. Once cooled slightly, 3.5µl of a 1mg/ml ethidium bromide solution was added before pouring into a mould. After the gel had set, it was covered in running buffer of 1x TBE containing the same concentration of ethidium bromide as the gel (0.1µg/ml). The desired volume of sample to be analysed was mixed with an appropriate volume of loading buffer before loading into a well in the gel. Samples were electrophoresed at a current of approximately 70mA for the required amount of time and fragments were detected by fluorescence of ethidium bromide intercalated in DNA under UV illumination. The result was documented in a digital photograph.

2.1.4 dHPLC analysis on the WAVE platform

Denaturing high-performance liquid chromatography (dHPLC) on the WAVE platform (Transgenomic Ltd., Glasgow, UK) was used to detect sequence changes in DNA PCR products. The process is outlined in Figure 2.1. In analysis by WAVE, the sample is injected into the flow path of buffer containing triethylammonium acetate (TEAA) and acetonitrile (ACN) and flows through a polystyrene-divinyl benzene copolymer DNA separation column located in an oven at the temperature required for analysis. TEAA acts as an ion-pairing agent and forms TEA+ ions in solution, which have both hydrophobic and hydrophilic portions. The positively charged TEA+ forms an association with the negatively charged phosphate backbone of the DNA creating a hydrophobic layer on the fragment that causes it to be attracted to the hydrophobic beads of the DNA separation column and leads to binding of the PCR products to the column. Over time, the proportion of ACN in the buffer is then increased in a specific gradient. As the ACN concentration increases, the ion-pairing properties of the TEAA decreases, eventually causing the elution of DNA from the column. Heteroduplexes containing mismatched bases adopt different conformations and bind less strongly to the column than the homoduplexes, and so are released from the column first followed by the homoduplexes. DNA eluting from the column passes through a UV detector, which records the absorbance at 260 nm. Results are displayed in the form of a chromatogram of absorbance over time. A sample containing only homoduplexes should produce a single peak on the chromatogram. The presence of additional heteroduplexes will create extra peaks on the chromatogram, for example the four peaks shown in Figure 2.1. This is dependent on the particular sequence and chromatograms can be characteristic of specific sequence alterations. All WAVE analysis runs performed in these studies contained PCR product with known wild-type (WT) sequence to enable a direct comparison between the WT WAVE chromatogram and that of test samples.

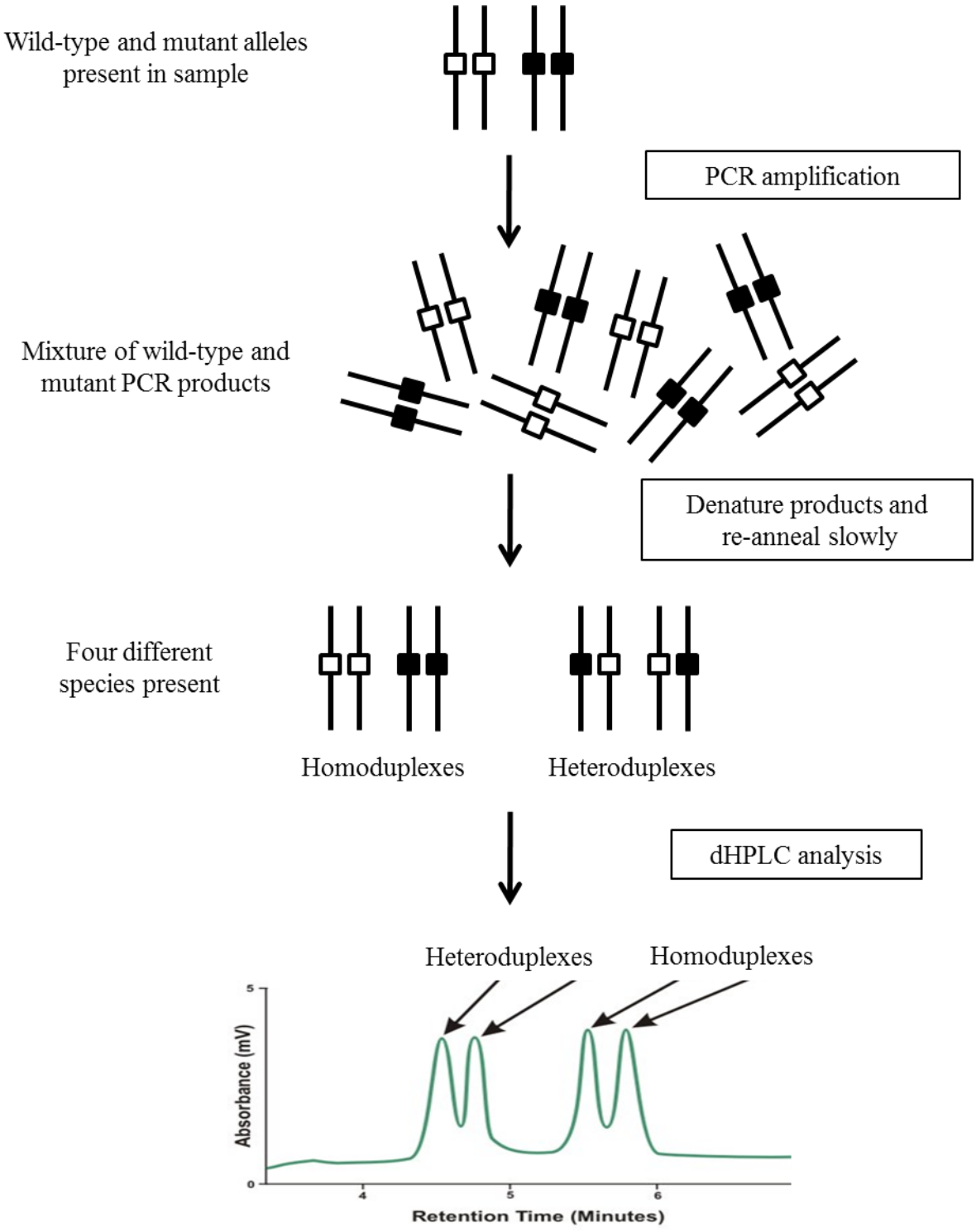


Figure 2.1 Overview of heteroduplex generation and dHPLC analysis for mutation detection. The polymerase chain reaction (PCR) was used to amplify the region of interest and the resulting amplicons were denatured and then re-annealed slowly to enable heteroduplex formation. These were then analysed by dHPLC at a specific temperature. Heteroduplexes eluted from the column and were detected first, followed by the homoduplexes.

PCR was used to amplify the region of interest from a template DNA or complementary DNA (cDNA), using a proof-reading enzyme, where possible, to reduce errors in base inclusion and thus reduce potential false heteroduplex formation. The presence of PCR products was checked by agarose gel electrophoresis. PCR products were denatured by incubation at 95°C for 5 minutes followed by 40 cycles of 1 minute incubations starting at 92°C and reducing by 1.5°C per cycle. This gradual cooling encouraged heteroduplex formation if a mutation was present, as shown in Figure 2.1. If only one type of allele, either wild-type (WT) or mutant, was present in the sample then just one homoduplex species was produced.

Navigator software (Transgenomic Ltd) was used to determine the temperature(s) at which to analyse the sample, which is dependent on the melting profile predicted by the software from the sequence of the amplicon. In general, a temperature was chosen at which the proportion of PCR product predicted to be helical was between 50% and 95%. The same sample may need to be analysed at more than one temperature depending on the percentage helicity across the amplicon.

Where enrichment of a particular species was required, a fraction collector was used in combination with WAVE analysis, by which specific fractions of the eluate from the column could be collected into separate vials. The fractions could then be further analysed by re-amplification by PCR, using the eluate as a template, for nucleotide sequencing. For samples where the heteroduplex species was present at a low level, WAVE analysis and collection was performed at a denaturing analysis temperature that enabled the collection and enrichment of the heteroduplex species. For samples with size changes where the sequence was difficult to determine by direct sequencing, WAVE analysis and collection at a non-denaturing temperature of 50°C allowed separation of different species by size, and so differently sized fragments could be collected in separate vials for further analysis.

2.1.5 Direct nucleotide sequencing

PCR products to be sequenced by standard Sanger sequencing were purified using the QIAquick PCR Purification Kit (QIAGEN, Crawley, UK). The concentration of DNA in the purified PCR product was quantified using a NanoDrop 1000 Spectrophotometer (Fisher Scientific UK Ltd., Loughborough, UK) and the product was then diluted with water to a concentration of 1ng/µl per 100 bp to be sequenced. This product was submitted, together with the appropriate primer at a concentration of 5pmol/µl, to the Scientific Support Service

at UCL Cancer Institute for direct nucleotide sequencing. Plasmids to be sequenced were submitted at a concentration of between 30ng/µl and 100ng/µl.

2.1.6 Restriction enzyme digests

All restriction enzymes were supplied by New England Biolabs (Hitchin, UK). Restriction digest reaction mixes and incubation temperatures were as according to manufacturers' instructions. Incubation times varied according to the required temperature: 37°C incubations were between 4 and 16 hours in duration, while incubations above 50°C were between 2 and 3 hours.

2.1.7 LB broth and agar plates

LB broth: 13 LB capsules in 500ml ddH_2O
LB agar plates: 13 LB capsules and 7.5g agar in 500ml ddH_2O

For both broth and plates, the reagents were autoclaved and cooled to below 50°C before addition of carbenicillin to a final concentration of 100µg/ml. Plates were then poured and left to set. Both broth and plates were stored at 4°C until required.

2.1.8 TOPO TA cloning of PCR products

The TOPO TA cloning kit (Invitrogen Life Technologies, Paisley, UK) enables cloning of PCR products produced by non-proof reading *Taq* DNA polymerases, such as BIOTAQ DNA Polymerase, because they add a single deoxyadenosine (A) to the 3´ ends of PCR products. The linearised pCR 2.1-TOPO vector is supplied complete with single 3´-thymidine (T) overhangs and topoisomerase I covalently bound to the vector. These "sticky ends" on PCR product and vector allow for easy ligation of the PCR product into the vector in a reaction catalysed by the topoisomerase. For the TOPO reaction, 1µl of PCR product was incubated with 1µl of the linearised pCR 2.1-TOPO vector, 1µl of the provided salt solution and 3µl of water at room temperature for 5 minutes.

2.1.9 Transformation of One Shot Max Efficiency DH5α-T1 *E.* coli competent cells

One vial of One Shot Max Efficiency DH5α-T1 *E. coli* cells was thawed on ice for each transformation. To each vial, 1µl of ligation reaction from section 2.1.8 or 10ng of circular

plasmid DNA, as required, was added and incubated on ice for 30 minutes. The cells were then heat-shocked for 30 seconds at 42°C, 250µl of SOC medium (2% tryptone, 0.5% yeast extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM glucose) was added per vial and the cells were incubated at 37°C in a shaking incubator for 1 hour. During this incubation, LB agar plates containing 100µg/ml carbenicillin were pre-warmed to 37°C. If the pCR 2.1-TOPO vector (section 2.1.8) was used for the transformation, plates were pre-coated with 40µl of a 40mg/ml X-gal solution in dimethylformamide to enable blue/white screening of colonies. In this system, the *LacZα* gene in the pCR 2.1-TOPO vector was disrupted if an insert was successfully ligated, as indicated by the formation of white colonies. If no insert was present, a functional *LacZα* gene was maintained and hydrolysis of X-gal caused the formation of blue colonies. Between 20µl and 100µl of transformed bacteria were spread per plate and plates were incubated overnight at 37°C to allow colony formation.

2.1.10 Identification and growth of transformed bacterial clones

To enable screening of individual bacterial colonies for the presence of particular inserts, colonies were plucked and seeded into 200µl of LB containing 100µg/ml carbenicillin in a well of a 96-well plate and incubated at 37°C overnight. The bacterial culture was then used as a template for PCR amplification, with specific primers and reaction conditions as given in the relevant chapters. Where plasmid isolation was required, the bacterial culture was used to inoculate 4ml of LB containing 100µg/ml carbenicillin, which was incubated overnight at 37°C with shaking. Glycerol stocks of clones were made by adding 150µl autoclaved glycerol to 850µl of the bacterial culture in a sterilised microfuge tube. This was vortexed thoroughly and stored at -80°C. Plasmid DNA from the remaining culture was extracted using the QIAprep Spin Miniprep Kit (QIAGEN, Crawley, UK) and the entire insert was sequenced to ensure that the constructs were correct. If larger amounts of plasmid DNA were required, the glycerol stock was used to inoculate 5ml of LB containing 100µg/ml carbenicillin and this was grown at 37°C in a shaking incubator for approximately 8 hours. This pre-culture was then added to 200ml of LB containing 100µg/ml carbenicillin and grown overnight. Plasmid DNA was extracted using the HiSpeed Plasmid Midi Kit (QIAGEN).

2.2 Cell Culture

2.2.1 Cell line

HEK 293T: Human embryonal kidney line, adherent, growth factor independent. Expresses the simian virus 40 (SV40) large T cell antigen

2.2.2 Cell culture general reagents

0.25% Trypsin- 0.02% EDTA solution (Sigma-Aldrich Company Ltd., Poole, UK)

Dulbecco's modified Eagle medium (DMEM) (PAA Laboratories Ltd., Yeovil, UK)

Foetal bovine serum (FBS) – heat inactivated (Sigma-Aldrich Company Ltd., Poole, UK)

GeneJuice Transfection Reagent (VWR International Ltd., Lutterworth, UK)

Penicillin (10,000 units/ml)-Streptomycin (10mg/ml) solution (Sigma-Aldrich Company Ltd., Poole, UK)

Phosphate buffered saline (PBS) (Sigma-Aldrich Company Ltd., Poole, UK)

2.2.3 Culture of 293T cells

293T cells were grown in 75cm^2 flasks (VWR International Ltd.) in DMEM supplemented with 10% FBS and 1% penicillin-streptomycin solution in a humidified atmosphere at 37°C with 5% CO_2. When cells reached 90-95% confluency, the growth medium was removed and the cell layer washed gently with DMEM. Trypsin-EDTA solution (3ml) was added to the cells and the flask incubated until cells dissociated from the flask. Cells were then re-suspended in 10ml of growth medium, transferred to a 15ml Falcon tube and centrifuged at 450g for 5 minutes. The supernatant was removed and the cells were re-suspended in 10ml of fresh growth medium. For maintenance, 2ml of the cell suspension was added to 18ml of growth medium in a fresh 75cm^2 flask.

2.2.4 Transient transfection of 293T cells

During passage of a 75cm^2 flask of 293T cells, 2ml of the cell suspension (section 2.2.3) was diluted in 12ml of growth medium per plate to be seeded. Each well of a 24-well or 6-well plate was seeded with 500µl or 2ml, respectively, of the diluted cell suspension and cells were left to adhere in a humidified atmosphere at 37°C with 5% CO_2 for at least 24hrs, until 50-80% confluent. For transfection of 6-well plates, 5µl of GeneJuice and 95µl of DMEM were used per well and for 24-well plates 1.25µl and 23.75µl, respectively. The total amount of GeneJuice and DMEM required for the whole experiment were mixed and incubated at room temperature for 5 minutes. Aliquots of appropriate volumes of GeneJuice in DMEM were prepared and the required volume of plasmid DNA added and incubated at room temperature for 15 minutes. The transfection mixture was added drop-wise to the plated cells. The plate was then gently rocked to mix and placed back in the incubator for 24 hours.

2.3 Cell lysis and immunoblotting

2.3.1 Reagents

30% (w/v) Acrylamide/ 0.8% bisacrylamide solution (National Diagnostics, Atlanta, USA)

Ammonium persulfate (APS) (Scientific Laboratory Supplies Ltd., Hessle, UK)

Bovine serum albumin (BSA) (PAA Laboratories Ltd., Yeovil, UK)

Bromophenol blue (Merck, Frankfurt, Germany)

Complete Mini Protease Inhibitor Cocktail Tablets (Roche Diagnostics Ltd., Burgess Hill, UK)

ECL Plus Western Blotting Detection Reagent (GE Healthcare UK Ltd., Little Chalfont, UK)

Full-range rainbow molecular weight markers (GE Healthcare UK Ltd., Little Chalfont, UK)

Detergent-Compatible (DC) Protein Assay (Bio-Rad Laboratories Ltd., Hemel Hempstead, UK)

Glycerol (VWR International Ltd., Lutterworth, UK)

Hybond-C Extra nitrocellulose membrane (GE Healthcare UK Ltd., Little Chalfont, UK)

Hyperfilm MP – High performance autoradiography film (GE Healthcare UK Ltd., Little Chalfont, UK)

2-Mercaptoethanol (Sigma-Aldrich Company Ltd., Poole, UK)

N,N,N′,N′-Tetramethylethylenediamine (TEMED) (Sigma-Aldrich Company Ltd., Poole, UK)

Non-fat milk powder (Marvel)

Nonidet-P40 (NP-40) (Sigma-Aldrich Company Ltd., Poole, UK)

Phosphate buffered saline (PBS) (Sigma-Aldrich Company Ltd., Poole, UK)

Sodium azide (Sigma-Aldrich Company Ltd., Poole, UK)

Sodium chloride (NaCl) (VWR International Ltd., Lutterworth, UK)

Sodium deoxycholate (VWR International Ltd., Lutterworth, UK)

Sodium dodecyl sulphate (SDS) (Scientific Laboratory Supplies Ltd., Hessle, UK)

Tris base (VWR International Ltd., Lutterworth, UK)

Tris-HCl (VWR International Ltd., Lutterworth, UK)

Tween-20 (Sigma-Aldrich Company Ltd., Poole, UK)

2.3.2 Buffers

Radioimmunoprecipitation (RIPA) buffer: 25mm Tris pH 7.6, 1% NP-40, 1% sodium deoxycholate, 150mM NaCl, 0.1% SDS

5x Laemmli sample buffer: 313mM Tris-HCl, 10% SDS, 50% glycerol, 25% 2-mercaptoethanol, 0.01% bromophenol blue

Running buffer: 25mM Tris, 250mM glycine, 0.1% SDS

Transfer buffer: 48mM Tris pH 9.1, 39mM glycine, 20% methanol, 0.038% SDS

TBST (Tris buffered saline Tween-20): 20mM Tris pH 7.5, 150mM NaCl, 0.05% Tween-20

Stripping buffer: 62.5mM Tris-HCl pH 6.7, 100mM 2-mercaptoethanol, 2% SDS

7x Protease inhibitor solution: 1 Complete Mini Protease Inhibitor Cocktail Tablet in 1.5ml water

2.3.3 Antibodies

Rabbit anti-C/EBPα #2295 (Cell Signaling Technology from New England Biolabs UK Ltd., Hitchin, UK)

Rabbit anti-C/EBPα antibody #2843 (Cell Signaling Technology from New England Biolabs UK Ltd., Hitchin, UK)

Murine anti-Tubulin (Sigma-Aldrich Company Ltd., Poole, UK)

Sheep anti-mouse IgG, ECL horseradish peroxidase-linked whole antibody (GE Healthcare UK Ltd., Little Chalfont, UK)

Donkey anti-rabbit IgG, ECL horseradish peroxidase-linked whole antibody (GE Healthcare UK Ltd., Little Chalfont, UK)

2.3.4 Preparation of cell lysates

Lysates were made from 293T cells that had been cultured in 6-well plates. All growth medium was removed and the cells washed twice with ice-cold PBS before the addition of 300µl per well of RIPA buffer containing an appropriate volume of protease inhibitor solution. The plate was placed on ice and rotated occasionally for 10 minutes. The cell lysates were then transferred to microcentrifuge tubes and centrifuged at 20,000g for 10 minutes at 4°C, after which the supernatants were transferred to clean microcentrifuge tubes and stored at -80°C until required.

2.3.5 Protein quantification

The concentration of protein in cell lysates was determined using the DC Protein Assay Reagents (Bio-Rad, Hemel Hempstead, UK). Protein standards of BSA in RIPA buffer at 6 different concentrations of between 0 and 1.5mg/ml were prepared. For every 1ml of the Assay Reagent A, 20µl of Reagent S was added to make Reagent A′, as the sample buffer contained detergent. In a 96-well plate, 25µl of Reagent A′ was added to 5µl of each standard or cell lysate. Reagent B, 200µl, was added to each well and samples were mixed and incubated for 15 minutes at room temperature. The absorbance was read at 750nm in a Varioskan Flash Multimode Reader (Thermo Fisher Scientific). Cell lysates were diluted 10-fold for quantification and each sample was measured in triplicate. The absorbance of the protein standards was used to generate a standard curve from which the protein concentration of cell lysates could be calculated.

2.3.6 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and semi-dry transfer

12.5% Running gel (for 2 gels): 2ml ddH_2O, 3.74ml 1M Tris pH 8.8, 4.2ml 30% (w/v) acrylamide/ 0.8% bisacrylamide solution, 100µl 10% SDS, 75µl 10% APS, 9µl TEMED

Stacking gel (for 2 gels): 3.49ml water, 625µl 1M Tris pH 6.8, 835ml 30% (w/v) acrylamide/ 0.8% bisacrylamide solution, 50µl 10% SDS, 38µl 10% APS, 7.5µl TEMED

Polyacrylamide gels were freshly made as required, with gel casting and electrophoresis performed using the Mini-PROTEAN Tetra Electrophoresis System (Bio-Rad Laboratories Ltd., Hemel Hempstead, UK). Reagents for a 12.5% running gel were mixed and pipetted into the glass plate casting module. This was covered with a layer of ddH_2O and the gel left to polymerise. Once set, the ddH_2O was poured off and the surface of the gel dried with fibre-free paper. The stacking gel was then pipetted on top and a comb inserted to form either 10 or 15 loading wells, as required, and left to polymerise. Cell lysates to be analysed were mixed with Laemmli sample buffer and boiled in a heat block for 5 minutes. The set gel was submerged in running buffer, and samples and a molecular weight marker were loaded into wells and electrophoresed at between 100V and 140V until the bromophenol blue dye front had travelled the required distance through the gel.

For semi-dry transfer of proteins, a nitrocellulose membrane and two pieces of Extra Thick Blot Paper (Bio-Rad) were cut to size and soaked in transfer buffer together with the gel for 10 minutes. These were then assembled in a sandwich in a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad) and the proteins were transferred from the gel to the membrane at 20V for 30 minutes.

2.3.7 Probing and detection

All membranes were blocked in 5% non-fat milk powder in TBST for between 1 and 2 hours at room temperature on a rocking platform. Incubation with the primary antibody was performed as specified and the membrane was then washed in TBST for at least 30 minutes, with the wash buffer changed at least 5 times. The appropriate secondary antibody, either anti-mouse IgG or anti-rabbit IgG, was used at a dilution of 1:10,000 in 3% non-fat milk powder in TBST and incubated with the membrane for 1-2 hours, then the membrane was washed as before. The membrane was then rinsed in PBS. Detection was performed using

the ECL Plus Western Blotting Detection Reagent kit and the enhanced chemiluminescence documented by exposure to autoradiography film. The film was then developed in an automated film developer (Konica Minolta SRX-101A).

2.3.8 Stripping

If required for re-probing, membranes were stripped to remove the original primary and secondary antibodies by incubation in stripping buffer for 10 minutes at 50°C. The membrane was then washed in TBST for at least 20 minutes, with the wash buffer changed at least 3 times. Blocking of the membrane was performed again before re-probing.

CHAPTER 3: DETECTION OF CEBPA MUTATIONS IN AML

3.1 Introduction

Transcription factors play a major role in the process of haemopoiesis and are up or down-regulated at specific points during differentiation as required for their target genes to be switched on or off. CCAAT enhancer binding protein alpha (C/EBPα) is one such transcription factor with a pivotal role in myelopoiesis.

3.1.1 C/EBPα structure and function

C/EBPα is coded for by the intronless gene *CEBPA*, found on the long arm of chromosome 19 at 19q13.1, and is the founding member of the C/EBP family of basic region leucine zipper (bZIP) transcription factors. The C/EBPα protein contains several domains that are characteristic of this family as shown in Figure 3.1. These include two transactivation domains (TAD1 and 2), and a DNA-binding domain (DBD) and leucine zipper domain (LZD) at the C-terminus. C/EBPα functions as a dimer, with dimerisation mediated by the LZD (Figure 3.2). This domain contains an α-helical structure that interacts with the α-helix in the binding partner to form a coiled-coil structure. C/EBPα can either form a homodimer or can heterodimerise with other family members, such as C/EBPβ and C/EBPδ, due to the conserved nature of the bZIP domain. Dimerisation is a pre-requisite for DNA binding, which is mediated by the DBD and occurs in the major groove of DNA (Figure 3.2) at sequence-specific sites, usually found in the promoters or enhancers of target genes. Once bound to DNA, the transactivation domains can function to activate transcription of target genes. There are two main isoforms of C/EBPα: the full-length 42 kDa protein (p42) and a truncated 30 kDa protein (p30). The p30 isoform is translated from an internal start site in the mRNA and the protein lacks the first 119 amino acids of the N-terminus, which includes TAD1. It has been found that the p30 protein has a lower transcriptional activation potential than the p42 protein (Pabst *et al*, 2001b).

C/EBPα is expressed in a number of different tissues including adipose, lung epithelium and liver, in which it plays several important roles, for instance in the generation of mature adipocytes (Ramji & Foka, 2002).

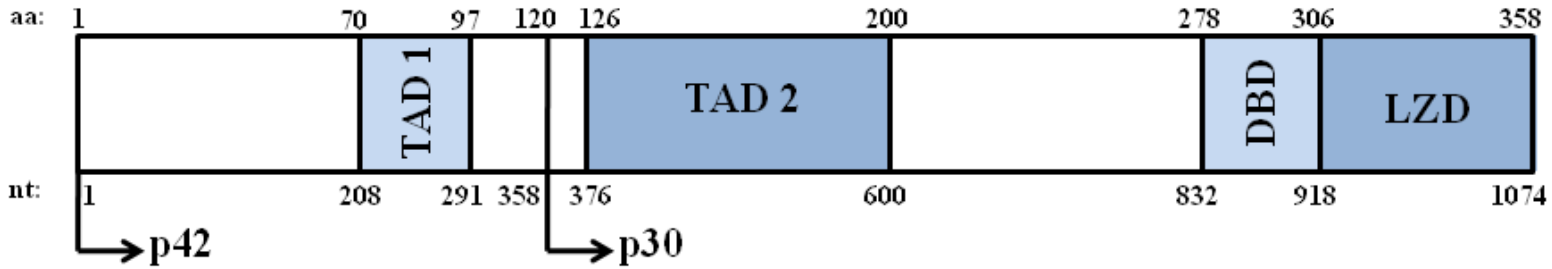


Figure 3.1 Functional domains of C/EBP*α*. C/EBPα contains functional domains that are typical of the C/EBP family of basic region leucine zipper transcription factors. These include transactivation domains 1 and 2 (TAD1 and TAD2), a DNA-binding domain (DBD) and a leucine zipper domain (LZD). Translational start sites for the two major isoforms p42 and p30, amino acid (aa) and nucleotide (nt) numbering are indicated. Adapted from Mueller *et al*, 2006.

It is also one of several transcriptions factors, including PU.1, C/EBPβ and C/EBPε, involved in myelopoiesis, where their expression is tightly controlled (Friedman, 2007). C/EBPα is predominantly expressed in early myeloid progenitors, in both common myeloid progenitors (CMP) and granulocyte/monocyte progenitors (GMP), and its level decreases through granulocytic differentiation (Ramji & Foka, 2002). By contrast, C/EBPε predominates in more mature granulocytes.

Studies of C/EBPα function *in vitro* and *in vivo* using mouse models have shed much light on the role it plays in haemopoiesis. Mice with a homozygous deletion of the gene coding for C/EBPα (*cebpa*) died within 8 hours of birth due to hypoglycaemia, caused by a lack of glycogen storage in the liver, and dramatically reduced lipid accumulation in either hepatocytes or adipose tissue (Wang *et al*, 1995). It was also observed that these mice had no mature granulocytes and an accumulation of myeloid blasts, although other haemopoietic lineages were unaffected (Zhang *et al*, 1997). Mice that were heterozygous for the deletion did not display this phenotype and were comparable to wild-type animals. Subsequent studies knocked out C/EBPα in adult mice and showed that the block in differentiation was at the CMP to GMP transition. However, C/EBPα was not required for differentiation after the GMP stage (Zhang *et al*, 2004).

As part of its role in granulopoiesis, there are a number of target genes that are regulated by C/EBPα in myeloid cells. These include genes that encode growth factor receptors such as the granulocyte colony-stimulating factor (G-CSF) receptor, granulocyte-macrophage colony-stimulating factor (GM-CSF) receptor and macrophage colony-stimulating factor (M-CSF) receptor to enable cells to respond to these cytokines. Other targets are those involved in granulocyte function, for instance myeloperoxidase, neutrophil elastase, lactoferrin and lysozyme, or genes associated with an acute phase response to a bacterial infection such as interleukin-6 or tumour necrosis factor-α (Ramji & Foka, 2002;Koschmieder *et al*, 2009).

C/EBPα not only promotes differentiation, it also arrests cell proliferation. This was initially observed in adipoblasts where C/EBPα expression caused a halt in mitotic growth (Umek *et al*, 1991), but subsequently has been described in other tissues.

[Copyright protected image removed]

Figure 3.2 Structure of the C/EBPα dimer-DNA complex. Crystal structure of the C-terminal amino acids (281 to 340) of rat C/EBPα bound as a dimer to a 21-mer DNA duplex. The leucine zipper (LZ) domain enables dimerisation through formation of a coiled-coil structure while the basic region (BR) mediates DNA sequence recognition and binding in the major groove of DNA. Figure from Miller *et al*, 2003.

The mechanism for this effect has been extensively studied and it was found that, in addition to activating the transcription of target genes, for instance early studies found that it increased p21 expression, a potent inhibitor of cyclin-dependent kinases (Timchenko *et al*, 1996), C/EBPα can interact directly with other proteins or complexes. Slomiany and colleagues reported that C/EBPα induced repression of E2F, a transcription factor that targets genes such as cyclins which drive cell cycle progression, and the two proteins were detected together in a complex (Slomiany *et al*, 2000). Additionally, C/EBPα that had been mutated so that the ability to repress E2F-dependent transcription was lost did not support either adipocyte or granulocyte differentiation in mice, indicating a crucial role for C/EBPα-induced E2F repression in terminal differentiation (Porse *et al*, 2001). C/EBPα also directly interacts with both cdk2 and cdk4, cyclin-dependent kinases important for progression through the cell cycle. A region of the TAD2 (Figure 3.1) of C/EBPα was found to mediate these interactions, which caused inhibition of the activity of these kinases (Wang *et al*, 2001), although this region was not required for C/EBPα-dependent growth regulation in a mouse model (Porse *et al*, 2006).

3.1.2 Mutations in the *CEBPA* gene in AML

A possible role for C/EBPα in AML, in which a key feature is a block in the differentiation of myeloid blasts, was investigated by Pabst and colleagues in 2001 given its involvement in myelopoiesis and cell cycle arrest (Pabst *et al*, 2001b). They sequenced the *CEBPA* gene in samples from 137 AML patients and found 8 (6%) cases with mutations predicted to cause non-synonymous amino acid changes. One patient had a nonsense point mutation and four had frame-shifting deletions in the N-terminus, one of whom also had an in-frame insertion in the C-terminus. These frame-shifts and the nonsense mutation all occurred before the translation initiation site for the truncated p30 isoform and it was shown that they caused loss of the full-length p42 and a concomitant increase in p30. Two patients had frame-shifting mutations after the p30 start site and one had a missense mutation in the N-terminus. All mutations, except the missense mutation in the N-terminus, caused either a complete lack of or reduced binding to a target DNA sequence, and most showed reduced transactivational potential compared to wild-type (WT) protein in a luciferase reporter assay. In patients with a single alteration, all mutations were found to be heterozygous, with one WT allele maintained. It was also shown *in vitro* that the p30 isoform had a dominant-negative effect over normal p42 in transactivation assays when they were co-expressed together.

After this first report of recurring mutations in *CEBPA* in AML patients, a number of other groups went on to replicate and extend these findings. Overall, it was found that mutations occurred at a frequency of between 6-18% in AML patients (Pabst *et al*, 2001b;Snaddon *et al*, 2003;Frohling *et al*, 2004;Bienz *et al*, 2005;Marcucci *et al*, 2008;Wouters *et al*, 2009;Pabst *et al*, 2009;Hou *et al*, 2009;Renneville *et al*, 2009a;Dufour *et al*, 2010). Putting these 10 studies together, of 3184 cases examined, 315 (10%) had at least one mutation in *CEBPA* and the ratio of cases with one mutation, hereafter called *CEBPA*-single, compared to those with two mutations (*CEBPA*-double) varied widely at between 1:2.1 and 6.7:1.

In general, two common types of mutations have been detected: out-of-frame insertions or deletions in the N-terminus predicted to cause loss of the p42 protein and an increase in p30 isoform expression, and in-frame insertions or deletions in the DBD or LZD, predicted to impair DNA-binding or dimerisation. *CEBPA*-double cases typically have both an N-terminal frame-shifting insertion or deletion and an in-frame insertion or deletion in the C-terminus.

In addition to the two classic types, mutations have also been found across the whole *CEBPA* gene and can be either point mutations or small or large insertions or deletions. The vast majority of mutations are heterozygous, affecting one allele only. However, homozygous mutations have been described in a small number of cases (6 of 2641, 0.2%) (Pabst *et al*, 2001b;Snaddon *et al*, 2003;Frohling *et al*, 2004;Bienz *et al*, 2005;Marcucci *et al*, 2008;Wouters *et al*, 2009;Pabst *et al*, 2009;Hou *et al*, 2009;Renneville *et al*, 2009a;Dufour *et al*, 2010), with uniparental disomy identified as a recurrent mechanism for their occurrence (Fitzgibbon *et al*, 2005;Wouters *et al*, 2007c).

CEBPA mutations have been shown to be relatively stable markers of disease with, in the majority of cases, the same *CEBPA* mutation(s) detected in samples taken at both presentation and relapse but not in remission (Tiesmeier *et al*, 2003;Lin *et al*, 2005;Shih *et al*, 2006). Studies of paired presentation and remission samples have also demonstrated germline *CEBPA* mutations in a small number of pedigrees with a history of familial AML. Of the 18 patients from 9 different kindreds who developed AML, all had a constitutional frame-shifting insertion or deletion in the N-terminus of *CEBPA*, with members of the same pedigree carrying the same mutation. At disease presentation, 14 (78%) of these patients had acquired an additional C-terminal mutation in the leukaemic cells. The acquisition of this second *CEBPA* mutation, and the specific DNA change involved, differed between affected members of the same pedigree (Smith *et al*, 2004;Sellick *et al*, 2005;Nanri *et al*,

2006;Corbacioglu *et al*, 2007;Pabst *et al*, 2008;Renneville *et al*, 2009b;Taskesen *et al*, 2011).

Single nucleotide polymorphisms (SNPs) are well known to occur in the *CEBPA* gene. The most common in the general population is the G/T SNP at nucleotide 690 (numbering from the A of the ATG for the p42 translation initiation codon as nucleotide 1), rs34529039, which codes for the silent T230T. The most prevalent allele in European populations is G, with G/T heterozygotes comprising about 25% (NCBI dbSNP). In addition, there are several reports of rarer point mutations also predicted to cause silent amino acid changes (Pabst *et al*, 2001b;Preudhomme *et al*, 2002;Frohling *et al*, 2004;Lin *et al*, 2005). Of greater interest is a duplication of 6 base pairs in a proline-rich region of TAD2 (584_589dup), predicted to cause H195_P196dup (also termed HP196_197ins or P194_H195dup). First recorded as a recurrent mutation (Frohling *et al*, 2004), it was found in both AML cases and 7 of 19 (39%) healthy volunteers by Lin *et al* (2005), who discounted it as pathogenic (Lin *et al*, 2005). This work was further extended by Wouters *et al* (2007b) who detected this change in 21 of 587 (4%) AML cases and 22 of 274 (8%) non-leukaemic blood samples, and found that the gene expression profiles of H195_P196dup AML cases did not cluster with that of other *CEBPA*-mutant cases (Wouters *et al*, 2007b). Additionally, H195_P196dup has been reported to lack prognostic value in AML (Biggio *et al*, 2008;Schnittger *et al*, 2011a), unlike *CEBPA* mutations, as will be discussed in chapter 4. The classification of H195_P196dup as a polymorphism has important implications for the assignment of nucleotide changes as pathogenic mutations, particularly when those changes are in-frame alterations in repetitive regions of the gene.

3.1.3 Techniques used to detect *CEBPA* mutations

There are several approaches that different groups have used to screen *CEBPA* for mutations, these are summarised in Table 3.1. The amino acid coding sequence of the *CEBPA* gene is a single exon of 1077 base pairs (bp) and most studies have used PCR to amplify the entire coding sequence in a small number of separate reactions to generate overlapping amplicons. Sanger sequencing of these PCR products, either directly or after cloning, has been the most commonly used technique to determine *CEBPA* mutation status. The advantage of nucleotide sequencing is that all types of mutation can be detected

Table 3.1 Studies of *CEBPA* mutation detection in patients with AML

Screening Technique	Study	Incidence of *CEBPA*-mutant (%)	Details
Sanger sequencing	Pabst *et al*, 2001b Lin *et al*, 2005 Hou *et al*, 2009	8 of 137 (6) 16 of 104 (15) 71 of 543 (13)	2 overlapping PCR products covering entire coding sequence generated and sequenced. 4 further products in cases of abnormal or ambiguous results.
	Gombart *et al*, 2002	8 of 78 (10)	4 overlapping PCR products covering entire coding sequence generated, cloned and sequenced.
	Preudhomme *et al*, 2002 Renneville *et al*, 2009a	15 of 135 (11) 53 of 638 (8)	PCR products generated as per Pabst *et al*, 2001b, cloned and sequenced.
	Frohling *et al*, 2004 Marcucci *et al*, 2008	33 of 236 (14) 32 of 175 (18)	4 overlapping PCR products covering entire coding sequence generated and sequenced.
	Bienz *et al*, 2005 Pabst *et al*, 2009	12 of 67 (18) 19 of 224 (8)	3 overlapping PCR products covering entire coding sequence generated and sequenced.
	Shih *et al*, 2006	22 of 149 (15)	2 overlapping PCR products covering entire coding sequence generated and sequenced.
Fragment analysis	Juhl-Christensen *et al*, 2008	20 of 446 (4)	4 overlapping fluorescently labelled PCR products covering entire coding sequence generated. Products for each patient pooled and fragment sizes analysed by capillary electrophoresis.
	Benthaus *et al*, 2008 Dufour *et al*, 2010	38 of 469 (8) 38 of 467 (8)	4 overlapping fluorescently labelled PCR products covering entire coding sequence generated in two multiplexed reactions. Products for each patient pooled and fragment sizes analysed by capillary electrophoresis.
Agarose gel analysis	Barjesteh *et al*, 2003	12 of 277 (4)	C-terminus only amplified and analysed by agarose gel electrophoresis for size changes. In positive cases, C-terminus and N-terminus amplified, cloned and sequenced.
PCR-SSCP analysis	Snaddon *et al*, 2003	8 of 99 (8)	4 overlapping PCR products covering entire coding sequence generated using primers as per Pabst *et al*, 2001b. Products analysed by SSCP and positive cases cloned and sequenced.
dHPLC	Wouters *et al*, 2009 Taskesen *et al*, 2011	41 of 598 (7) 151 of 1182 (13)	3 overlapping PCR products covering entire coding sequence generated. Heteroduplex analysis on dHPLC WAVE platform. Mutations confirmed by sequencing.

Abbreviations: PCR, polymerase chain reaction; dHPLC, denaturing high-performance liquid chromatography; SSCP, single-strand conformational polymorphism. Note: In all studies, cases with H195_P196dup or synonymous sequence changes only have been classified as *CEBPA*-WT.

whether they are heterozygous, homozygous, point mutations, insertions or deletions. However, the process is relatively labour-intensive and expensive, especially for a large cohort of patient samples. In addition, it can be easy to miss mutations in cases with low mutant level, for instance where the mutation was present in only a sub-clone of the tumour population.

Size separation of products by agarose gel electrophoresis to detect insertions or deletions has been utilised as a quick method of screening the C-terminus of *CEBPA* (Barjesteh van Waalwijk van Doorn-Khosrovani *et al*, 2003). However, this technique lacks sensitivity for smaller size changes and would not detect point mutations at all. Indeed, the proportion of *CEBPA*-mutant cases found in this report was lower than would be expected (4% compared to 10% in other studies). A more sensitive method of looking at size changes in the *CEBPA* gene is PCR with a labelled primer and fragment analysis by capillary electrophoresis. This technique has been utilised by several groups (Table 3.1) and has been shown to be effective at detecting low-level size changes and even single base pair insertions or deletions. However, mutations that do not cause an alteration in product length, both point mutations and indels that do not change the overall size, will be missed and these types of mutations are present in a significant number of *CEBPA*-mutant patients (Ahn *et al*, 2009). PCR-SSCP analysis has been utilised in a small study of AML patients (Snaddon *et al*, 2003), however it is technically demanding. A more sensitive method to detect different species within a PCR product is denaturing high-performance liquid chromatography (dHPLC).

The technique of dHPLC on the WAVE platform (Transgenomic Ltd, Glasgow, UK) for mutation detection relies on heteroduplex analysis and has been used effectively to determine *CEBPA* mutation status in a large series of AML samples (Wouters *et al*, 2009). Both point mutations and size changes can be detected with an appropriate degree of sensitivity and with a reasonably high throughput, given that analysis of a single PCR product takes between 3 and 7 minutes, depending on the DNA separation column used, as detailed in chapter 2. In addition, only samples with chromatograms different to that of a known WT sample will need to be subjected to direct nucleotide sequencing or another confirmatory method, minimising cost. Therefore, the method chosen to detect *CEBPA* mutations in this work was dHPLC on the WAVE platform.

This chapter outlines the optimisation of an effective method for screening the *CEBPA* gene in patient samples using dHPLC and the application of this method to a large, well-characterised cohort of younger adult AML patients.

3.2 Materials and Methods

3.2.1 Optimising PCR protocols

The initial step for detecting *CEBPA* mutations in patient samples by dHPLC was to obtain suitable PCR products from genomic DNA (gDNA) or complementary DNA (cDNA). This can be challenging because the GC content of the coding sequence is high at 75%, causing secondary structure formation that can interfere with DNA polymerase activity. Issues with GC-rich templates can be overcome through the use of PCR additives such as betaine, dimethyl sulfoxide (DMSO) and proprietary formulations e.g. PCRx Enhancer System (Invitrogen Life Technologies, Paisley, UK). However, for products to be analysed by WAVE, the DNA separation cartridge warranty limits the final concentration of betaine to 2.5M, DMSO to 10% and rules out the use of proprietary formulas. Another consideration for WAVE analysis is the preferred use of high-fidelity DNA polymerases that possess both 5′ to 3′ and 3′ to 5′ exonuclease activity. This is because small errors in DNA replication, especially in an early PCR cycle, can lead to spurious heteroduplex formation, which may interfere in the detection of heteroduplexes caused by mutations.

In order to look for mutations across the entire coding sequence, three pairs of primers were designed to generate overlapping amplicons called fragments 1, 2 and 3 so that all the coding sequence and a portion of the 5′ and 3′ untranslated regions were covered, as shown in Figure 3.3 A. The first primer pairs designed were CEBPA/1F and CEBPA/1R, CEBPA/2F and CEBPA/2R, and CEBPA/3F and CEBPA/3R for fragments 1, 2 and 3, respectively. All primer sequences are given in Appendix Table 1. Initial tests of the primer pairs were performed with a non-proof-reading enzyme BIOTAQ DNA Polymerase (Bioline Ltd, London, UK) using a standard reaction mix, apart from the addition of reagents to aid amplification of GC-rich templates, and standard cycling conditions as detailed in chapter 2. These initial PCRs tested various annealing temperatures and the addition of either betaine or DMSO at different final concentrations of up to 2.5M and 10% respectively. The aim was to generate a single band of the correct size and strong intensity as detected by agarose gel electrophoresis and ethidium bromide staining. The initial primer pairs were unsuccessful for both fragments 1 and 2 and it was necessary to design further forward and reverse primers for each fragment and test these with the original primers using a mix and match approach. In addition, touch-down PCR cycling protocols and the use of nested primers in two rounds of PCR were also investigated. After satisfactory preliminary primer pair tests using BIOTAQ DNA Polymerase were obtained, two DNA polymerases with proof-reading capabilities were tried: Optimase Polymerase (Transgenomic Ltd., Glasgow, UK) and

Phusion Hot Start High-Fidelity DNA Polymerase (New England Biolabs, Hitchin, UK). Once PCR protocols using a proof-reading DNA polymerase had been optimised, Navigator software (Transgenomic Ltd., Glasgow, UK) was utilised to generate amplicon-specific temperatures at which to analyse the products by dHPLC. For fragment 3, PCR products from only a single primer pair were tested by dHPLC, but for fragments 1 and 2 products from more than one primer pair were run on the WAVE at their specific analysis temperatures. A primer pair was selected that generated a WAVE chromatogram consisting of a single, sharp peak using DNA from a cell line with known wild-type (WT) *CEBPA*, for instance HL60 or NB4.

The PCR reaction mix, cycling conditions and WAVE protocols were optimised for each fragment and this information is detailed below as well as summarised in Table 3.2. Examples of the PCR products generated, as viewed on a 2% agarose gel stained with ethidium bromide under UV illumination, are given in Figure 3.3 B. The conditions used were the same for gDNA and cDNA.

3.2.2 Detection of mutations in fragment 1

Fragment 1 contains the N-terminus including the translation initiation sites for p42 and p30, TAD1 and part of TAD2 (Figure 3.3 A). Amplicons of 548 bp were generated from primers CEBPA/1F and CEBPA/1R3 using Phusion Hot Start High-Fidelity DNA Polymerase. The standard Phusion Hot Start High-Fidelity DNA Polymerase reaction mix detailed in chapter 2 (section 2.1.2) was scaled up for a total reaction volume of 25µl and 5% DMSO was included. The standard cycling conditions were used with an annealing temperature of 62°C. Products were checked on a 2% agarose gel stained with ethidium bromide and, where present, were denatured (see section 2.1.4). Three temperatures were selected for WAVE analysis: 66.7°C, 68.1°C and 69.9°C. Figure 3.4 shows the sequence-specific melting profile across fragment 1 at each specified temperature, as generated by the Navigator software. Each WAVE run also contained a PCR product from a cell line with known WT *CEBPA*, such as HL60 or NB4.

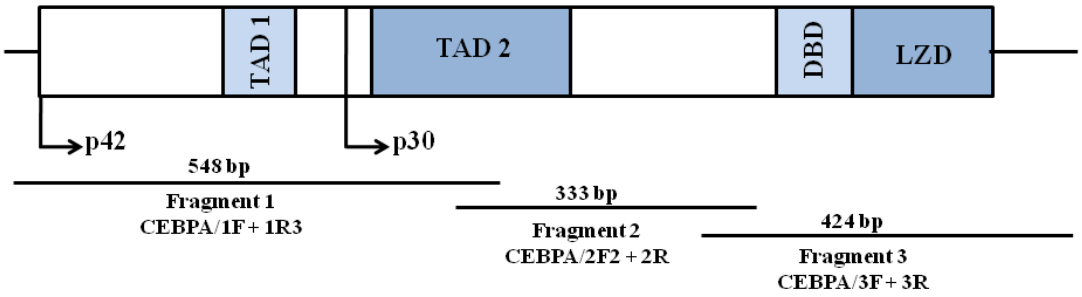
Table 3.2 Summary of optimised *CEBPA* PCR protocols

Amplicon name	Forward primer	Reverse primer	Size of product (bp)	DNA Polymerase	Annealing temperature	Additive in PCR
Fragment 1	CEBPA/1F	CEBPA/1R3	548	Phusion HS HF	62°C	5% DMSO
Fragment 2	CEBPA/2F2	CEBPA/2R	333	BIOTAQ	62°C	5% DMSO
Fragment 3	CEBPA/3F	CEBPA/3R	424	Phusion HS HF	64°C	5% DMSO
G/T nt 690 confirmation	CEBPA/2F2	CEBPA/C693G/ R(MM)	242	BIOTAQ	62°C	5% DMSO

All primer sequences are given in Appendix Table 1.

Abbreviations: bp, base pairs; PCR, polymerase chain reaction; nt, nucleotide; HS, hot start; HF, high-fidelity; DMSO, dimethyl sulfoxide; MM, mismatch.

A



B

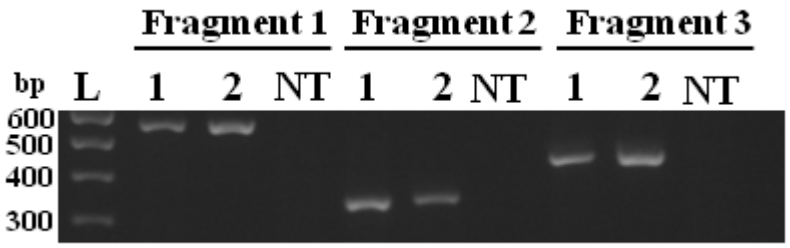


Figure 3.3 PCR fragments used for the detection of *CEBPA* mutations

A. Diagram of the functional domains in CEBPA with the locations of fragments 1, 2 and 3 for WAVE analysis.

B. PCR products from 2 DNA samples (1 and 2) for fragments 1 (548 base pairs [bp]), 2 (333 bp) and 3 (424 bp) and the no template (NT) negative controls, separated by agarose gel electrophoresis. The first lane is a DNA size standard ladder (L) with band sizes as indicated.

3.2.3 Detection of mutations in fragment 2

Fragment 2, 333 bp, was generated from primers CEBPA/2F2 and CEBPA/2R using BIOTAQ DNA Polymerase. Although this is a non-proof-reading enzyme, despite extensive optimisation it was not possible to obtain products from proof-reading polymerases that gave WAVE chromatograms. Nevertheless, heterozygosity for the common G/T SNP at nucleotide (nt) 690 present in fragment 2 was easily detected using this enzyme, giving confidence that lack of a proof-reading activity did not appear to hamper the detection of sequence alterations in this case. The basic BIOTAQ DNA Polymerase reaction mix given in chapter 2 (section 2.1.2) was altered to include 5% DMSO and scaled up to a total volume of 25µl. Cycling conditions were as standard, with an annealing temperature of 62°C. Products were checked on a 2% agarose gel stained with ethidium bromide and, where present, were denatured. WAVE analysis was performed at 65.5°C, 69.2°C and 70.1°C (Figure 3.4). Each WAVE run contained PCR products from a known *CEBPA*-WT case and one heterozygous for the G/T SNP at nt 690 to allow for direct comparison.

3.2.4 Detection of mutations in fragment 3

Fragment 3 covers the DBD and LZD in the C-terminus and amplicons of 424 bp were generated from primers CEBPA/3F and CEBPA/3R using Phusion Hot Start High-Fidelity DNA Polymerase and 5% DMSO in a total volume of 20µl. The standard Phusion Hot Start High-Fidelity DNA Polymerase reaction mix was used, except the addition of 5% DMSO, and standard cycling conditions with an annealing temperature of 64°C. Products were checked on a 2% agarose gel stained with ethidium bromide and, where present, denatured. WAVE analysis was performed at 66.2°C and 68.9°C (Figure 3.4), with each run containing a known WT PCR product.

PCR products from all three fragments were analysed by WAVE for all patients in the cohort and all WAVE chromatograms were scored by two individuals independently. Those resembling the WT control for that fragment were scored as negative. All others were scored as positive and investigated further as detailed in sections 3.2.5 and 3.2.6.

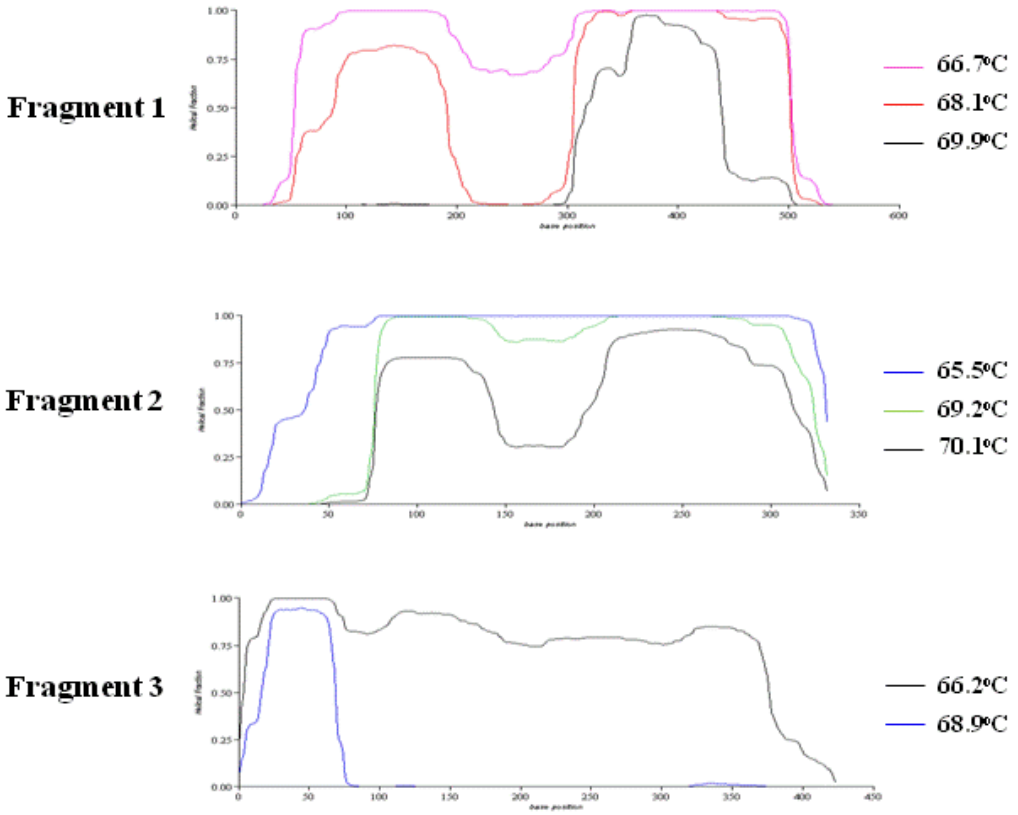


Figure 3.4 Amplicon melting curve profiles. Melting curve profiles for fragments 1, 2 and 3 at the temperatures selected for WAVE analysis, displayed as the helical fraction against the base position in the PCR product.

3.2.5 Confirmation of common polymorphisms

Samples heterozygous for the synonymous G/T SNP at nt 690 were identified by a characteristic WAVE chromatogram of fragment 2 (Figure 3.5 A). The presence of this polymorphism was confirmed by sequencing and/or by PCR and restriction enzyme digestion. For the PCR, a 242 bp amplicon was generated using the screening primer for fragment 2, CEBPA/2F2, and primer CEBPA/C693G/R(MM) (Appendix Table 1), which introduced a mismatched base at nucleotide 693. PCR products were generated using the standard 20µl BIOTAQ DNA Polymerase reaction mix, with the addition of 5% DMSO, and cycling conditions, with an annealing temperature of 62°C. The introduced mismatch meant that digestion with restriction enzyme *Sgr*AI discriminated between alleles in a nucleotide-specific manner, giving bands of 22 and 220 bp for G-alleles and a single band of 242 bp for the T-alleles (Figure 3.5 B). The H195_P196dup polymorphism in fragment 2 was also identified by a characteristic WAVE chromatogram (Figure 3.5 C) and was confirmed by sequencing and/or by capillary electrophoresis as detailed in chapter 5. Cases with these polymorphisms as the sole change were scored as WT.

3.2.6 Confirmation of *CEBPA* mutations

All fragments with an abnormal WAVE chromatogram not scored as one of the two common polymorphisms were re-amplified using the appropriate screening primers, purified using the QIAquick PCR Purification Kit (Qiagen) and sent to UCL Cancer Institute sequencing service for direct nucleotide sequencing.

In the case of size changes that were difficult to determine from the sequence due to the relative level of the mutant compared to WT sequence or due to a large insertion/deletion, a PCR product was generated as before and run on the WAVE at 50°C, a non-denaturing temperature that separates different species in the sample by size, with smaller fragments eluting first. Specific fractions of the eluate were collected, re-amplified and sequenced. Fraction collection was also performed for samples with low level mutations that could not be identified from direct sequencing but that gave clear heteroduplex peaks. In these cases the analysis temperature at which the heteroduplex species was most clearly visible in the WAVE chromatogram was selected for the fraction collection run and heteroduplex peaks were collected, re-amplified and sequenced.

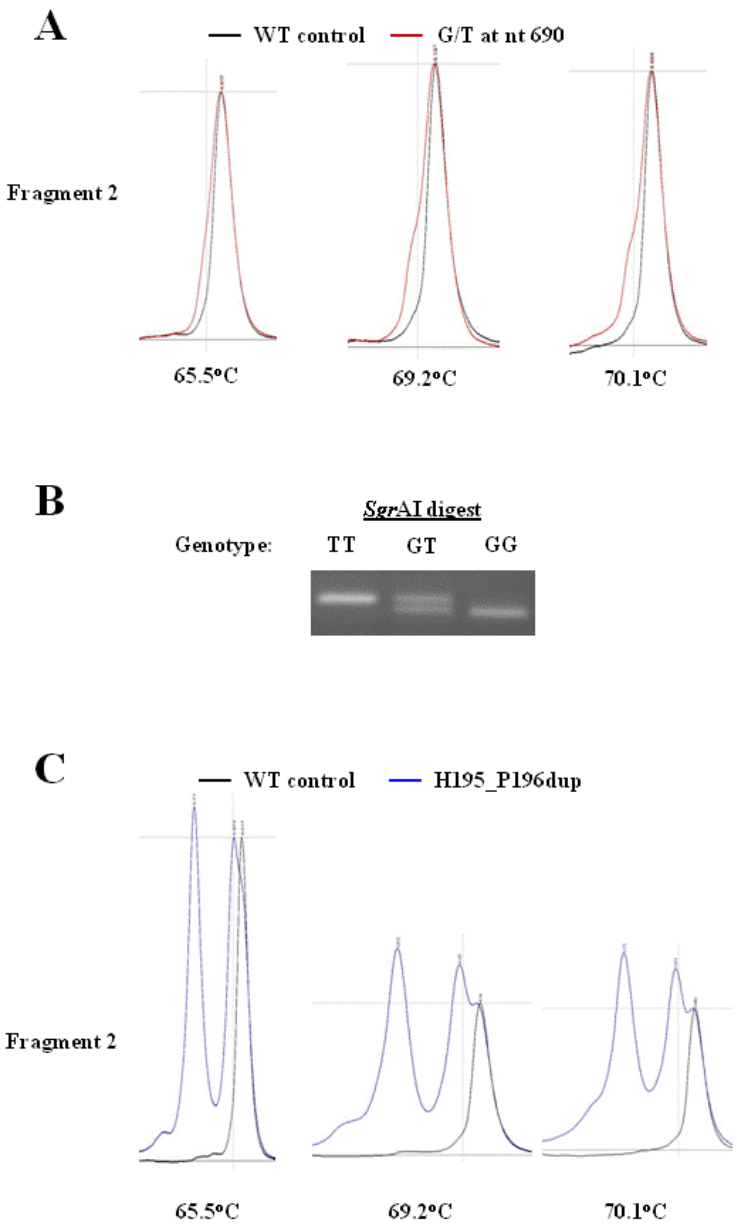


Figure 3.5 Identification and confirmation of common *CEBPA* polymorphisms

A. Characteristic fragment 2 WAVE chromatograms of a case that is heterozygous (G/T) for the synonymous SNP at nucleotide (nt) 690. B. Genotyping of G/T alleles at nt 690 by amplification with a mismatch primer designed to introduce an *Sgr*AI digestion site in G-alleles but not T-alleles. The result of the *Sgr*AI digest is shown as visualised on a 2% agarose gel stained with ethidium bromide. C. Characteristic fragment 2 WAVE chromatograms of a case with the H195_P196dup polymorphism.

3.2.7 Detection of homozygous *CEBPA* mutations

Homozygous *CEBPA* mutations have been described in the literature, however these may be harder to detect using dHPLC as the WAVE platform relies upon the formation of heteroduplexes of WT and mutated PCR products. If there were only a few contaminating normal cells or non-mutated leukaemic cells in the sample, then a homozygous mutation may be missed. In order to detect homozygous mutations, PCR products were generated using the same PCR protocols given above from both patient samples and NB4 cell line DNA, which is WT for *CEBPA* mutations. The products were mixed in equal quantities before denaturing and WAVE analysis at the appropriate temperatures. All fragments with abnormal chromatograms were re-amplified using the appropriate primers and sequenced. This additional screen was only performed for fragments 1 and 3, containing the mutation hotspots, of samples from normal karyotype patients scored as having neither a mutation nor a polymorphism. Patients with a heterozygous mutation or polymorphism were assumed not to have uniparental disomy at the *CEBPA* locus.

3.3 Results

DNA (gDNA = 1382, cDNA = 45) was available from peripheral blood or bone marrow samples taken at diagnosis from 1427 non-APL patients entered into the United Kingdom Medical Research Council AML 10 (n = 510) or AML 12 (n = 917) trials between 1988 and 2002. Ethical approval for the trials and tissue collection for research was obtained from the Multi-Centre Research Committee of Wales and informed consent obtained in accordance with the Declaration of Helsinki. The clinical and molecular characteristics of the cohort are described in detail in chapter 4.

Due to the large number of patients in the study, the cohort was split into two groups for *CEBPA* mutation detection: cases with a normal karyotype (n = 583) and those with an abnormal or unknown karyotype (n = 844). I am very grateful to Kenneth Koo who performed PCRs, WAVE analysis and mutation confirmation for the majority of the latter group.

3.3.1 Detection of *CEBPA* sequence alterations by WAVE analysis

Of the 1427 patients investigated, 469 (33%) had one or more fragments with abnormal WAVE chromatograms. In 382 cases, a characteristic fragment 2 pattern indicated one of the two common polymorphisms (G/T at nt 690 and H195_P196dup), which were confirmed as detailed in section 3.3.2. In a further 12 cases, direct nucleotide sequencing showed either a synonymous coding change (n = 5) or a substitution in the 3′ untranslated region (n = 7). All patients in whom these were the only changes were scored as *CEBPA*-WT.

In the remaining 107 (7%) patients, at least one coding change predicted to be non-synonymous (*CEBPA*-mutant) was identified and, in the case of most mutations (85%), it was possible to determine the change by direct nucleotide sequencing. There were 23 mutations (14%) where it was difficult to identify the nucleotide change from the direct sequencing due to either a low proportion of mutant or to a large size change. The fraction collection facility on the WAVE was used in these cases to either purify the heteroduplex species at a denaturing temperature (n = 13) or to separate differently sized fragments at a non-denaturing temperature (n = 10). The remaining mutations (n = 2) were determined by sequencing cloned PCR products (see chapter 5 for cloning method). Mutations were detected in all three fragments and included point mutations and size changes. Figure 3.6 A and B show representative WAVE chromatograms from typical heterozygous mutations of these types. Additional screening for homozygous mutations by dHPLC of mixed patient and known WT PCR products from fragments 1 and 3 was performed in 373 of the 393 (95%) patients with a normal karyotype lacking either a heterozygous mutation or polymorphism. PCR products for mixing could not be obtained for the remaining 20 patients. Homozygous mutations were detected in three patients that had not been seen in analysis of their unmixed fragments, an example of which is shown in Figure 3.6 C. In one case the WAVE chromatogram indicated a low level mutant (Figure 3.6 D) which could only be detected in the sequencing chromatogram after fraction collection (Figure 3.6 E). In 31 of 34 (91%) cases where a mutation was only detected in either fragment 1 or 3, the other fragment was sequenced, regardless of the WAVE analysis result. No additional mutations were detected. In the remaining three cases either further PCR product could not be obtained for sequencing or the sequencing reaction failed.

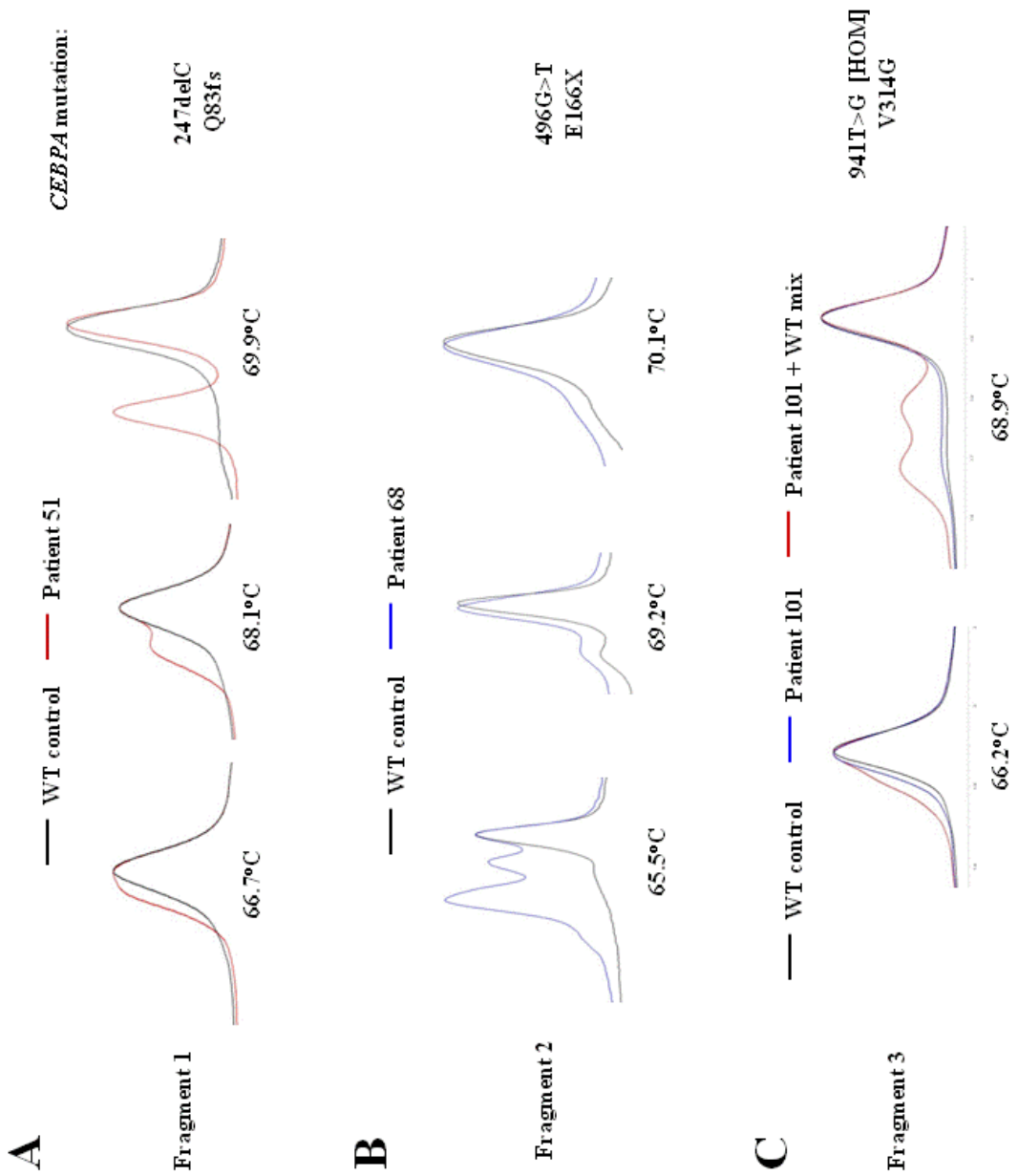


Figure 3.6 Detection of *CEBPA* mutations by WAVE analysis (A to C)

A to C. Representative WAVE chromatograms of patient samples normalised and compared to that of a known wild-type (WT) control. A. A 1 base pair deletion in fragment 1. B. A heterozygous point mutation in fragment 2. C. A homozygous point mutation in fragment 3, only detected when patient PCR product was mixed with WT PCR product prior to WAVE analysis and not detected in un-mixed sample.

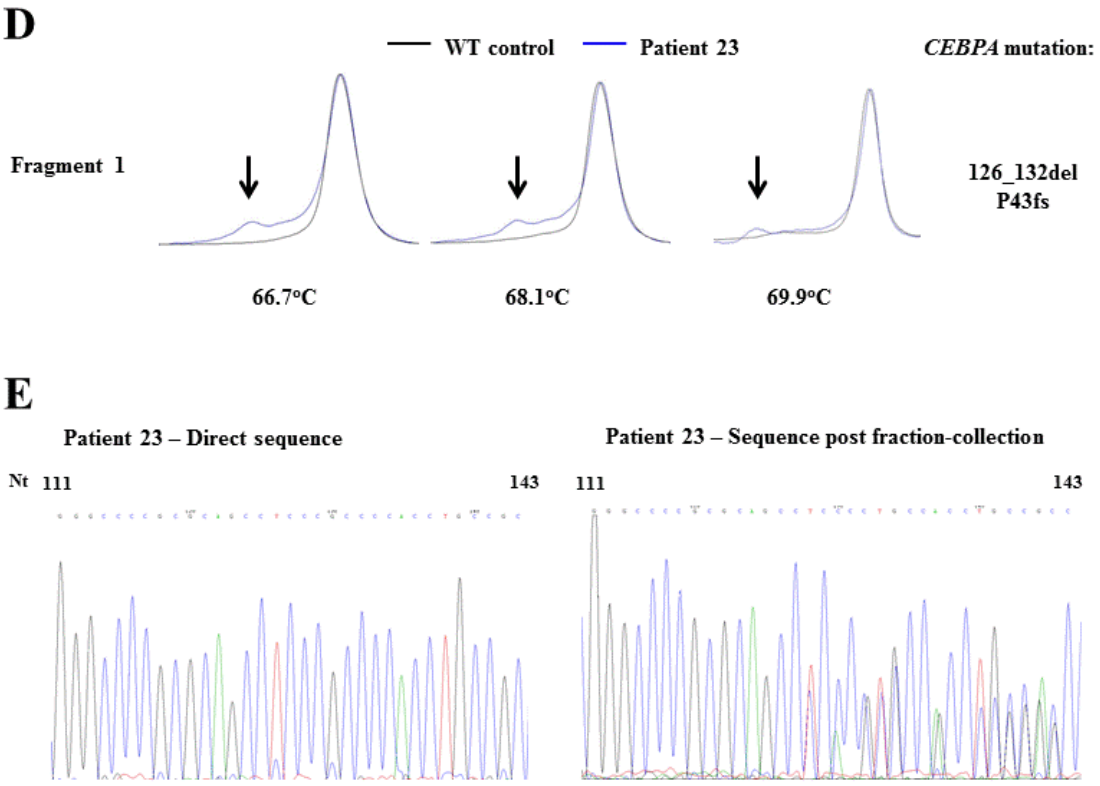


Figure 3.6 continued (D and E)

D. WAVE chromatograms of fragment 1 from patient 23 normalised and compared to that of a known WT control with a low level mutant indicated by the arrows. E. Direct sequencing of the fragment from patient 23 did not show a mutation. The fraction collection facility of the WAVE was therefore used to purify the heteroduplex, which was amplified and re-sequencing showed a 7 base pair deletion. The relevant section of the sequencing chromatograms is shown, with Nt indicating the nucleotides displayed, numbered from the p42 ATG.

3.3.2 Confirmation of common *CEBPA* polymorphisms

There were 316 (22%) cases which were scored as heterozygous for the G/T SNP at nt 690 from the characteristic WAVE chromatogram of fragment 2 (Figure 3.5 A). This was confirmed in 98% of these cases by an alternative method, 13 by direct nucleotide sequencing, 288 by PCR with a mismatch primer and *Sgr*AI digestion (Figure 3.5 B) and 9 cases by both techniques. PCR product either for sequencing or for *Sgr*AI digestion could not be obtained in the 6 patients where confirmation was not performed. The incidence of heterozygotes for this SNP did not differ between *CEBPA*-WT and *CEBPA*-mutant cases (288 of 1320 [22%] vs. 28 of 107 [26%], $P = .30$, Pearson's Chi-square). The H195_P196dup polymorphism was present in 68 (5%) patients, as scored from the WAVE chromatogram (Figure 3.5 C). This was confirmed in 97% of these cases by sequencing (n = 1), fragment analysis of labelled PCR products (n = 57, described in detail in chapter 5) or by both methods (n = 8). PCR product could not be obtained in the two cases where confirmation was not performed. The incidence of H195_P196dup was 64 of 1320 (5%) in *CEBPA*-WT and 4 of 107 (4%) in *CEBPA*-mutant cases, which was not significantly different ($P = .18$, Fisher's exact test). There were two cases who were both G/T heterozygotes at nt 690 and H195_P196dup.

3.3.3 Characteristics of *CEBPA* mutations

Overall, of the 107 *CEBPA*-mutant cases, 48 (45%) were *CEBPA*-single and 59 (55%) were *CEBPA*-double. A complete list of all sequence changes, apart from the two common polymorphisms, is given in Appendix Table 2 along with the classification of the cases as *CEBPA*-single, *CEBPA*-double or SNP. One case (patient 66) had two heterozygous mutations on the same allele, as determined by cloning of PCR products (see chapter 5 for cloning method), and so was scored as *CEBPA*-single. One further case (patient 4) had two mutations each of approximately an intermediate level on different alleles, however the sequencing chromatogram showed WT allele present at about the 50% level. Consequently it was judged more likely that the mutations were in different cells than present as compound heterozygote frame-shifting N-terminal mutations in the same cell, a situation not described in the literature, and this case was classified as *CEBPA*-single. Cases with a homozygous mutation were scored as *CEBPA*-double.

There were notable differences between *CEBPA*-single and *CEBPA*-double cases in terms of the types and locations of mutations detected. The location s of all the mutations are shown in Figure 3.7 and a summary of this information is given in Table 3.3. Mutations in *CEBPA*-

single patients were distributed across the entire gene: 28 patients (58%) had classical N- or C-terminal hotspot mutations, 12 patients (25%) had changes in the middle region (between the p30 start codon and the DBD) that were predicted to cause non-functional proteins through frame-shifting or nonsense mutations and eight patients (17%) had changes of unknown consequence in the middle region, such as in-frame insertions/deletions and missense mutations. By contrast, mutations in *CEBPA*-double patients clustered quite tightly in the N- and C-terminal hotspots. Most *CEBPA*-double cases (83%) had both an out-of-frame insertion in the N-terminus in combination with a DBD or LZD (bZIP) mutation in the C-terminus, and a further four patients (7%) had an N-terminal change with a nonsense mutation or frame-shift alteration in the middle region. Assuming that these mutations were on different alleles, they were all predicted to cause p30 translation from one allele and a non-functional or truncated protein from the other. Of the remaining six *CEBPA*-double patients, two had two different bZIP mutations, one had two changes in the middle region and three had homozygous mutations in the C-terminus.

Samples taken during complete remission (CR) were available from nine *CEBPA*-mutant cases, seven *CEBPA*-double including one homozygous mutant and two *CEBPA*-single patients, as indicated in Appendix Table 2. No *CEBPA* mutations were detected in any of these samples, indicating that they were acquired rather than germline.

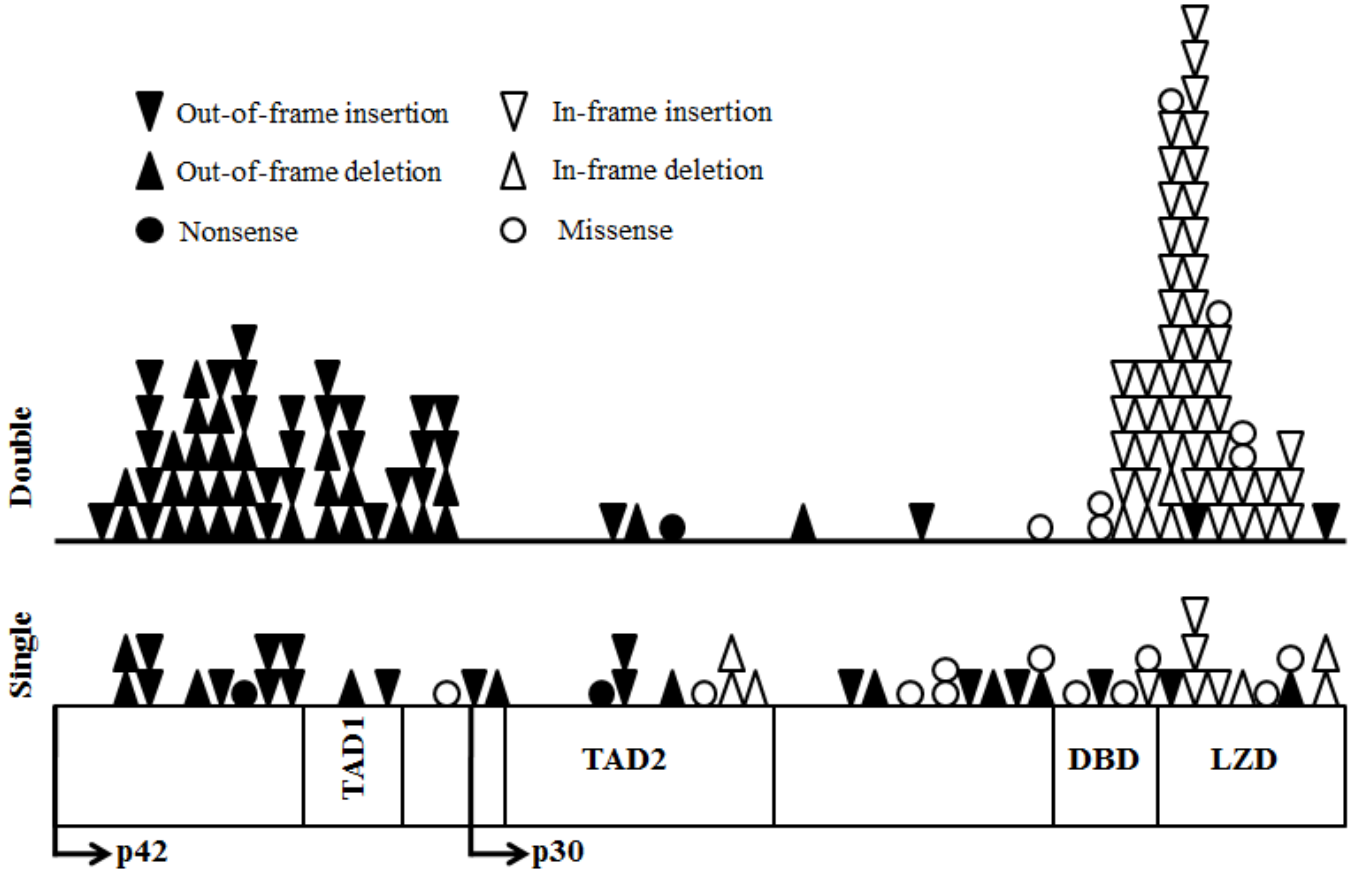


Figure 3.7 Location of all *CEBPA* mutations detected. Location and type of mutations detected in patients that were either *CEBPA*-single or *CEBPA*-double mutant-positive. Translation initiation sites for the p42 and p30 isoforms are indicated as well as the major functional domains: transactivation domains 1 and 2 (TAD1 and TAD2), DNA-binding domain (DBD) and leucine zipper domain (LZD).

Table 3.3 Summary of *CEBPA* mutations detected

	Mutation 1			Mutation 2		
	N-terminus 1st to 2nd start sites AA 1-120	**Middle** 2nd start site to DBD AA 121-277	**C-terminus** DBD and LZD AA 278-358	**Middle** 2nd start site to DBD AA 121-277	**C-terminus** DBD and LZD AA 278-358	**No. of patients**
Single:	Out-of-frame ins/del					11
	Nonsense					1
		Out-of-frame ins/del				11
		Nonsense				1
		In-frame ins/del				3
		Missense				5
			Out-of-frame ins/del			3
			In-frame ins/del			8
			Missense			5
Total						**48**
Double:	Out-of-frame ins/del				In-frame ins/del	46
	Out-of-frame ins/del				Out-of-frame ins/del	1
	Out-of-frame ins/del				Missense	2
	Out-of-frame ins/del			Out-of-frame ins/del		3
	Out-of-frame ins/del			Nonsense		1
		Out-of-frame ins/del		Missense		1
			Out-of-frame ins/del		In-frame ins/del	1
			In-frame ins/del		Missense	1
			Missense-HOM			2
			In-frame ins/del-HOM			1
Total						**59**

Unless indicated otherwise, all mutations were heterozygous. Abbreviations: DBD, DNA-binding domain; LZD, leucine zipper domain; AA, amino acid; ins, insertion; del, deletion; HOM, homozygous mutation.

3.4 Discussion

There are a number of recurrent gene mutations that have been identified in AML and effective methods to detect these markers in patient samples are required. The data presented in this chapter shows the development of a method to screen for *CEBPA* mutations and the application of this method to a large cohort of young adult AML patients. PCR amplification of three overlapping fragments and analysis by dHPLC on the WAVE platform was selected because it enabled sensitive detection of both point mutations and size changes in patient samples with minimal processing of PCR products. The PCR protocols required extensive optimisation due to the GC-rich nature of *CEBPA*, however a method was developed that was practical for a large number of samples.

In total, of the 1427 cases analysed by dHPLC on the WAVE platform, 107 (7%) patients had one or more *CEBPA* mutation(s). This incidence is significantly lower than that reported in 10 other studies combined together (315 of 3184, 10%, $P = .009$ Pearson's Chi-square) (Pabst *et al*, 2001b;Snaddon *et al*, 2003;Frohling *et al*, 2004;Bienz *et al*, 2005;Marcucci *et al*, 2008;Wouters *et al*, 2009;Pabst *et al*, 2009;Hou *et al*, 2009;Renneville *et al*, 2009a;Dufour *et al*, 2010). However, several of these studies only looked at cases with normal karyotypes, where *CEBPA* mutations are more common (see chapter 4). When compared to those studies that included all karyotypes (including APL), our incidence is not different (192 of 2140, 9%, $P = .120$ Pearson's Chi-square) (Pabst *et al*, 2001b;Wouters *et al*, 2009;Pabst *et al*, 2009;Hou *et al*, 2009;Renneville *et al*, 2009a). The results obtained indicated that the selected method was appropriate and as efficient at detecting *CEBPA* mutations as publications using other methods, including sequencing. Further evidence in support of this was that no additional mutations were found when direct sequencing was performed on the other end of the gene in 31 cases scored as *CEBPA*-single on WAVE analysis with a fragment 1 or 3 mutation. Of note, it was found that running each fragment at more than one temperature was required, as indicated by the melting profiles of the amplicons. Although heteroduplex peaks were visible at all analysis temperatures for some sequence changes (Figure 3.5 C), there were cases where the heteroduplex was only visible at one of the temperatures (Figure 3.6 B).

The frequency of homozygous mutations in this cohort (3 of 1427, 0.2%) is equivalent to that reported in other studies (6 of 2641, 0.2%) (Pabst *et al*, 2001b;Snaddon *et al*, 2003;Frohling *et al*, 2004;Bienz *et al*, 2005;Marcucci *et al*, 2008;Wouters *et al*, 2009;Pabst *et al*, 2009;Renneville *et al*, 2009a;Dufour *et al*, 2010). Although the WAVE platform relies

upon heteroduplex formation for mutation detection, past experience in the department has shown that homozygous mutations in other genes are usually still detected by dHPLC of un-mixed PCR products (Mead *et al*, 2007), due to a low proportion of either normal or non-mutated leukaemic cells in the sample. However, this was not the case in this study as the three patients with homozygous *CEBPA* mutations were only detected through mixing of patient and WT PCR products (Figure 3.6 C). The additional screen was only performed in fragments 1 and 3 of normal karyotype cases without a heterozygous mutation or polymorphism. Mutations are most commonly associated with a normal karyotype (chapter 4) and of the 6 cases of homozygous mutations reported in the literature, none have been in fragment 2, therefore it was considered appropriate not to include this region in the additional screen. It is possible that homozygous mutations in other cases may have been missed, but it is likely to be a very limited number as the reported incidence overall is so low (0.2%). Of the three cases with a homozygous *CEBPA* mutation, two had data available from genotyping arrays (Gupta *et al*, 2008). Both of these cases were found to have uniparental disomy on chromosome 19 at the *CEBPA* locus, in line with published data on the mechanism of homozygous mutation acquisition in this gene (Fitzgibbon *et al*, 2005;Wouters *et al*, 2007c).

In addition to detecting *CEBPA* mutations, WAVE analysis also identified the two common polymorphisms (rs34529039 and H195_P196dup) at similar incidences to those previously reported for other European cohorts (NCBI dbSNP) (Wouters *et al*, 2007b). The WAVE chromatogram for both of these polymorphisms was highly characteristic and, where confirmation was performed (376 of 382, 98%), results were concordant. The incidence of either of these polymorphisms did not differ between *CEBPA*-WT and *CEBPA*-mutant cases, although Schnittger *et al* (2011a) found no co-incidence of an H195_P196dup and a *CEBPA* mutation in two separate cohorts of patients totalling 2266 cases and suggested that the H195_P196dup may have a protective effect against the acquisition of *CEBPA* mutations.

Of the 107 *CEBPA*-mutant cases detected in this study, 48 (45%) were *CEBPA*-single and 59 (55%) were *CEBPA*-double, giving a ratio of *CEBPA*-single to *CEBPA*-double cases of 1:1.2. As discussed previously, the reported ratio of *CEBPA*-single to *CEBPA*-double cases has varied widely across 10 different studies at between 1:2.1 and 6.7:1 (Pabst *et al*, 2001b;Snaddon *et al*, 2003;Frohling *et al*, 2004;Bienz *et al*, 2005;Marcucci *et al*, 2008;Wouters *et al*, 2009;Pabst *et al*, 2009;Hou *et al*, 2009;Renneville *et al*, 2009a;Dufour *et al*, 2010). Overall, however, the average proportion of *CEBPA*-mutant cases that were *CEBPA*-single or *CEBPA*-double was 57% and 43%, respectively (1.3:1). This is broadly in

agreement with the data from the current cohort, and indicates that *CEBPA*-mutant cases are split roughly equally between *CEBPA*-single and *CEBPA*-double.

There were striking differences noticed between *CEBPA*-single and *CEBPA*-double cases in terms of the types and locations of mutations detected (Figure 3.7). Of the *CEBPA*-double cases, 83% had both a classical frame-shifting size change in the N-terminus in combination with a classical mutation in the C-terminus (Table 3.3), predicted to cause p30 isoform expression from one allele and interfere with DNA binding or dimerisation of the protein from the other allele. This predominance of the "classic" *CEBPA* mutations is in line with data from other reported cohorts. However, while 58% of the *CEBPA*-single cases had one of either of these classic mutations or a nonsense mutation in the N-terminal, also predicted to cause p30 isoform expression, there were 20 cases with a non-classical mutation in the middle of the gene (Table 3.3). Of these, 12 were either frame-shifting or caused a stop codon, leaving 8 cases with an in-frame size change or missense mutation of unknown consequence. It is of note that a significant minority of *CEBPA*-mutant cases (20 of 107, 19%) had non-classical mutations, and the implications of this will be investigated further in chapter 6.

Germline mutations of *CEBPA* have previously been reported in rare pedigrees of patients with familial AML, although there is some data to suggest that the incidence of this may be more common than first assumed (Pabst *et al*, 2008). In this study, no *CEBPA* mutations were detected in samples taken during CR from patients who were *CEBPA*-mutant at diagnosis. However, it should be noted that remission material was only available from 9 of 107 (8%) *CEBPA*-mutant cases in the cohort.

The data presented in this chapter therefore shows that analysis of PCR products on the WAVE platform is a robust technique for the detection of mutations in the *CEBPA* gene in a large cohort of patient samples. Mutations were found in all fragments and of all different types, including point mutations and homozygous mutations, and at a frequency that is comparable to other studies. The method used in this chapter for further screening of homozygous mutations in a selected cohort only was appropriate for this large retrospective study. However, if samples were being screened prospectively then this should be extended to all cases and fragments to ensure that mutations are not missed. The cohort of patients studied here is large and well-characterised, both in terms of other molecular markers and in the availability of clinical data, including long follow-up. Determining the *CEBPA* mutation status in these patients enabled a thorough study of the characteristics of *CEBPA*-mutant

AML, the impact mutations have on patient outcome and how this is affected by the presence, or lack, of other gene mutations, and these topics are the subject of chapter 4.

CHAPTER 4: PROGNOSTIC IMPACT OF CEBPA MUTATIONS IN YOUNG ADULT PATIENTS

4.1 Introduction

CEBPA mutations are commonly acquired in AML patients, as described in chapter 3, and are known to be associated with certain clinical characteristics and to impact on patient outcome.

4.1.1 Clinical characteristics of AML patients with *CEBPA* mutations

A number of different groups have investigated the characteristics of *CEBPA*-mutant AML through screening for mutations in cohorts of patients, and these are summarised in Table 4.1. There are several features found to be associated with *CEBPA*-mutant AML that are in common between the different studies. For instance, where FAB type has been studied, all groups have reported an association with the M1 and M2 subtypes. In addition, all groups have found mutations to be more frequent in intermediate-risk karyotype patients, the largest proportion of whom have an NK. Indeed, several studies have focussed entirely on *CEBPA* mutations in NK AML, and when these studies are combined the incidence of mutations in this subgroup is 266 of 2127 (13%) (Frohling *et al*, 2004;Bienz *et al*, 2005;Marcucci *et al*, 2008;Dufour *et al*, 2010;Taskesen *et al*, 2011), which is significantly higher than that found in 8 studies where karyotype was not selected (245 of 2528, 10%, $P = .002$ Pearson's Chi-square) (Pabst *et al*, 2001b;Preudhomme *et al*, 2002;Lin *et al*, 2005;Shih *et al*, 2006;Wouters *et al*, 2009;Pabst *et al*, 2009;Hou *et al*, 2009;Renneville *et al*, 2009a). *CEBPA* mutations have also been specifically associated with deletions of chromosome 9q not in a complex karyotype (Frohling *et al*, 2005). Most groups have also reported no difference in the age, sex or presenting white blood cell count (WBC) of *CEBPA*-mutant patients compared to *CEBPA*-WT. This broadly remained the same when *CEBPA*-single and *CEBPA*-double cases were analysed separately, apart from single reports that *CEBPA*-single cases were more likely to be female (Dufour *et al*, 2010) and *CEBPA*-double cases to be of lower age (Taskesen *et al*, 2011) mutated patients.

While these associations have been consistently noted across studies, other characteristics have been more variably reported, in particular the coincidence of *CEBPA* mutations with other molecular markers. When comparing patients with *CEBPA* mutations to those without, several studies reported no difference in the incidence of *FLT3*/ITDs between the two groups

(Preudhomme *et al*, 2002;Lin *et al*, 2005;Shih *et al*, 2006;Renneville *et al*, 2009a), whereas others reported a lower incidence of *FLT3*/ITDs in *CEBPA*-mutant AML (Barjesteh van Waalwijk van Doorn-Khosrovani *et al*, 2003;Frohling *et al*, 2004;Marcucci *et al*, 2008). In addition, a lower incidence of *NPM1* mutations in *CEBPA*-mutant patients compared to *CEBPA*-WT has also been described (Marcucci *et al*, 2008;Pabst *et al*, 2009;Renneville *et al*, 2009a). Interestingly, differences in the coincidence of *CEBPA* mutations with other molecular markers are more striking in studies where *CEBPA*-single and *CEBPA*-double patients have been analysed separately. In comparison to *CEBPA*-singles, *CEBPA*-double cases have consistently fewer *NPM1* mutations and *FLT3*/ITDs (Wouters *et al*, 2009;Hou *et al*, 2009;Dufour *et al*, 2010;Taskesen *et al*, 2011).

4.1.2 Impact of *CEBPA* mutations on response to therapy and patient outcome in AML

Over the last decade, there have been numerous studies on the clinical impact of *CEBPA* mutations in AML and these are summarised in Table 4.2. Initial studies were performed in relatively small cohorts and, at the time at which the work in this thesis was commenced, there were 4 reports, together describing outcome in 72 *CEBPA*-mutant cases of 625 patients studied (12%) (Preudhomme *et al*, 2002;Barjesteh van Waalwijk van Doorn-Khosrovani *et al*, 2003;Frohling *et al*, 2004;Bienz *et al*, 2005). By 2009, 2 further studies of NK-only patients had been published (Schlenk *et al*, 2008;Marcucci *et al*, 2008), the largest of which had screened 509 patients for *CEBPA* mutations. In terms of complete remission (CR) rate, all of these studies found no impact of a *CEBPA* mutation, apart from Schlenk *et al* (2008) who reported an increase in the CR rate in *CEBPA*-mutant cases. However, all reported that *CEBPA*-mutant patients had a significantly better outcome than *CEBPA*-WT, both in terms of overall survival (OS) and of endpoints reflecting relapse rates such as event-free survival (EFS), disease-free survival (DFS) or relapse-free survival (RFS) depending on the study. All groups found that mutated *CEBPA* was an independent favourable factor for outcome in multivariate analyses.

Table 4.1 Studies reported on the characteristics of AML patients with *CEBPA* mutations

Reference	Total no. in study	Median age, years (range)	No. mutants (% of total)	Cytogenetics †	FAB type †	Clinical Characteristics and Immunophenotype †	Co-incidence with other mutations †
Pabst *et al*, 2001b	137	NR	8[1] (6)	Increased: NK	Increased: M1 and M2	NR	NR
Preudhomme *et al*, 2002	135	45 (15-65)	15 (11)	All intermediate risk (*P*=.06)	Increased: M1 (*P*=.02)	No difference: Age, sex and WBC	No difference: *FLT3*/ITD
Barjesteh *et al*, 2003	277	43 (15-60)	12 (4)	All intermediate risk	92% M1 or M2	No difference: Age, sex and WBC	Only 1 of 12 *FLT3*/ITD mutant
Frohling *et al*, 2004	236 NK only	47 (16-60)	33[2] (14)	N/A	72% M1 or M2	Increased: Hb (*P*=.01) % PB blasts (*P*=.008) Decreased: Plt (*P*=.04) No difference: Age, sex, WBC	Decreased: *FLT3*/ITD or D835 or both (*P*=.01)
Lin *et al*, 2005	104	46 (1-85)	16 (15)	Increased: NK Intermediate risk (*P*=.004)	Increased: M1 and M2 (*P* = 0.001)	Increased: CD7 (*P*<.001), CD15 (*P*=.005), CD34 (*P*=.008), HLA-DR (*P*=.034), Hb (*P*=.036) No difference: Age, sex, WBC	No difference: *FLT3*/ITD
Bienz *et al*, 2005	67 NK only	49 (18-71)	12 (18)	N/A	Increased: M1 and M2	Increased: CD7 (*P*=.0023) Decreased: LDH (*P*=.015), WBC (*P*=.18) No difference: Age and sex	NR
Shih *et al*, 2006	149	(0.4-74)	22 (15)	Increased: Intermediate risk (*P*=.002)	Increased: M1 and M2 (*P*=.022)	No difference: Age, sex, Hb, and WBC	No difference: *FLT3*/ITD, *FLT3*/TKD, *NRAS*, *KRAS*
Marcucci *et al*, 2008	175 NK only	(18-59)	32 (18)	N/A	NR	Increased: Hb (*P*=.02) Decreased: Plt (*P*=.009) No difference: Age, sex and WBC	Decreased: *NPM1* (*P*=.0001) *FLT3*/ITD (*P*=.07) No difference: *FLT3*/TKD
Wouters *et al*, 2009 *	598	46.5 (15-77)	41 (7) 13 Single 28 Double	Increased: Intermediate risk	Increased: M1 and M2	No difference: Age, sex, WBC	Double vs. Single Decreased: *NPM1* (*P*=.026), *FLT3*/ITD (*P*=.081), *FLT3*/TKD (*P*=.094)

Table 4.1 Continued

Reference	Total no. in study	Median age, years (range)	No. mutants (% of total)	Cytogenetics	FAB type	Clinical Characteristics and Immunophenotype	Co-incidence with other mutations
Pabst *et al*, 2009 *	224	53 (<61)	19 (8) 7 Single 12 Double	Increased: Intermediate risk	All M1 or M2	Decreased: WBC (*P*=.012), LDH (*P*=.032) No difference: Age, sex	Decreased: *FLT3*/ITD (*P*=.023), *NPM1* (*P*=.008)
Renneville *et al*, 2009a	638	(16-70)	53 (8)	Increased: Intermediate risk (*P*=.05), NK (*P*=.01)	NR	Increased: Males (*P*=.06) No difference: Age, WBC	Decreased: *NPM1* (*P*=.008) No difference: *FLT3*/ITD
Hou *et al*, 2009 *	543	NR	71 (13) 24 Single 47 Double	Single vs. Double No difference	Single vs. Double No difference	Single vs. Double Increased: CD56 (*P*=.038) Decreased: HLA-DR (*P*=.0014), CD7 (*P*=.006), CD15 (*P*<.0001) No difference: Age, sex, WBC, LDH	Single vs. Double Increased: *NPM1* (*P*=.0109), *FLT3*/ITD (21% vs. 11%, N.S.)
Dufour *et al*, 2010 *	467 NK only	61 (17-85)	38 (8) 18 Single 20 Double	N/A	Increased: M1 and M2 (*P*=.009)	Across WT, Single and Double No difference: Age, WBC Singles – Increased females (*P*=.006) Doubles – Increased Hb (*P*=.035)	Double vs. Single Decreased: *NPM1* (*P*=.002), *FLT3*/ITD (*P*=.007)
Taskesen *et al*, 2011 *	1182 NK only	(16-60)	151 (13) 60 Single 91 Double	N/A	NR	No difference: Age, sex, WBC Double vs. WT Decreased: Age (*P*=.04), Plt (*P*<.0001)	Double vs. Single Decreased: *NPM1* (*P*<.0001), *FLT3*/ITD (*P*<.001) Single vs. WT Decreased: *NPM1* (*P*=.018)

† All comparisons are made for *CEBPA*-mutant vs. *CEBPA*-wild-type (WT) cases. [1] Two cases with synonymous point mutations alone have been excluded from the *CEBPA*-mutant cases given in this report. [2] Three cases with H195_P196dup alone have been excluded from the *CEBPA*-mutant cases given in the article. *In these reports, characteristics for *CEBPA*-single and *CEBPA*-double cases were available separately. Characteristics in the table are still given as *CEBPA*-mutant versus *CEBPA*-wild-type (WT) but where differences between *CEBPA*-single (Single) and *CEBPA*-double (Double) cases were reported these are indicated specifically. Abbreviations: NK, normal karyotype; NR, not reported; N/A, not applicable; FAB, French-American-British; Hb, haemoglobin; WBC, white blood cell count; PB, peripheral blood; LDH, lactate dehydrogenase; Plt, platelets; N.S., not significant

The first analysis of *CEBPA*-single and *CEBPA*-double patients separately was published by Wouters *et al* in 2009. They obtained gene expression profiles (GEP) on 524 AML cases with known *CEBPA* mutant status, including 12 *CEBPA*-single and 26 *CEBPA*-double cases, and found that, whilst they could derive a signature for cases with *CEBPA* mutations, it had limited ability to detect all *CEBPA*-mutant cases. However, the misclassification was found to be of *CEBPA*-single cases, with *CEBPA*-double cases predicted accurately. In their outcome analysis, as previously reported, *CEBPA*-mutant cases had an improved clinical outcome over *CEBPA*-WT cases. However, the heterogeneity in the GEP between *CEBPA*-single and *CEBPA*-double patients led them to analyse these groups separately and they found that the benefit in outcome was restricted to *CEBPA*-double patients only, with a *CEBPA*-single mutant having no impact. Whilst this heterogeneity in outcome between *CEBPA*-single and *CEBPA*-double cases was replicated in another smaller cohort of 224 patients (Pabst *et al*, 2009), a larger study of 638 patients enrolled in the Acute Leukaemia French Association trials (ALFA) (Renneville *et al*, 2009a) found little difference between 29 *CEBPA*-single and 24 *CEBPA*-double patients in terms of outcome. No significant differences were found for CR rate, RFS or DFS and only a trend for better OS in *CEBPA*-double patients (P=.09), with no significance found in multivariate analysis. Therefore, numerous questions remained regarding the impact of *CEBPA* mutations in AML and whether one or two mutations were necessary to see a clinical benefit.

4.1.3 Factors modifying the impact of *CEBPA* mutations in AML

There are a number of other factors that are known to affect patient outcome in AML (chapter 1), which need to be considered when investigating the impact of *CEBPA* mutations. The effect of karyotype in *CEBPA*-mutant AML has not been well studied, in part because a large number of groups have focussed solely on NK patients, and also because mutations tend to occur only in either NK patients or those with an AK that is neither adverse nor favourable risk. One group has reported that the benefit of a *CEBPA* mutation is limited to those with an NK and that *CEBPA*-mutant patients with an AK should be considered as intermediate rather than favourable risk (Renneville *et al*, 2009a), however, this has not been confirmed in other cohorts.

Table 4.2 Studies reported on the impact of *CEBPA* mutations on clinical outcome in AML

	Cohort			CR rate		Outcome compared to *CEBPA*-WT			
Reference	**Total no. in study**	**Median age, years (range)**	***CEBPA* Mutant n (%)**	**Mutant vs. WT**	**Single vs. Double**	***CEBPA* mutant**	***CEBPA* single**	***CEBPA* double**	**Overall impact on outcome**
Preudhomme *et al*, 2002	135	45 (15-65)	15 (11) 7 Single 8 Double	No difference	NR	Increased: OS (*P*=.04), EFS (*P*=.04), DFS (*P*=.05) Multivariate: Increased OS (*P*=.05), DFS (*P*=.05)	NR	NR	Mutant favourable
Barjesteh *et al*, 2003	[1]187 IK only	43 (15-60)	12 (6) 12 Double	No difference	NR	Increased: OS (*P*=.03), EFS (*P*=.02) Multivariate: Increased OS (*P*=.04), EFS (*P*=.03)	NR	NR	Mutant favourable
Frohling *et al*, 2004	236 NK only	47 (16-60)	[2]33 (14) 18 Single 15 Double	No difference	NR	Increased: Remission duration (*P*=.01), OS (*P*=.05) Multivariate for WT: Decreased remission duration (*P*=.01), OS (*P*=.04)	NR	NR	Mutant favourable
Bienz *et al*, 2005	67 NK only	49 (18-71)	12 (18) 4 Single 8 Double	No difference	NR	Increased: OS (*P*=.0007), DFS (*P*=.0017) Multivariate: Increased OS (*P*=.0005), DFS (*P*=.0012)	NR	NR	Mutant favourable
Schlenk *et al*, 2008	509 NK only	[3]48 (16-60)	67 (13) Single/ Double NR	Increased	NR	Increased: OS (*P*<.001[4]), RFS (*P*<.001[4]) Multivariate: Increased OS, RFS	NR	NR	Mutant favourable

Table 4.2 Continued

	Cohort			CR rate		Outcome compared to *CEBPA*-WT			
Reference	**Total no. in study**	**Median age, years (range)**	***CEBPA* Mutant n (%)**	**Mutant vs. WT**	**Single vs. Double**	***CEBPA* mutant**	***CEBPA* single**	***CEBPA* double**	**Overall impact on outcome**
Marcucci *et al*, 2008	175 NK only	NR (18-59)	32 (18) 18 Single 14 Double	No difference	NR	Increased: EFS (*P*=.017), DFS (*P*=.075). No SD in OS (*P*=.10) Multivariate: Increased EFS (*P*=.007), DFS (*P*=.014), OS (*P*<.001)	NR	NR	Mutant favourable
Wouters *et al*, 2009	598	46.5 (15-77)	41 (7) 13 Single 28 Double	NR	NR	Increased: OS (*P*=.027), EFS	Multivariate: No difference for OS (*P*=.65), EFS (*P*=.16)	Increased: OS (*P*=.004), EFS (*P*=.005) Multivariate: Increased OS (*P*<.001), EFS (*P*<.001)	Double favourable
Pabst *et al*, 2009	224	53 (<61)	19 (8) 7 Single 12 Double	Increased	No difference	NR	No difference: OS, DFS Multivariate: No difference for OS (*P*=.52), DFS (*P*=.28)	Increased (vs. Single): OS (*P*=.006), DFS (*P*.013) Multivariate (vs. WT): Increased OS (*P*<.001), DFS (*P*<.001)	Double favourable
Renneville *et al*, 2009a	638	NR (16-70)	53 (8) 29 Single 24 Double	No difference	No difference	Increased: RFS (*P*=.04), DFS (*P*=.03). No SD for OS (*P*=.11)	Single vs. Double: Decreased OS (*P*=.09) No difference: RFS, DFS		Mutant favourable in NK without *FLT3*/ITD

Table 4.2 Continued

	Cohort			CR rate		Outcome compared to *CEBPA*-WT			
Reference	**Total no. in study**	**Median age, years (range)**	***CEBPA* Mutant n (%)**	**Mutant vs. WT**	**Single vs. Double**	***CEBPA* mutant**	***CEBPA* single**	***CEBPA* double**	**Overall impact on outcome**
Hou *et al*, 2009	[5]397	NR	60 (15) 16 Single 44 Double	NR	Decreased (P = .0051)	NR	Multivariate: No difference for OS (P=.227), DFS (P=.629)	Increased: OS (P=.013), DFS (P=.016) Multivariate: Increased OS (P=.004), DFS (P=.001)	Double favourable
Dufour *et al*, 2010	467 NK only	61 (17-85)	38 (8) 18 Single 20 Double	Single vs. WT: No difference Double vs. WT: Increased (P=.071)		Increased: OS (P=.028)	No difference: OS (P=.506), EFS (P=.771), RFS (P=.685) Multivariate: No difference for OS (P=.433), EFS (P=.736)	Increased: OS (P=.018), EFS (P=.064). No difference: RFS (P=.209) Multivariate: Increased OS (P=.003), EFS (P=.008)	Double favourable
Taskesen *et al*, 2011	1182 NK only	NR (16-60)	151 (13) 60 Single 91 Double	Single vs. WT: No difference Double vs. WT: Increased (P = .002)		NR	Increased: OS (P=.05), RFS (P=.02) No difference: EFS (P=.22) Multivariate: No difference for OS (P=.1), EFS (P=.4), RFS (P=.3)	Increased: OS (P<.0001), EFS (P<.0001), DFS (P=.05) Multivariate: Increased OS (P<.0001), EFS (P<.0001), RFS (P.001)	Double favourable

P values ≤ 0.05 were considered significant. Abbreviations: Single, *CEBPA*-single; Double, *CEBPA*-double; NR, not reported; NK, normal karyotype; IK, intermediate-risk karyotype; CR, complete remission; WT, wild-type; OS, overall survival; DFS, disease-free survival; RFS, relapse-free survival; EFS, event-free survival; *FLT3*/ITD, internal tandem duplications of the *FLT3* gene; SD, significant difference. [1] Total cohort studied was 277 cases, outcome data only given for cases with intermediate-risk karyotypes. [2] Three cases with H195_P196dup polymorphism alone have been excluded from the *CEBPA*-mutant cases given in the article. They are still counted as *CEBPA*-mutant for outcome analyses, however *CEBPA*-mutant remained favourable for outcome in a later analysis with these cases re-classified as *CEBPA*-WT (Frohling *et al*, 2007). [3] Age reported for entire cohort (n=872), rather than the subset who had *CEBPA* mutation analysis. [4] *P*-value across *CEBPA*-mutant, *NPM1*-mutant without *FLT3*/ITD and other genotypes. [5] Total cohort studied was 543 patients (24 *CEBPA*-single, 47 *CEBPA*-double). Only patients receiving standard chemotherapy were enrolled into survival analysis.

A small number of groups have looked at the effect of co-incident *FLT3*/ITD mutations, with some conflicting data. Both studies on patients in the ALFA trials found that the benefit for a *CEBPA* mutation was lost in the presence of a *FLT3*/ITD (Preudhomme *et al*, 2002;Renneville *et al*, 2009a), however a report from the German-Austrian AML Study Group saw no impact of a *FLT3*/ITD in *CEBPA*-mutant AML (Frohling *et al*, 2004). Analysis from the Cancer and Leukemia Group B in the group classified as molecular high-risk (those with a *FLT3*/ITD and/or *NPM1*-WT) determined that a *CEBPA* mutation was still favourable in these patients. Even after adjustment for *FLT3*/ITD status, a *CEBPA* mutation still independently predicted longer DFS (P = .008) and longer OS (P = .009) in this study (Marcucci *et al*, 2008). It has also been suggested that a *FLT3*/TKD mutation could further improve the benefit in outcome seen in *CEBPA*-mutant patients (Bacher *et al*, 2008). However, this study compared a very small number of four *FLT3*/TKD-mutant/*CEBPA*-mutant patients to 41 *FLT3*/TKD-WT/*CEBPA*-mutant patients and, whilst better estimates for both OS and EFS were described for the prior group, no significant differences were reported. Clarification of the role of *FLT3* mutations in *CEBPA*-mutated AML is therefore still required and questions on the impact of *NPM1* mutations in this context are yet to be explored.

This chapter reports the results of an investigation into the impact of *CEBPA* mutations in a large cohort of non-APL young adult AML patients, in terms of patient characteristics, response to therapy and long-term outcome. The interaction with other factors such as karyotype and *FLT3*/ITD or *NPM1* mutations are also explored.

4.2 Patients, Materials and Methods

4.2.1 Patients

As detailed in chapter 3, diagnostic samples from 1427 patients with known *FLT3*/ITD and *NPM1* status were available for *CEBPA* mutation screening and all had been entered into the United Kingdom MRC AML 10 (n = 510) or AML 12 (n = 917) trials between 1988 and 2002. The clinical information was made available for this study by the chief investigator of the trials, Professor Alan K. Burnett, and all statistical analysis was performed by the trials' statistician, Dr. Robert K. Hills. The details of the clinical and molecular characteristics of the cohort are given in Table 4.3. The cohort was mostly young adult cases (median age at trial entry was 43 years, range 15-68 years, only 35 cases were aged 60 years or more) and 92% had *de novo* AML. Patients with APL were excluded from this study due to the distinct

biological basis and different patient treatment of this disease, as discussed in chapter 1. When compared to the 2341 non-APL adults in the trials who were not included in the study cohort, there were no significant differences in age, sex, presence of secondary disease or cytogenetics. Patients in the cohort analysed for *CEBPA* mutations did have higher presenting WBCs ($P < .0001$), which probably reflects a bias in samples available for tissue banking. Although patients included in the cohort had a marginally worse performance status ($P = .04$), OS was not significantly different.

4.2.2 Therapy in AML 10 and AML 12

The MRC AML 10 trial (Figure 4.1) (Hann *et al*, 1997) recruited patients between 1988 and 1995, and was primarily for patients less than 56 years of age, with older patients entered if considered suitable for the intensive therapy. The initial randomisation in this trial was between two different induction regimens. One entailed 2 courses of DAT chemotherapy (daunorubicin [50 mg/m^2 by slow intravenous push on days 1, 3 and 5]; cytarabine [100 mg/m^2 12-hourly by IV push on days 1 to 10 for the first course and 1 to 8 for the second course] and 6-thioguanine [100 mg/m^2 12-hourly, orally days 1 to 10 for the first course and 1 to 8 for the second course]). The other induction schedule consisted of ADE chemotherapy (daunorubicin and cytarabine as before and etoposide [100 mg/m^2 by 1 hour IV infusion days 1 to 5]). After induction, patients who entered CR received 2 courses of consolidation therapy: MACE (amsacrine [100 mg/m^2 by 1 hour IV infusion days 1 to 5]; cytarabine [200 mg/m^2 by continuous IV infusion days 1 to 5]; etoposide [100 mg/m^2 by 1 hour IV infusion days 1 to 5]) then MiDAC (Mitoxantrone [10 mg/m^2 by short IV infusion days 1 to 5]; cytarabine [1.0 g/m^2 by short IV infusion days 1 to 5]). A second randomisation after completion of 4 courses of chemotherapy compared high-dose therapy with autologous bone marrow transplant (BMT) rescue versus no further therapy. Patients who had an HLA-matched sibling donor were scheduled for allogeneic BMT after 4 courses of chemotherapy and were not randomised to receive autologous-BMT or not.

The MRC AML 12 trial (Figure 4.1) (Burnett *et al*, 2010), which recruited patients between 1994 and 2002, also randomised patients between different induction schedules. Initially, patients either received 2 cycles of ADE chemotherapy as used in AML 10 or 2 cycles of MAE (mitoxantrone [12 mg/m^2 by IV infusion on days 1, 3 and 5]; cytarabine [100 mg/m^2 12-hourly by IV push on days 1 to 10 for the first course and 1 to 8 for the second course]; and etoposide [100 mg/m^2 by 1 hour IV infusion days 1 to 5]). Of these patients, 29% were also randomly assigned to receive G-CSF or a placebo in course 1. From November 1998, a protocol amendment allowed randomisation between 2 courses of DAT induction

chemotherapy which included either standard-dose cytarabine (S-DAT, as for AML 10 DAT schedule) or high-dose cytarabine (H-DAT, as for S-DAT except a cytarabine dose of 200 mg/m^2, rather than 100 mg/m^2). In this induction arm, 89% were also randomly assigned to receive all-*trans*-retinoic acid or not. In terms of consolidation therapy, patients entering CR received MACE (as AML 10 schedule) and then were randomised to receive either one or two further courses (four or five in total). In either case, those without favourable cytogenetics were randomly allocated between BMT (allogeneic where a matched sibling donor was available and autologous otherwise) and chemotherapy as a final course. Patients with favourable cytogenetics were not eligible for randomisation to receive a BMT. If only one further chemotherapy course was assigned, then MiDAC (as AML 10 schedule) was received. If two further chemotherapy courses were assigned then patients received ICE (idarubicin [10 mg/m^2 on days 1 to 3]; cytarabine [100 mg/m^2 by IV push days 1 to 5 every 12 hours]; etoposide [100 mg/m^2 by 1 hour IV infusion days 1 to 5]) followed by MiDAC. Where BMT had been assigned, patients proceeded to transplant after recovery from MACE for those allocated BMT as course 4 and after ICE for those allocated BMT as course 5. Poor risk patients were treated on an MRC relapsed/refractory protocol. Assigned chemotherapy and allocation of transplantation in first remission for the patients included in this study are given in Table 4.3.

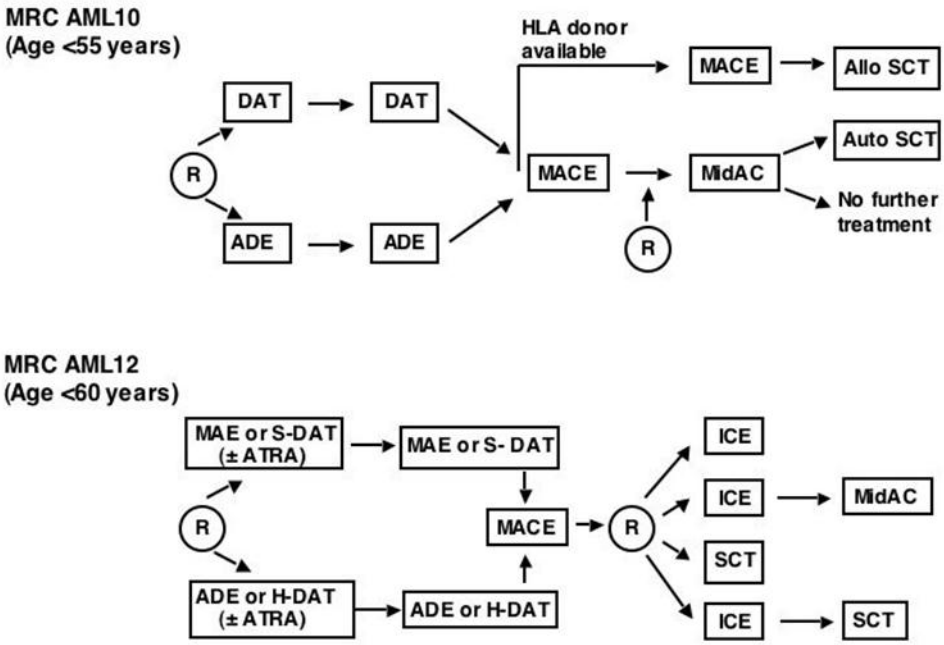


Figure 4.1 Outline of the treatment protocols for the United Kingdom MRC AML 10 and AML 12 trials. For details of the trials and chemotherapy schedules see section 4.2.2. Abbreviations: R, randomisation event; DAT, daunorubicin, cytarabine and 6-thioguanine; ADE, cytarabine, daunorubicin and etoposide; MACE, amascrine, cytarabine and etoposide; MidAC, mitoxantrone and cytarabine; HLA donor, human leukocyte antigen matched donor; SCT, stem-cell transplantation; MAE, mitoxantrone, cytarabine and etoposide; S-DAT, cytarabine at standard dose, daunorubicin and 6-thioguanine; ATRA, *all-trans*-retinoic acid; H-DAT, cytarabine at twice standard dose, daunorubicin and 6-thioguanine; ICE, idarubicin, cytarabine and etoposide. Adapted from Virappane *et al* (2008).

4.2.3 Clinical end points

CR was defined as a normocellular bone marrow containing less than 5% blasts and showing evidence of normal maturation of other marrow elements. Persistence of myelodysplastic features did not preclude the diagnosis of CR. Of the cases defined as CR, 91% achieved a neutrophil count of $1x10^9$/L and a platelet count of $100x10^9$/L. Remission failures were classified by the clinicians as either due to induction death (ID; death related to treatment and/or hypoplasia within 30 days of trial entry) or resistant disease (RD; failure to eliminate disease, including partial remission with 5% to 15% blasts in the bone marrow). Where the clinician's evaluation was not available, deaths within 30 days of entry were classified as ID and deaths later than 30 days after entry as RD. OS was the time from randomisation to death. For patients achieving CR, RFS was the time from the date of first CR to an event (death in first CR or relapse) and cumulative incidence of relapse (CIR) was the incidence of relapse after CR with death in CR as a competing risk; the cumulative incidence of death in CR (CIDCR) was the incidence of death without relapse, with relapse as a competing risk.

4.2.4 Statistical methods

The Mantel-Haenszel test for trend (for ordinal data) and chi-squared tests were used to test for differences in clinical and demographic data by *CEBPA* status. Kaplan-Meier life tables were constructed for survival data and compared by means of the log-rank test. Multivariate Cox models (Cheson *et al*, 2003) were used to analyse CIR, RFS and OS. Models were fitted with variables added successively to the model, in order of significance, if adding the variable provided a significant improvement in fit (in terms of deviance) at $P < .05$. Initially, all main effects were added and the model checked to determine whether any effects added ceased to be significant. Interactions between those covariates entering the model were then explored. Tests for heterogeneity were used to compare subgroups (Early Breast Cancer Trialists' Collaborative Group, 1990). Odds ratios (ORs), hazard ratios (HRs) and 95% confidence intervals (CI) are quoted for all main endpoints. In all cases, a ratio <1 indicates benefit for *CEBPA* mutation(s). All *P*-values are two-tailed.

4.3 Results

Surviving patients in the cohort of 1427 analysed for *CEBPA* mutations were censored on 26th October 2008, with follow-up complete for 97% of patients. Median follow-up was 11.7 years (range 3.0 to 20.4 years). In total, 107 (7%) were *CEBPA*-mutant, comprising 48 (45%) classified as *CEBPA*-single and 59 (55%) as *CEBPA*-double. Details of the mutations detected are given in chapter 3 and Appendix Table 2.

4.3.1 Patient characteristics according to *CEBPA* genotype

Details of patient characteristics stratified by *CEBPA*-WT, *CEBPA*-single and *CEBPA*-double status are given in Table 4.3. There was no difference between these groups in terms of sex, presenting WBC or type of leukaemia (*de novo* or secondary). However, there was an association of younger age with a greater number of mutations ($P = .02$), in particular for *CEBPA*-double patients compared to *CEBPA*-WT and *CEBPA*-single cases together ($P = .003$). Consistent with the literature, *CEBPA* mutations were associated with FAB types M1 and M2, with 78 (73%) *CEBPA*-mutant patients classified as M1 or M2 compared with 600 (45%) *CEBPA*-WT ($P < .001$, Pearson's Chi square). Cytogenetic risk groups were allocated according to the MRC classification (Grimwade *et al*, 1998) and, in line with published data there was an association of *CEBPA* mutations with intermediate-risk karyotypes. No *CEBPA* mutations were found in any of the 167 patients with favourable-risk cytogenetics and, of the 141 cases with adverse-risk karyotypes, only four *CEBPA*-single and no *CEBPA*-double mutations were detected. The greatest number of *CEBPA*-mutant patients had an NK (60 of the 84 cases where karyotype was available, 71%), however the incidence of *CEBPA* mutations was similar between NK and intermediate-risk AK patients: 4% and 6% of NK patients were *CEBPA*-single and *CEBPA*-double respectively compared with 3% and 5% of those with an intermediate-risk AK. In the intermediate-risk AK patients, there were no significant differences in the coincidence of particular *CEBPA*-mutant genotypes and abnormalities such as +8, +21 or del(13q). However, there was an association between del(9q) and *CEBPA*-double mutations: 3 of 13 (23%) intermediate-risk AK *CEBPA*-double patients had a del(9q) abnormality compared with 12 of 260 (5%) intermediate-risk AK *CEBPA*-WT or *CEBPA*-single patients ($P = .03$, Fisher's exact test), an association which has been described previously (Frohling *et al*, 2005).

Table 4.3 Characteristics of patients studied according to *CEBPA* genotype

Characteristic	Total No.	*CEBPA*-WT No. (%)	*CEBPA*-single No. (%)	*CEBPA*-double No. (%)	*P* across *CEBPA*-WT/ *CEBPA*-single/ *CEBPA*-double	*P* *CEBPA*-mutant vs. *CEBPA*-WT	*P* *CEBPA*-double vs. *CEBPA*-WT + *CEBPA*-single
No. of patients	1427	1320 (93)	48 (3)	59 (4)			
Trial:							
AML 10	510	467 (92)	21 (44)	22 (37)	0.5*	0.3†	0.8†
AML 12	917	853 (93)	27 (56)	37 (63)			
Chemotherapy:							
ADE	501	460 (35)	24 (50)	17 (29)	0.2†	0.4†	0.5†
MAE	243	231 (18)	4 (8)	8 (14)			
S-DAT	449	412 (31)	14 (29)	23 (39)			
H-DAT	234	217 (16)	6 (13)	11 (19)			
Age, years:							
15-29	281	252 (19)	8 (17)	21 (36)	0.02*	0.13*	0.003*
30-39	292	272 (21)	8 (17)	12 (20)			
40-49	418	388 (29)	15 (31)	15 (25)			
50-59	401	377 (29)	15 (31)	9 (15)			
60 or more	35	31 (2)	2 (4)	2 (3)			
Median	43	43	46.5	35			
range	15-68	15-68	16-62	16-67			
Sex:							
Female	728	675 (51)	23 (48)	30 (51)	0.8*	0.7†	1.0†
Male	699	645 (49)	25 (52)	29 (49)			

Table 4.3 Continued

Characteristic	Total No.	*CEBPA*-WT No. (%)	*CEBPA*-single No. (%)	*CEBPA*-double No. (%)	*P* across *CEBPA*-WT/ *CEBPA*-single/ *CEBPA*-double	*P* *CEBPA*-mutant vs. *CEBPA*-WT	*P* *CEBPA*-double vs. *CEBPA*-WT + *CEBPA*-single
Diagnosis:							
de Novo	1317	1215 (92)	44 (92)	58 (98)	0.12*	0.2†	0.08†
Secondary	110	105 (8)	4 (8)	1 (2)			
WBC, X 10^9/L:							
<10	434	412 (32)	11 (23)	11 (19)	0.19*	0.13*	0.4*
10-19.9	222	200 (15)	9 (19)	13 (22)			
21-49.9	310	286 (22)	9 (19)	15 (26)			
50-99.9	215	195 (15)	8 (17)	12 (21)			
100+	216	199 (15)	10 (21)	7 (12)			
Missing	30	28	1	1			
Median range	22.9 0.4-559	22.2 0.4-559	33.0 0.7-294	19 3.6-480			
FAB Type:							
M0	52	49 (4)	2 (4)	1 (2)	0.01†	0.0002†	0.004†
M1	268	233 (19)	15 (33)	20 (34)			
M2	410	367 (30)	16 (36)	27 (47)			
M4	353	339 (27)	7 (16)	7 (12)			
M5	171	166 (13)	3 (7)	2 (3)			
M6	42	41 (3)	1 (2)	0			
M7	22	22 (2)	0	0			
Bilineage	1	1 (<0.5)	0	0			
RAEB-t	21	19 (2)	1 (2)	1 (2)			
Other/Unknown	87	83	3	1			

Table 4.3 Continued

Characteristic	Total No.	*CEBPA*-WT No. (%)	*CEBPA*-single No. (%)	*CEBPA*-double No. (%)	*P* across *CEBPA*-WT/ *CEBPA*-single/ *CEBPA*-double	*P* *CEBPA*-mutant vs. *CEBPA*-WT	*P* *CEBPA*-double vs. *CEBPA*-WT + *CEBPA*-single
[1]**Cytogenetics:**							
Favourable	167	167 (15)	0	0	<0.0001†	<0.0001†	0.0004†
Normal Karyotype	583	523 (48)	26 (70)	34 (72)			
Other Intermediate	273	253 (23)	7 (19)	13 (28)			
Adverse	141	137 (13)	4 (11)	0			
Unknown	263	240	11	12			
***FLT3*/ITD:**							
Wild-type	1060	974 (74)	34 (71)	52 (88)	0.04*	0.13†	0.01†
Mutant	367	346 (26)	14 (29)	7 (12)			
***NPM1*:**							
Wild-type	889	804 (61)	28 (58)	57 (97)	<0.0001*	0.0001†	<0.0001†
Mutant	538	516 (39)	20 (42)	2 (3)			
ITD/*NPM1*:							
Wild-type/Wild-type	747	675 (51)	21 (44)	51 (86)			
Wild-type/Mutant	313	299 (23)	13 (27)	1 (2)			
Mutant/Wild-type	142	129 (10)	7 (15)	6 (10)			
Mutant/Mutant	225	217 (16)	7 (15)	1 (2)			

Table 4.3 Continued

Characteristic	Total No.	*CEBPA*-WT No. (%)	*CEBPA*-single No. (%)	*CEBPA*-double No. (%)	*P* across *CEBPA*-WT/ *CEBPA*-single/ *CEBPA*-double	*P* *CEBPA*-mutant vs. *CEBPA*-WT	*P* *CEBPA*-double vs. *CEBPA*-WT + *CEBPA*-single
SCT in first remission:							
None	1110	1026 (78)	38 (79)	46 (78)	0.9†	0.5†	0.6†
Sibling allograft	154	144 (11)	5 (10)	5 (8)			
Autograft	136	123 (9)	5 (10)	8 (14)			
MUD/UD	25	25 (2)	0	0			
Other/unknown	2	2 (<1)	0	0			

* Test for trend, † Test for heterogeneity. [1] Cytogenetic risk groups as per Grimwade *et al* (1998). Abbreviations: ADE, cytarabine, daunorubicin and etoposide; MAE, cytarabine, mitoxantrone and etoposide; S-DAT, cytarabine at standard dose, daunorubicin and thioguanine; H-DAT, cytarabine at twice standard dose, daunorubicin and thioguanine; RAEB-t, refractory anaemia with excess blasts in transformation; SCT, stem-cell transplantation; MUD, matched unrelated donor; UD, unrelated donor. For details of chemotherapy schedules see text.

In terms of the co-incidence with other gene mutations, when comparing *CEBPA*-mutant to *CEBPA*-WT patients there was no difference in the incidence of *FLT3*/ITDs (20% vs. 26%; *P* = .13), but *CEBPA*-mutant patients were less likely to have an *NPM1* mutation (21% vs. 39%; *P* = .0001). However, if *CEBPA*-single and *CEBPA*-double patients were analysed separately, it was clear that, while the incidence of a *FLT3*/ITD was similar in *CEBPA*-WT (26%) and *CEBPA*-single (29%) patients, there was a lower incidence in *CEBPA*-double patients (12%; *P* = .04). In addition, *CEBPA*-double cases were highly unlikely to have an *NPM1* mutation compared with *CEBPA*-WT and *CEBPA*-single patients (3%, 39% and 42% respectively; *P* < .0001).

4.3.2 Response to therapy and clinical outcome by *CEBPA* genotype

There were no differences in the therapy received, either chemotherapy or transplantation, for *CEBPA*-WT, *CEBPA*-single or *CEBPA*-double patients (Table 4.3). For response to therapy measured by CR rate, RD or ID, there were no statistically significant differences when comparing *CEBPA*-mutant to *CEBPA*-WT cases, or if *CEBPA*-single and *CEBPA*-double patients were analysed separately (Table 4.4). However, the highest CR rate was in the *CEBPA*-double patients (83%, 77% and 90% for *CEBPA*-WT, *CEBPA*-single and *CEBPA*-double respectively; *P* = .15 for *CEBPA*-double vs. *CEBPA*-WT and *CEBPA*-single together) as well as the lowest incidence of RD (11%, 13% and 5%; *P* = .17).

When long term outcome was analysed, with all survival probabilities given at 8 years, OS was significantly better in *CEBPA*-mutant than *CEBPA*-WT patients (44% vs. 34%; *P* = .02; Figure 4.2 A and Table 4.4), as was RFS (41% vs. 34%; *P* = .05). However, this benefit was limited to *CEBPA*-double patients, with *CEBPA*-single patients having a similar outcome to *CEBPA*-WT in terms of both OS and RFS. OS was 34%, 31% and 54% for *CEBPA*-WT, *CEBPA*-single and *CEBPA*-double respectively (*P* = .004; Figure 4.2 B and Table 4.4) and RFS was 34%, 35% and 45% (*P* = .04; Table 4.4). Similar results were obtained if patients were censored at the time of transplantation (OS: 49%, 40% and 69% for *CEBPA*-WT, *CEBPA*-single and *CEBPA*-double, respectively; *P* = .02), and if only patients in the intermediate cytogenetic risk group were analysed (OS: 34%, 30% and 51%, respectively; *P* = .02). Although *CEBPA*-double cases had the lowest CIR, this was not statistically significant (50%, 60% and 44%; *P* = .2; Figure 4.2 C and Table 4.4).

Table 4.4: Unadjusted results for response to therapy and outcome at 8 years

	CEBPA				*CEBPA*			Dose effect		*CEBPA*-double vs. *CEBPA*-WT+*CEBPA*-single	*P*
	WT	MUT	OR/HR (95% CI)	*P*	WT	Single	Double	OR/HR (95% CI)	*P*		
No. of patients	1320	107			1320	48	59				
Response to therapy											
CR (OR)	83%	84%	0.91 (0.54-1.53)	0.7	83%	77%	90%	0.85 (0.60-1.21)	0.4	0.60 (0.30-1.20)	0.15
RD (OR)	11%	8%	0.80 (0.42-1.52)	0.5	11%	13%	5%	0.78 (0.49-1.24)	0.3	0.55 (0.24-1.30)	0.17
ID (OR)	7%	7%	1.13 (0.52-2.49)	0.8	7%	10%	5%	0.99 (0.61-1.61)	0.9	0.76 (0.27-2.15)	0.6
Outcome at 8 years											
OS (HR)	34%	44%	0.77 (0.62-0.97)	0.02	34%	31%	54%	0.83 (0.72-0.95)	0.004	0.79 (0.57-1.11)	0.17
RFS (HR)	34%	41%	0.79 (0.62-1.01)	0.05	34%	35%	45%	0.86 (0.74-0.99)	0.04	0.73 (0.54-1.00)	0.05
CIR (HR)	50%	50%	0.87 (0.65-1.15)	0.3	50%	60%	44%	0.89 (0.75-1.05)	0.2	0.81 (0.55-1.17)	0.3
CID in CR (HR)	17%	9%	0.59 (0.36-0.97)	0.03	17%	5%	12%	0.77 (0.58-1.03)	0.06	0.71 (0.38-1.32)	0.3

Abbreviations: CR, complete remission; RD, resistant disease; ID, induction death; OS, overall survival; RFS, relapse-free survival; CIR, cumulative incidence of relapse; CID in CR, cumulative incidence of death in complete remission; WT, Wild-type; MUT, mutant; OR, odds ratio; HR, hazard ratio; CI, confidence intervals.

Multivariate analysis was performed considering the known prognostic factors: age, sex, type of leukaemia, cytogenetic risk group, WBC, performance status, *FLT3*/ITD and *NPM1* mutant status, and the data is shown in Table 4.5. *CEBPA*-double remained an independent favourable prognostic factor for both OS (P = .004) and RFS (P = .02). Although in univariate analysis *CEBPA*-double had not had a statistically significant beneficial impact on CIR, in multivariate analysis it was an independent favourable factor for relapse (P = .02), suggesting that other factors such as the relative lack of coincident *NPM1* mutations had obscured its benefit previously.

Mutations in *CEBPA* can be classified not only by their number but also by their type, and the possible impact of mutation location was therefore also investigated. Only patients with either or both of the two "classic" mutations, as described in detail in chapter 3, were included in this analysis, i.e. mutations predicted to cause p30 isoform translation (N-terminal) and/or those predicted to cause loss of the C-terminal DNA-binding/leucine-zipper domain (C-terminal). The 14 patients excluded from this analysis, including the 3 with homozygous mutations, are indicated in Appendix Table 2. The outcome for both N-terminal and C-terminal *CEBPA*-single patients was similar to that of *CEBPA*-WT patients and was significantly worse than that of *CEBPA*-double patients. Estimates for OS were 34%, 17%, 39% and 58% for *CEBPA*-WT, N-terminal alone, C-terminal alone and *CEBPA*-double, respectively (P = .008 for heterogeneity; Figure 4.2 D) and for CIR were 50%, 90%, 57% and 41% (P = .05). Although N-terminal alone patients had both the lowest OS (17%) and the highest CIR (90%), there were no statistical differences in outcome between *CEBPA*-WT, N-terminal alone or C-terminal alone patients in pair-wise comparisons using the log-rank test (OS: P = .8 for *CEBPA*-WT vs. N-terminal alone, P = .6 for *CEBPA*-WT vs. C-terminal alone, P = .9 for N-terminal alone vs. C-terminal alone; CIR: P = .5, .6 and .3, respectively).

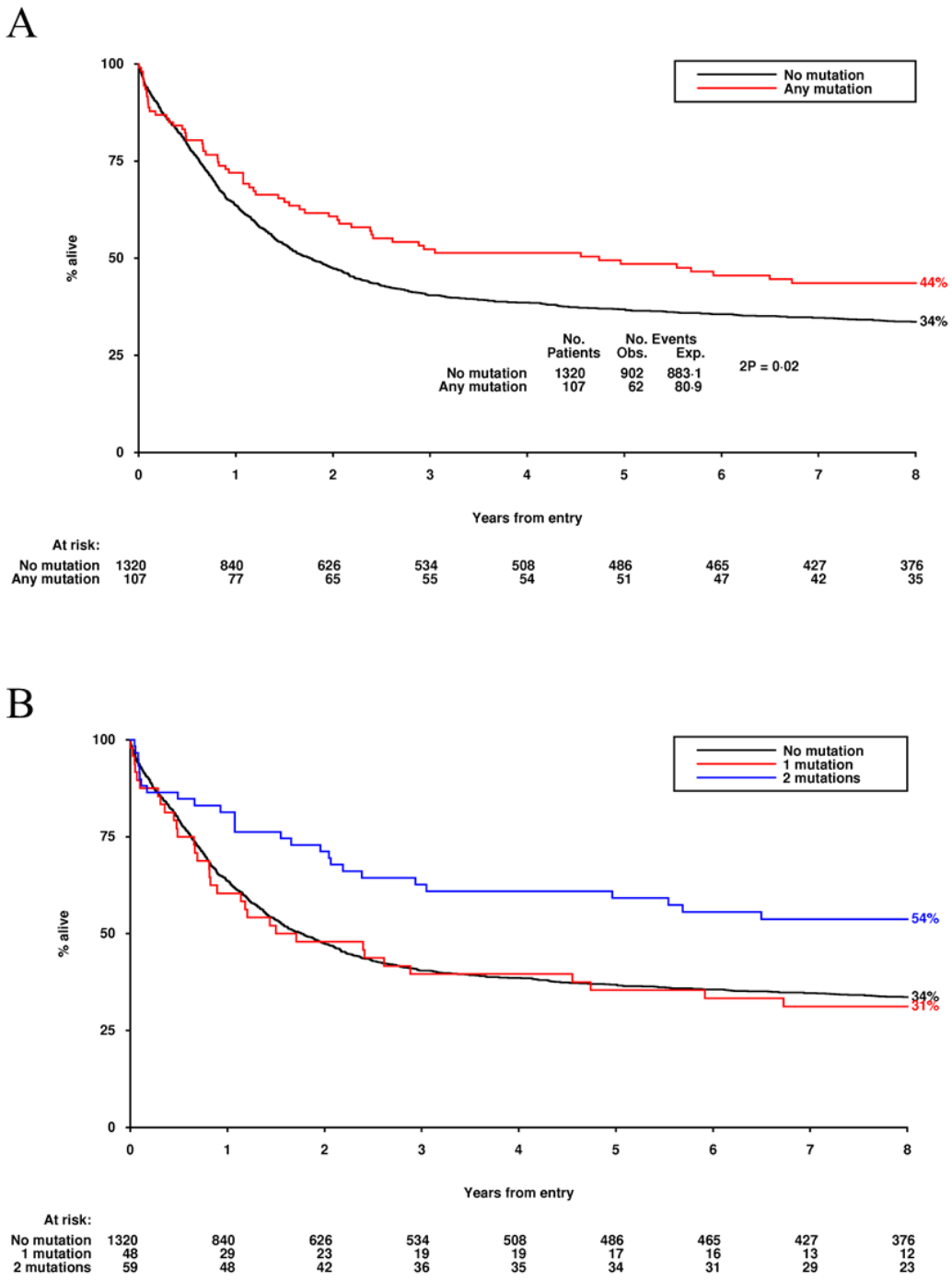
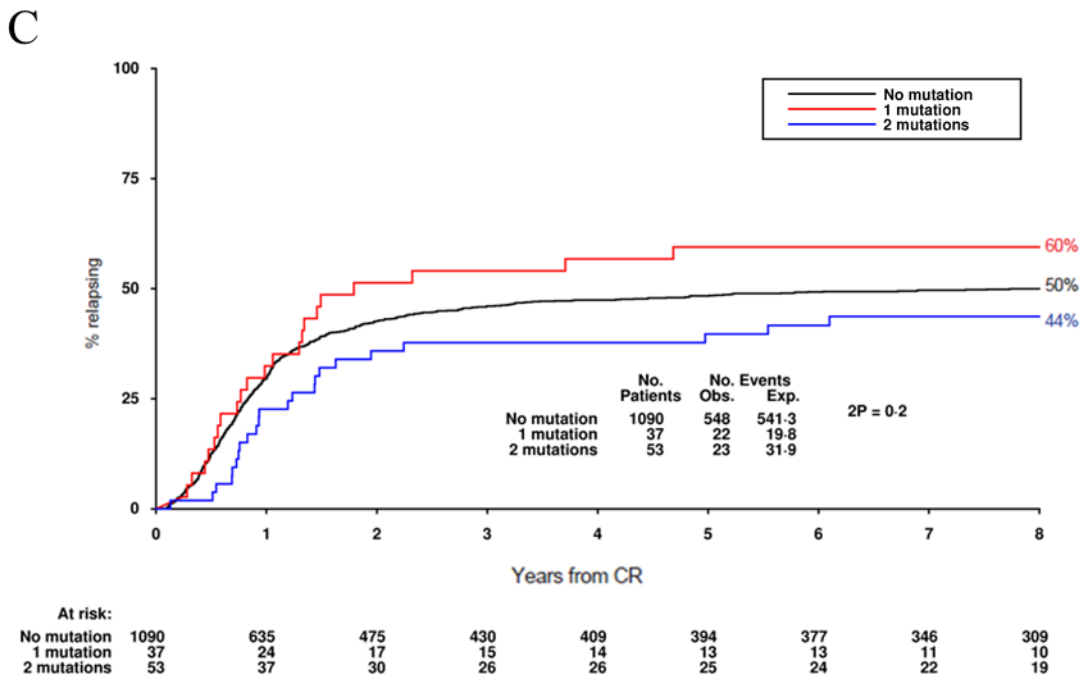


Figure 4.2 The impact of *CEBPA* mutant status on clinical outcome. Kaplan-Meier curves for overall survival stratified by (A) *CEBPA*-Wild-type (WT, no mutation) or *CEBPA*-mutant (any mutation) status and (B) *CEBPA*-WT, *CEBPA*-single (1 mutation) or *CEBPA*-double (2 mutations) mutant status. (C) Cumulative incidence of relapse by *CEBPA*-WT, *CEBPA*-single or *CEBPA*-double mutant status. (D) OS stratified by *CEBPA*-WT, single N-terminal (term), single C-term, or double N-term plus C-term mutant status.



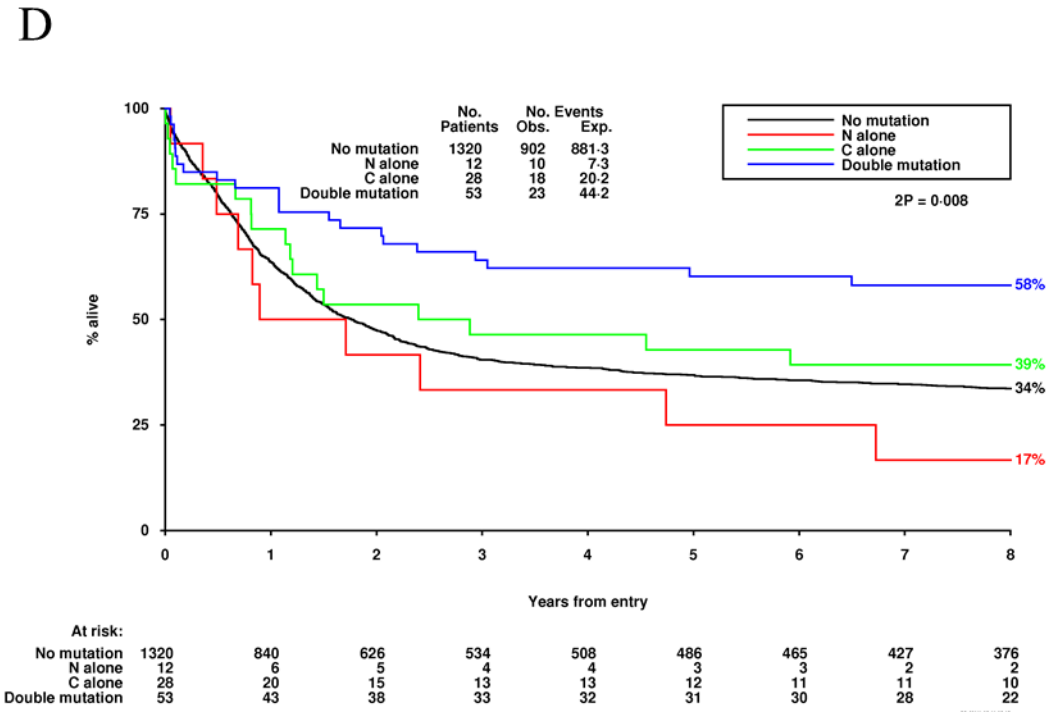


Figure 4.2 Continued

Table 4.5: Multivariate analysis of outcome

	Odds or hazard ratios (95% confidence intervals)*				
Variable	**CR**	**OS**	**RFS**	**Relapse**	**Death in CR**
Age	1.04 (1.04-1.06)	1.02 (1.01-1.03)	1.01 (1.01-1.02)	1.01 (1.00-1.02)	1.03 (1.01-1.04)
WBC	1.006 (1.003-1.009)	1.002 (1.001-1.003)	1.001 (1.000-1.002)	1.001 (1.000-1.003)	1.001 (0.999-1.004)
Performance Status	1.36 (1.16-1.60)	1.10 (1.03-1.18)	1.03 (0.95-1.11)	1.03 (0.94-1.12)	1.04 (0.89-1.22)
Male	1.03 (0.72-1.49)	1.11 (0.96-1.28)	1.05 (0.90-1.23)	1.01 (0.85-1.21)	1.17 (0.85-1.61)
Cytogenetic group	4.64 (3.20-6.71)	1.90 (1.65-2.18)	1.65 (1.41-1.94)	1.89 (1.58-2.27)	1.08 (0.78-1.50)
Secondary	2.29 (1.35-3.91)	1.23 (0.96-1.58)	1.09 (0.79-1.51)	1.10 (0.76-1.59)	1.09 (0.56-2.12)
***FLT3*/ITD**	1.24 (0.79-1.94)	1.59 (1.34-1.88)	1.69 (1.41-2.03)	2.09 (1.70-2.57)	0.82 (0.54-1.25)
***NPM1*-mutant**	0.27 (0.17-0.43)	0.46 (0.39-0.55)	0.48 (0.40-0.57)	0.42 (0.34-0.52)	0.69 (0.48-0.99)
***CEBPA*-double**	0.54 (0.18-1.61) *P* = .3	0.53 (0.35-0.81) *P* = .004	0.60 (0.40-0.91) *P* = .02	0.57 (0.346-0.93) *P* = .02	0.74 (0.32-1.70) *P* = .5

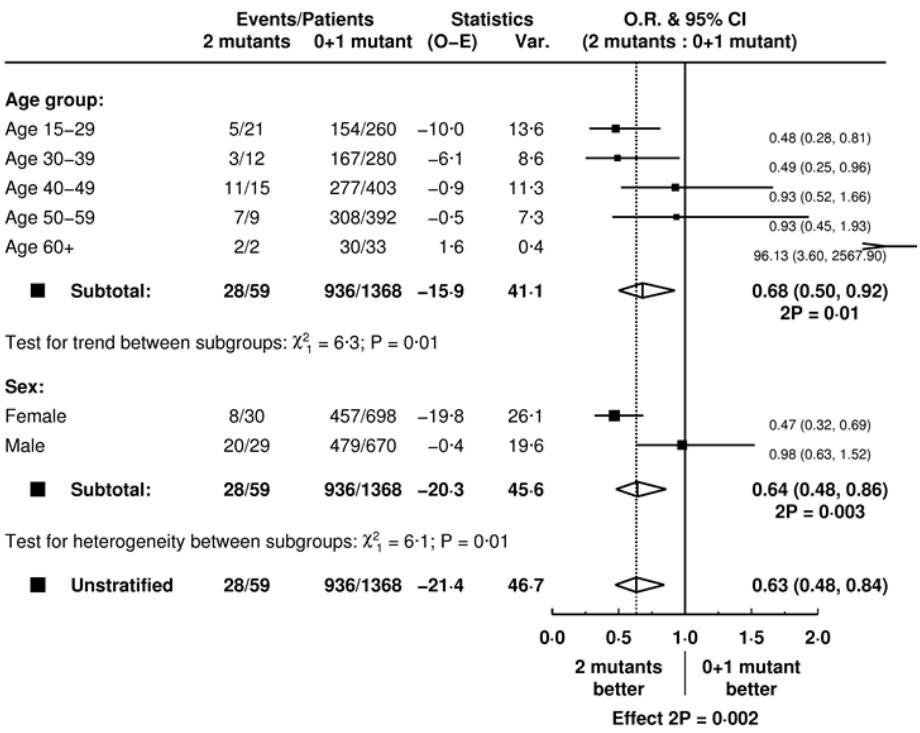
*Factors entered into the model were age, sex, type of leukaemia, cytogenetics, WBC, performance status, *FLT3*/ITD and *NPM1* mutant status. Abbreviations: CR, complete remission; OS, overall survival; RFS, relapse-free survival; WBC, white blood cell count

4.3.3 Modifying factors for outcome in *CEBPA* mutant AML

Analyses were performed in order to investigate patient characteristics that could modify the beneficial impact of double *CEBPA* mutations. For these analyses, *CEBPA*-WT and *CEBPA*-single patients were grouped together, as no difference in outcome had been found between them. Tests for heterogeneity indicated that the benefit of *CEBPA*-double mutations was predominantly limited to younger patients and female patients (Figure 4.3 A). The impact of karyotype was also explored in tests for heterogeneity with patients stratified by cytogenetic risk group, with the intermediate-risk patients further divided into those with an NK and those with an AK. This analysis showed that the benefit of *CEBPA*-double mutations was at least as great in the intermediate-risk AK patients as those with an NK (P = .7 for OS and P = .1 for CIR; Figure 4.3 B).

The interaction between *CEBPA* mutations and *FLT3*/ITDs or *NPM1* mutations was also explored. In terms of interaction with *FLT3*/ITDs, it was found that the favourable outcome of *CEBPA*-double patients was lost in the presence of a *FLT3*/ITD, OS for *CEBPA*-WT, *CEBPA*-single and *CEBPA*-double was 36%, 35% and 59%, respectively, in *FLT3*/ITD-negative (P = .002) and 26%, 21% and 14%, respectively, in *FLT3*/ITD-positive patients (P = .5) (Figure 4.4 A and B). In addition, there was no evidence that the presence of a *CEBPA*-single mutation further improved the favourable outcome of an *NPM1* mutation, with no difference in outcome between *NPM1*-mutant/*CEBPA*-WT, *NPM1*-mutant/*CEBPA*-single and *NPM1*-WT/*CEBPA*-double patients: OS was 45%, 44% and 56%, respectively (P =.2; Figure 4.4 C). Only two patients were both *NPM1*-mutant and *CEBPA*-double and so this combination was not analysed. This result was not different if patients were also stratified according to *FLT3*/ITD status, in *FLT3*/ITD-negative patients OS for *NPM1*-mutant/*CEBPA*-WT, *NPM1*-mutant/*CEBPA*-single and *NPM1*-WT/*CEBPA*-double was 55%, 54% and 60%, respectively, (P = .6) and in *FLT3*/ITD-positive patients 32%, 21% and 17%, respectively, (P = .8).

A



B

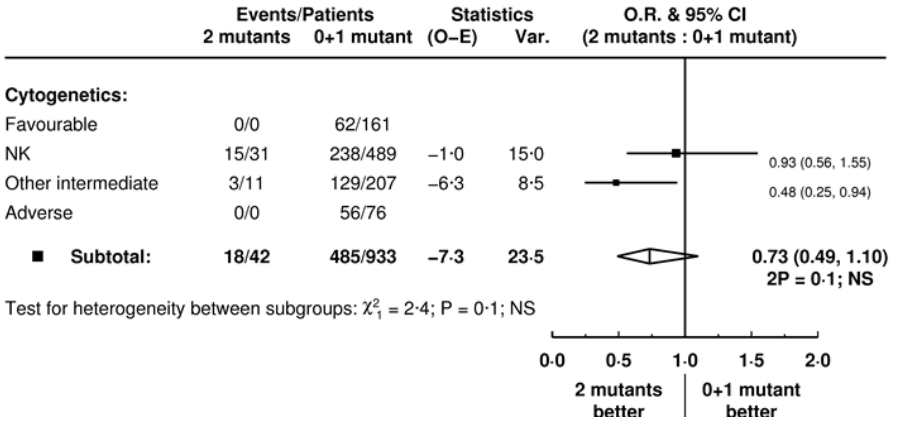


Figure 4.3 Factors modifying the impact of *CEBPA* mutant status on clinical outcome. Mantel-Byar analyses for *CEBPA*-double (2 Mutants) cases compared to *CEBPA*-WT and *CEBPA*-single together (0+1 Mutant) of (A) overall survival stratified by age and sex and (B) relapse risk stratified by cytogenetic risk group. Abbreviations: O-E, observed minus expected; Var, variance; HR, hazard ratio; CI, confidence interval; OR, odds ratio; NK, normal karyotype; NS, not significant.

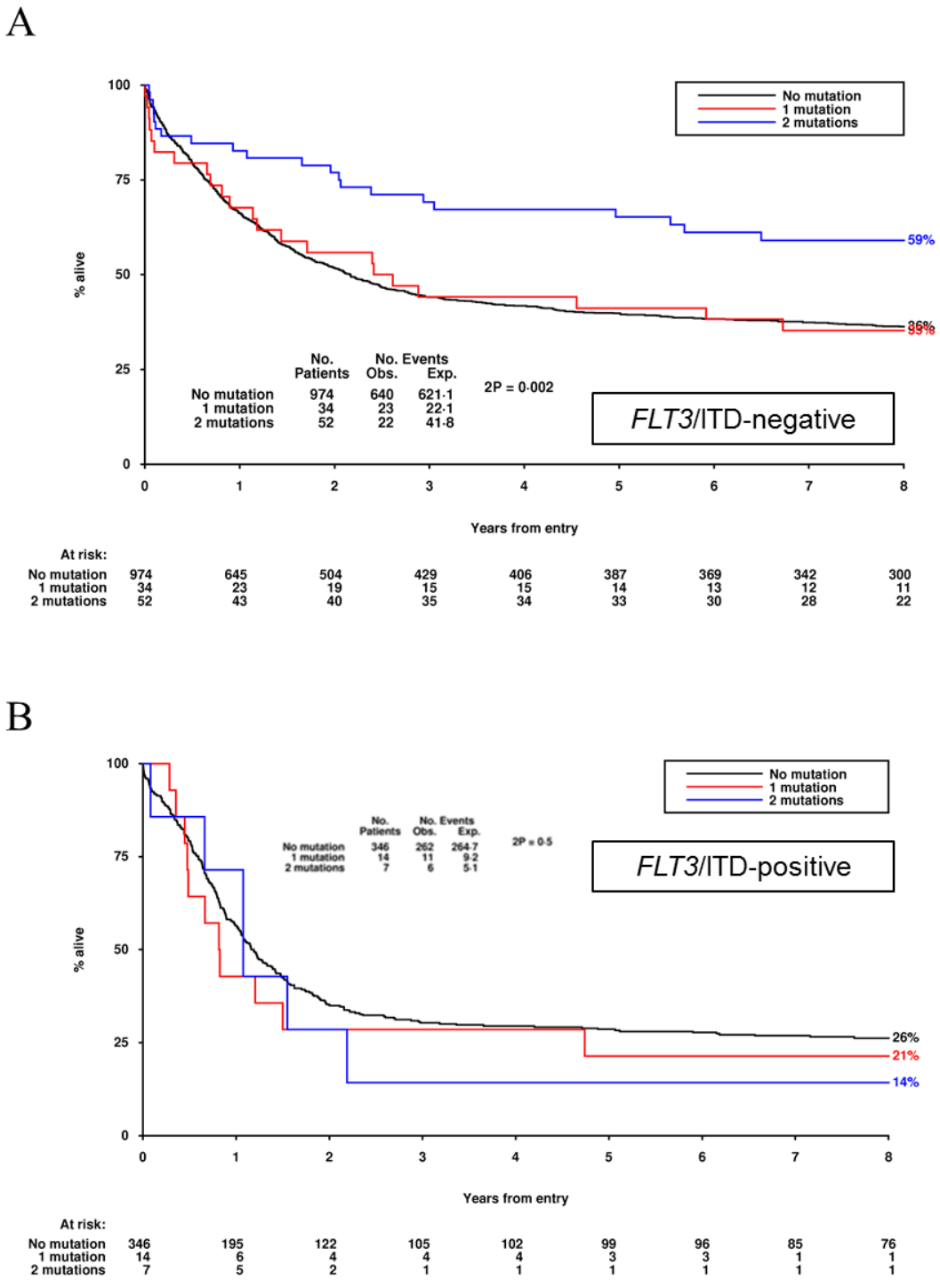


Figure 4.4 Interaction of *CEBPA* mutations with *FLT3*/ITD and *NPM1* mutations
Kaplan-Meier curves for overall survival (OS) stratified by *CEBPA* mutant status in (A) *FLT3*/ITD negative and (B) *FLT3*/ITD positive patients. (C) OS in patients stratified by *NPM1* and *CEBPA* status (0, 1 and 2 mut for *CEBPA*-WT, *CEBPA*-single and *CEBPA*-double cases, respectively).

C

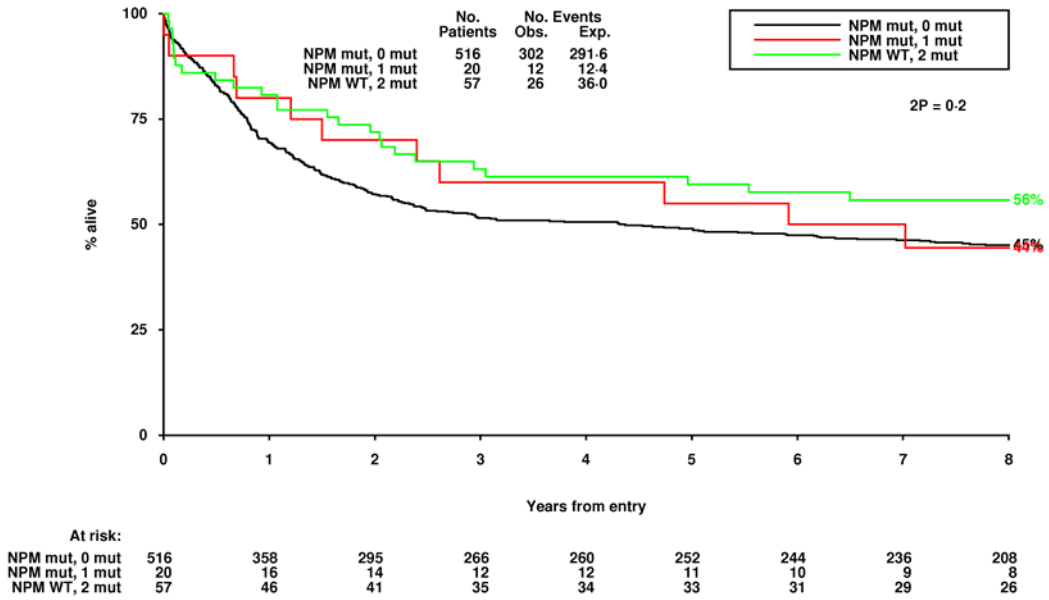


Figure 4.4 Continued

4.4 Discussion

The data presented in this chapter shows an investigation of *CEBPA*-mutated AML in terms of patient characteristics and the impact of mutations on outcome in a large cohort of well-characterised young adult AML patients with long term follow-up. It was found that *CEBPA*-double, but not *CEBPA*-single, mutated patients had an improved outcome over those without *CEBPA* mutations, with this benefit being lost in the presence of a *FLT3*/ITD.

There have been many reports on the characteristics of AML patients with *CEBPA* mutations and, in general, there is a great deal of agreement, both across these different reports and with the data presented in this chapter, which is indicative of a strong association with particular characteristics in this subgroup of patients. For example, most studies have also found no differences between different *CEBPA* mutation groups for sex or presenting WBC (Preudhomme *et al*, 2002;Barjesteh van Waalwijk van Doorn-Khosrovani *et al*, 2003;Frohling *et al*, 2004;Shih *et al*, 2006;Marcucci *et al*, 2008;Wouters *et al*, 2009;Hou *et al*, 2009;Taskesen *et al*, 2011) and all studies, including this one, have consistently found an association between *CEBPA* mutations and the M1 and M2 FAB-types, and with intermediate-risk karyotypes, in particular an NK (Tables 4.1 and 4.3). It is of note that no *CEBPA* mutations were detected in 167 patients with a favourable-risk karyotype, which may be a reflection of the disruption of C/EBPα activity by alternative mechanisms. For instance, C/EBPα protein translation has been shown to be suppressed by the product of the oncogenic CBFB-MYH11 fusion in inv(16) AML via a calreticulin-dependent mechanism (Helbling *et al*, 2005). Additionally, in t(8;21) AML the product of the RUNX1-RUNX1T1 fusion (also known as AML1-ETO) has been shown to suppress *CEBPA* transcription, potentially by inhibition of positive auto-regulation of the *CEBPA* promoter (Pabst *et al*, 2001b). Therefore in patients with these particular chromosomal abnormalities, the acquisition of a *CEBPA* mutation may be functionally redundant.

Whilst there is much agreement between different study groups on the features of *CEBPA*-mutated AML, the overall incidence of either *CEBPA*-single or *CEBPA*-double mutations is relatively rare (3% and 4%, respectively in this cohort), and this means that some features may only be discerned when very large numbers of patients have been studied. For instance, previous reports found no differences between *CEBPA*-WT, *CEBPA*-single and *CEBPA*-double patients in terms of age (Wouters *et al*, 2009;Pabst *et al*, 2009;Hou *et al*, 2009;Dufour *et al*, 2010), but in this study *CEBPA*-double cases were found to be significantly younger than other patients ($P = .003$), and this was also noted in a recent analysis of 1182 NK cases (Taskesen *et al*, 2011). However, in contrast with the study of

Taskesen *et al* (2011), which reported that *CEBPA*-single patients were less likely to have *NPM1* mutations than those without *CEBPA* mutations, there were no differences in the incidences of *FLT3*/ITD or *NPM1* mutations between *CEBPA*-WT and *CEBPA*-single cases in the current cohort. By comparison, *CEBPA*-double patients were significantly less likely to have a *FLT3*/ITD than *CEBPA*-WT and *CEBPA*-single patients, and highly unlikely to have an *NPM1* mutation, which is consistent with other reports (Wouters *et al*, 2009;Hou *et al*, 2009;Dufour *et al*, 2010;Taskesen *et al*, 2011).

Given that, unlike *CEBPA*-single cases, patients with *CEBPA*-double mutations generally lack mutations in other genes that are commonly mutated in AML, interesting questions are raised regarding the sequence of events leading to leukaemic transformation. In patients with two *CEBPA* mutations these could affect the same allele, monoallelic mutations, or different alleles, biallelic mutations, with loss of normal allele, and this is explored further in chapter 5. However, evidence from mouse models suggests that one disrupted *CEBPA* allele is not sufficient for leukaemogenesis. Kirstetter and colleagues (2008) generated a *CEBPA* allele lacking the N-terminus (L), which allowed p30 but not p42 isoform expression and mimicked the effect of an N-terminal frame-shifting mutation. Mice homozygous for the L allele had GMPs with vastly increased self-renewal capacity and developed a transplantable AML with complete penetrance. Heterozygous mice, L/+, did not develop AML and had similar haemopoiesis profiles to WT mice. However, there was a decrease in the p42 to p30 isoform ratio, and myeloid progenitors from the bone marrow were found to have mildly increased proliferative capacity compared with those from WT mice. A follow-up study (Bereshchenko *et al*, 2009) combined the p42-null L allele with a K allele containing a classic in-frame C-terminal mutation that has been found in a number of AML patients (K313dup) in a competitive transplantation model. All mice receiving cells with two affected *CEBPA* alleles (K/K, K/L or L/L) developed a transplantable leukaemia, with the most efficient leukaemogenesis in mice receiving K/L cells. The authors hypothesised that the efficient leukaemogenesis observed in mice receiving K/L cells was caused by premalignant haemopoietic expansion induced by the K allele combined with residual myeloid commitment maintained by the L allele, and that this provided a potential explanation for the observation of both types of mutation commonly found together in patients.

Further evidence that a single *CEBPA* mutation may be an early step in the process, but not leukaemogenic *per se*, is the observation in familial AML with heterozygous germline *CEBPA* mutations that, at disease presentation, an acquired mutation on the normal allele is very common, occurring in 78% of patients (Smith *et al*, 2004;Nanri *et al*, 2006;Corbacioglu

et al, 2007;Pabst *et al*, 2008;Renneville *et al*, 2009b;Taskesen *et al*, 2011). In addition, there can be a long latency before disease on-set of up to 46 years (Pabst *et al*, 2008). Of note, there are no reported patients with germline *CEBPA* mutations and a family history of AML with acquired *FLT3*/ITD or *NPM1* mutations at disease presentation, although one case has been described with both *KRAS* and *WT1* mutations (Taskesen *et al*, 2011). Biallelic *CEBPA* mutations lead to the development of AML in a mouse model with full penetrance indicating that, in the mouse at least, further initiating events are not required. However, further complexity is likely for leukaemogenesis in human patients and, whilst *NPM1* and *FLT3*/ITD mutations are not common in *CEBPA*-double cases, recent evidence has shown a strong specific association with mutations in *GATA2*, a gene coding for a transcription factor that is important for megakaryocytic differentiation and the proliferation and maintenance of haemopoietic stem/progenitor cells (Greif *et al*, 2012).

In terms of patient outcome, as has been historically reported (Preudhomme *et al*, 2002;Barjesteh van Waalwijk van Doorn-Khosrovani *et al*, 2003;Frohling *et al*, 2004;Bienz *et al*, 2005;Schlenk *et al*, 2008;Marcucci *et al*, 2008) it was found that *CEBPA*-mutant cases had a better outcome than *CEBPA*-WT. However, the benefit seen in the present cohort was restricted to *CEBPA*-double mutant patients, with very similar outcomes between *CEBPA*-single and *CEBPA*-WT cases (Figure 4.2 B). At the time the work presented in this chapter was completed, this finding was in agreement with a study from Wouters *et al* (2009) and a report of a smaller cohort (Pabst *et al*, 2009), but Renneville *et al* (2009a) had not found this clear difference in outcome. Subsequently, there have been three further studies, two restricted to only NK patients, that have also found that the benefit for *CEBPA*-mutations is limited to *CEBPA*-double cases (Hou *et al*, 2009;Dufour *et al*, 2010;Taskesen *et al*, 2011) (Table 4.2). It was also important to determine that it was the presence of *CEBPA*-double mutations, and not simply the presence of a particular type of mutation, such as a classic N-terminal or C-terminal mutation, that was impacting upon outcome. Whilst 90% of *CEBPA*-double cases had both types of mutation, only 25% and 58% of *CEBPA*-single patients had a classic N-terminal or C-terminal mutation, respectively. In addition, it has been suggested that there may be a difference in outcome between *CEBPA*-single cases with an N-terminal and those with a C-terminal mutation (Taube *et al*, 2009), and that the GEPs of *CEBPA*-single cases with a C-terminal mutation have a tendency to be potentially less distinct from GEPs of *CEBPA*-double cases than do those from *CEBPA*-single patients with an N-terminal mutation (Wouters *et al*, 2009). However, when the impact of mutation location on outcome was investigated in the present cohort, no significant differences were found (Figure 4.2 D), and this is in line with a more recent study by Taskesen *et al* (2011). AML with mutated *CEBPA* has been added as a provisional entry to the updated WHO classification of AML

(Swerdlow *et al*, 2008;Vardiman *et al*, 2009), however there is now a growing body of evidence that, rather than a *CEBPA* mutation *per se* defining a separate entity, it is the presence of *CEBPA*-double mutations that defines a group of patients with particular characteristics and GEPs (Wouters *et al*, 2009;Dufour *et al*, 2010;Taskesen *et al*, 2011).

Many studies of *CEBPA* mutations in AML have focussed solely on NK patients (Tables 4.1. and 4.2), however this excludes intermediate-risk AK patients who accounted for a significant proportion of this cohort (273 of 1427, 19%), and in whom a *CEBPA*-double mutation status could be very informative for outcome. In addition, the incidence of *CEBPA*-double mutations was similar in intermediate-risk AK and NK patients (5% and 6%, respectively). There were, therefore, sufficient intermediate-risk AK cases in this large cohort of patients to analyse whether karyotype modified the impact of *CEBPA* mutations on outcome. It was found that the benefit for *CEBPA*-double mutants was at least as great in intermediate-risk AK cases as in NK patients (Figure 4.3 B). This is in contrast to a study on ALFA trial patients (Renneville *et al*, 2009a), where the benefit of *CEBPA* mutations was mostly seen in those with an NK. This difference may be due to the way in which patients were grouped for analyses. In the present study the comparison was made between patients with and without *CEBPA*-double mutations, whereas in the study of Renneville *et al* (2009a), *CEBPA*-mutant patients were divided into three subsets defined by karyotype and *FLT3*/ITD status (NK and *FLT3*/ITD-negative, NK and *FLT3*/ITD-positive, and AK) and not by the number of *CEBPA* mutations.

The interaction between different cooperating events in leukaemogenesis and the impact on outcome of the particular combination present is important if molecular markers are to be used for risk-stratification of patients for therapy. For instance, *FLT3*/ITD and *NPM1* mutations are, respectively, known adverse and favourable factors for patient prognosis, with some controversy over the impact on prognosis if both are present (see chapter 1). The impact of *FLT3*/ITD mutations in *CEBPA*-mutant AML has been studied previously with differing results: either no effect reported (Frohling *et al*, 2004;Marcucci *et al*, 2008) or a negative impact on outcome (Renneville *et al*, 2009a). In the present cohort, the benefit in outcome for *CEBPA*-double mutations was completely lost in the presence of a *FLT*/ITD (Figure 4.4 A and B), the first analysis of the impact of *FLT3*/ITD mutations in this subgroup. By contrast with the strong impact of a *FLT3*/ITD, there was no evidence that a *CEBPA*-single mutation further improved the favourable outcome of *NPM1*-mutant AML (Figure 4.4 C), which has not been reported by others. Only 2 of 1427 (0.1%) patients investigated were both *CEBPA*-double and *NPM1* mutated, and therefore this combination was not investigated for outcome. In line with this observation and the finding that a

FLT3/ITD had a poor impact on outcome regardless of *CEBPA* genotype (Figure 4.4 A and B), a recent large collaborative study of 1182 NK patients found that the outcome of *CEBPA*-single patients was influenced by their *FLT3*/ITD and *NPM1* mutant status (Taskesen *et al*, 2011). However, it was also noted in this report that the outcome for *CEBPA*-single patients was better, although not significantly so, than *CEBPA*-WT patients in all equivalent groups defined by *FLT3*/ITD and *NPM1* status. There was no evidence of this in the present cohort, with no difference in outcome seen between *CEBPA*-single and *CEBPA*-WT patients either in the whole cohort or in groups defined by *FLT3*/ITD or *NPM1* mutation status. Outcome in *CEBPA*-mutant AML therefore appears to be dependent on what the particular cooperating event(s) may be. Whilst many *CEBPA*-single cases have an additional *NPM1* or *FLT3*/ITD mutation, in the current cohort, 21 of 48 (44%) *CEBPA*-single patients lacked either alteration. In these patients alternative factors such as cytogenetics and other gene mutations with prognostic implications, such as those in *DNMT3A* and *IDH2*, may be important.

There are a large number of genes known to be commonly mutated in AML and several of them, such as *FLT3* and *NPM1*, are now routinely screened in the diagnostic work-up of new cases. As the number of markers increases, some rationalisation will no doubt be required to target molecular analyses to cases where knowledge of mutation status would be most useful for risk stratification of therapy. The data presented here suggest that *CEBPA* mutation screening could be restricted to intermediate-risk karyotype patients lacking a *FLT3*/ITD or *NPM1* mutation, although it should not be limited to just those with an NK. Using such a strategy for the current cohort, only a third of patients would have required *CEBPA* screening.

Several studies have suggested that *CEBPA* mutant-positive patients should be grouped in the favourable risk category and should therefore not receive a transplant in first remission (Preudhomme *et al*, 2002;Barjesteh van Waalwijk van Doorn-Khosrovani *et al*, 2003;Marcucci *et al*, 2008;Schlenk *et al*, 2008). The improved outcome for *CEBPA*–double patients in our cohort was most apparent for OS. Although this may reflect the trend for better remission rate in *CEBPA*-double patents and the reduced relapse rate, it also suggests that these patients are more likely to respond to salvage therapy after relapse. This has implications for the use of transplantation in these patients and warrants further investigation, but, with the relatively low incidence of *CEBPA*-double patients, this will require a meta-analysis of several cohorts.

CHAPTER 5: QUANTIFICATION OF CEBPA MUTANT LEVEL AND CONFIRMATION OF BIALLELIC MUTATIONS

5.1 Introduction

Mutations in *CEBPA* were detected in 107 patients in the present cohort and features of these alterations are explored further in this chapter.

5.1.1 Mutant level in AML

Mutant level can be defined as the amount of mutant allele present in a DNA sample as a proportion of the total alleles, and is therefore a reflection of the overall allelic composition of the cells in that sample. There are three main states for a given locus in a cell, homozygous, heterozygous and hemizygous. Homozygosity indicates the presence of two identical alleles for that locus on both homologous chromosomes, heterozygosity indicates different alleles and hemizygosity indicates that only one copy of the locus is present. In a genetically identical population of cells, therefore, the mutant level of a homozygous mutation or a hemizygous mutation in all cells would be 100%. For a mutation in a heterozygous state with wild-type (WT) allele in all cells, the mutant level would be 50%. In AML, studies have shown that mutant level can impact on clinical outcome. It may also be informative for determining the order of mutation acquisition.

For example, the mutant level of a *FLT3*/ITD, a mutation which is associated with adverse patient outcome, has been shown to be important for prognosis. One early report of 23 patients with an NK and a *FLT3*/ITD found that only the cases with more mutant allele than WT, as assessed by agarose gel electrophoresis of PCR products, had a significantly poorer outcome than NK patients without a *FLT3*/ITD (Whitman *et al*, 2001). A second study of 59 cases with a *FLT3*/ITD and an intermediate-risk karyotype found that a mutant level higher than the median level of 44% was associated with an adverse prognosis, while patients with a *FLT3*/ITD mutant level less than this had a similar outcome to *FLT3*/ITD-negative cases (Thiede *et al*, 2002). A report from our own group on a large cohort of 354 *FLT3*/ITD-positive cases from the MRC AML 10 and 12 trials split patients into those with a low (less than 25%), intermediate (25% to 50%) or high (greater than 50%) mutant level. In line with the two smaller studies, the poorest patient outcomes were found in those with a high level *FLT3*/ITD mutant level. However, prognosis was still significantly worse than *FLT3*/ITD-

negative cases in those with an intermediate level mutation and, even in the low mutant level group, relapse risk was still significantly higher (Gale *et al*, 2008).

Data on mutant levels can be used to infer something about the biology and order of mutation acquisition in leukaemogenesis. One of the hallmarks of AML is the clonal expansion of abnormal myeloid cells, however, there is a general consensus that leukaemogenesis, as with other malignancies, probably requires several genetic hits over a period of time (Kelly & Gilliland, 2002;Ashworth *et al*, 2011). The sequencing of individual cancer genomes has revealed large numbers of somatically acquired mutations, from less than 1,000 in some cancers to more than 100,000 in others, and these may be divided into driver and passenger mutations (see chapter 1) (Stratton *et al*, 2009). The number of driver mutations within an individual cancer probably varies between cancer types and it has been suggested it may be around 5 to 7, or as many as 20 (Stratton *et al*, 2009). These cancer-associated mutations may confer a selection advantage by alteration of a number of different cellular pathways, such as by increasing proliferation or evasion of normal apoptosis processes. This accumulation of genetic damage results in a transformed phenotype, with cells of that clone out-growing normal cells. However, the linear acquisition of mutations in a single clone is an over-simplification, with several lines of evidence for intra-tumour genetic heterogeneity, both in leukaemias and other malignancies (Greaves & Maley, 2012;Ding *et al*, 2012;Gerlinger *et al*, 2012). As depicted in Figure 5.1, some genetic hits may occur as early or primary events in leukaemogenesis and, as such, are present in the majority of blast cells at disease presentation. Other hits may be acquired in only a sub-clone of cells at a later stage of the evolution of the leukaemia. Whilst these hits confer a selection advantage to that sub-clone in the environmental conditions pertaining at that time, it is likely that at disease presentation the secondary hit will only be present in a minority of blast cells, which will be reflected in the mutant level.

The use of mutant levels in making inferences about the order of mutation acquisition can be exemplified by the combination of *FLT3*/ITD and *NPM1* mutations. In the study of the MRC AML 10 and AML 12 trial patients, the *NPM1* mutant level of 503 *NPM1*-mutant cases clustered quite tightly around the median mutant level of 43%, with only 8% of cases having a low level (less than 25%) and just 3% a high level (greater than 50%) mutant (Gale *et al*, 2008). Therefore, in the majority of *NPM1*-mutant patients the mutation was present in between

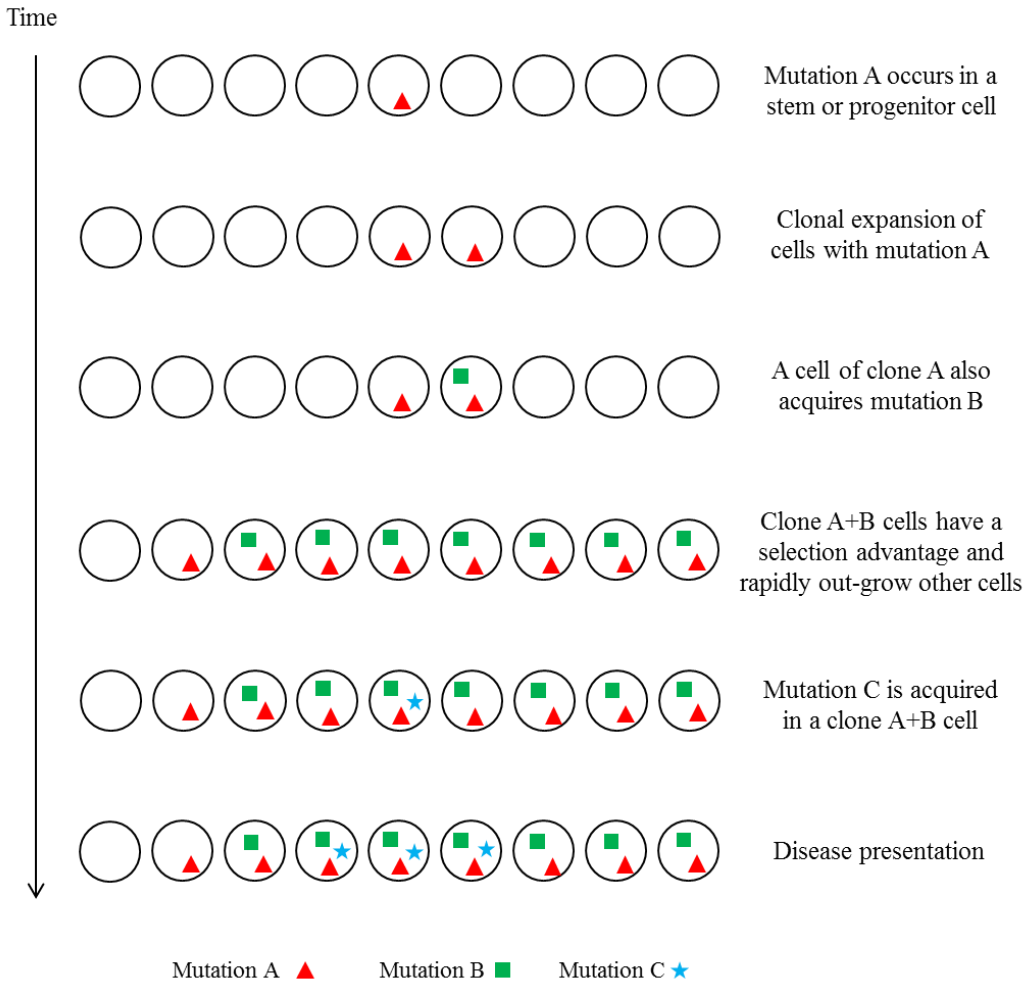


Figure 5.1 Multi-step pathogenesis of leukaemia and the biological basis of mutation level. In this example, at disease presentation, mutations A and B are present in nearly all cells and mutation C is only present in a small proportion of the population. In genomic DNA extracted from this cell population, and assuming all mutations are heterozygous, mutations A and B would have an intermediate mutant level of 39% and 44%, respectively. Mutation C would have a low mutant level of 17%.

25% and 50% of alleles (intermediate level), which is consistent with a heterozygous mutation present in most of the cells, indicating that the acquisition of the *NPM1* is likely to be an early event in leukaemogenesis.

By contrast, the *FLT3*/ITD level in the 354 *FLT3*/ITD-positive patients studied varied widely between individuals. Although the median total mutant level was 35%, this ranged between 1% and 96%, with 29% of cases classified as low, 56% as intermediate and 15% as high level mutant. In 208 patients with both a *FLT3*/ITD and an *NPM1* mutation, quantification of both mutant levels showed no correlation between the two. For the 53 patients with a high level *FLT3*/ITD, a mutant level above 50% indicated the presence of the *FLT3*/ITD in the majority of cells, with a loss of the WT allele in at least some of the cells. In previous reports, fluorescence in-situ hybridisation (FISH) analysis has found no loss or gain of chromosomal material at the *FLT3* locus on chromosome 13, however loss of heterozygosity as a result of uniparental disomy (UPD) has been indicated by SNP array data and analysis of short tandem repeats, consistent with a homozygous, rather than hemizygous, mutation (Thiede *et al*, 2002;Raghavan *et al*, 2005;Griffiths *et al*, 2005). For patients with an intermediate level *FLT3*/ITD, there are several alternatives for the allelic composition of the cell population. For example, a 50% mutant level could be the result of a heterozygous mutation in all cells, a homozygous mutation in half of the cells, or a mixed population of both. Further study of 34 of the intermediate mutant level cases was carried out using a polymorphism to investigate the genetic composition in more detail (Green *et al*, 2008). In 32 (94%) of 34 cases, no evidence of a homozygous mutation could be detected, consistent with the presence of a heterozygous mutation in most cells. In only two cases, mutant homozygosity in at least some cells was indicated, with one patient having more homozygous than heterozygous mutant cells. A low level *FLT3*/ITD suggested that less than half of the cells in the sample had a heterozygous mutation. In most of these low level *FLT3*/ITD cases it could be shown, by either X-chromosome inactivation pattern analysis for a clonal population or presence of an *NPM1* mutation at greater than 25% of total *NPM1* alleles, that a low *FLT3*/ITD level was not the result of the leukaemic cells comprising only a small proportion of the total cells analysed. Therefore, in some patients, the *FLT3*/ITD mutation was occurring in only a sub-clone of the leukaemia. In addition, a low-level *FLT3*/ITD was frequently found in combination with an intermediate-level *NPM1* mutation suggesting that, while most of the leukaemic cells carried the *NPM1* mutation, only a sub-clone also had the *FLT3*/ITD. This indicated that the *FLT3*/ITD was probably acquired after the *NPM1* mutation.

5.1.2 *CEBPA* mutant level

There is very little information available in the literature on mutant level in *CEBPA*. One study of 8 *CEBPA*-mutant cases reported only that the mutant level ranged between 33% and 60%, with no further information available on individual levels or methodology (Pabst *et al*, 2001b). In addition, although a number of groups have used a screening method for *CEBPA* length mutations that would give quantitative data, only details of the detected *CEBPA* mutations were reported, with no information on the mutant level in these patients (Lin *et al*, 2006;Juhl-Christensen *et al*, 2008;Benthaus *et al*, 2008). The level of *CEBPA* mutations is of interest as the mutations are generally considered to be early events in leukaemogenesis due to their presence in the germline of some pedigrees with familial AML and their relative stability as a marker of disease (chapter 3), and would therefore be predicted to be present at around the 50% level. However, there has been no comprehensive validation of this hypothesis and it is unknown whether this is true for all *CEBPA*-mutant cases, and further investigation is required.

As described in chapter 3, two different heterozygous mutations can be detected in approximately half of *CEBPA*-mutant patients, and there are interesting questions regarding the allelic and cellular context of the mutations in these cases. In genes where two different heterozygous mutations can be detected, they may either affect the same allele (monoallelic) or different alleles (biallelic). Figure 5.2 depicts the three possible allelic compositions for a population of leukaemic cells in a patient with two mutations, each with a mutant level of 50%. First, the different mutations could be in different cells, therefore each mutation would need to be homozygous in half of the cells in order to constitute 50% of the total alleles. Additionally, all cells could have both mutations, either monoallelically, on the same allele, or biallelically, on different alleles. Mutant level quantification in isolation would not be sufficient to confirm which is the case, and cloning of alleles is required to determine if the different mutations are on the same, or different, alleles (section 5.1.4).

5.1.3 Techniques for quantifying *CEBPA* mutant level

There are several different techniques that could be used to quantify *CEBPA* mutant level, one of which is fragment analysis of fluorescently-labelled PCR products. Although there is a lack of reported data on *CEBPA* mutant level, nevertheless, as mentioned earlier, several groups have successfully used this method to detect *CEBPA* length

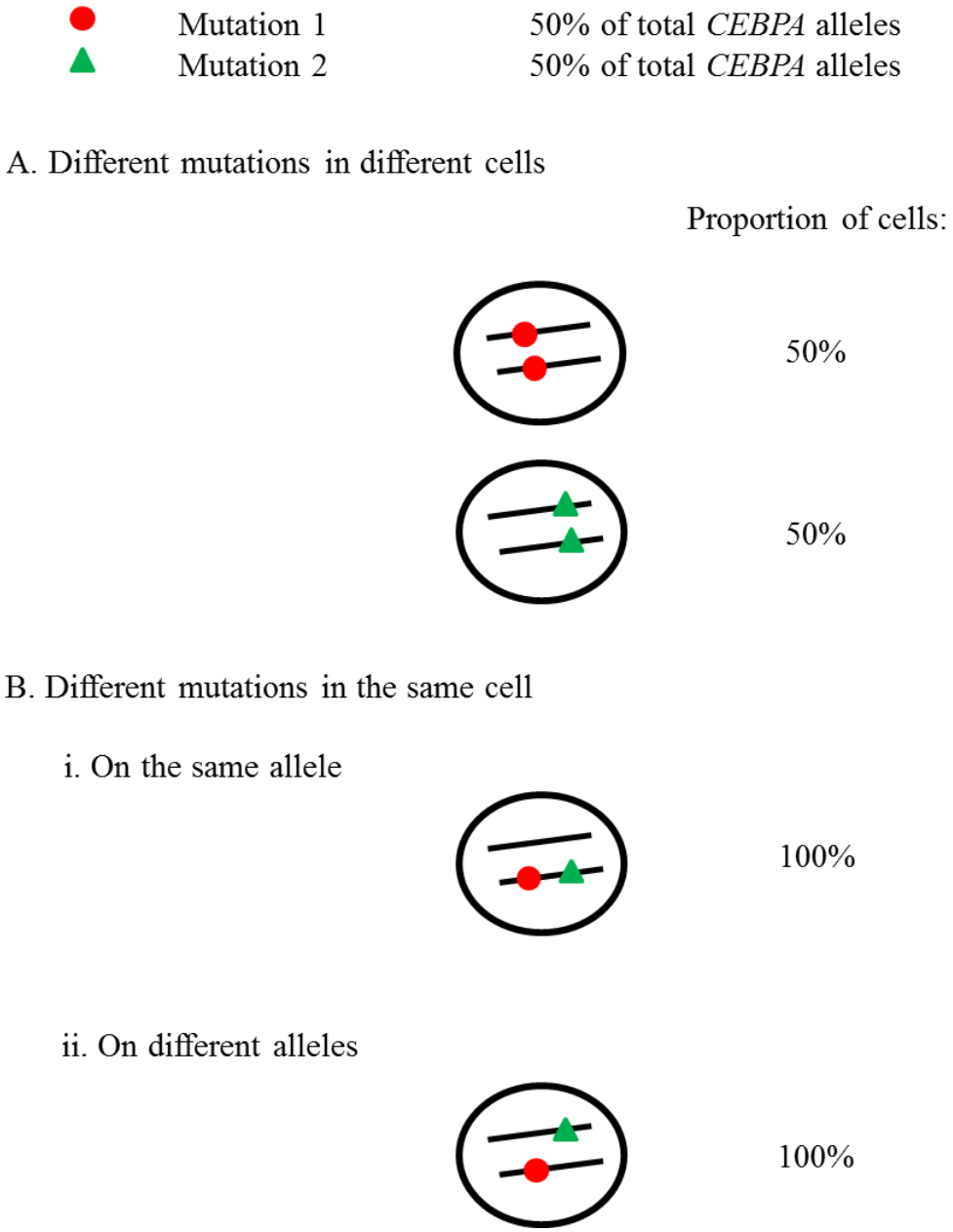


Figure 5.2 Possible allelic composition of cells in a *CEBPA*-double case.
Each mutation constitutes 50% of the total *CEBPA* alleles. Cellular composition if the different mutations occur in (A) different cells and (B) the same cell. Note, a combination of these options could also occur. For simplicity, the possibility of cells without a *CEBPA* mutation, either normal or leukaemic, has been excluded.

mutations (Lin *et al*, 2006;Juhl-Christensen *et al*, 2008;Benthaus *et al*, 2008). This technique can give a quantitative assessment of mutant level by using the area under the relevant peaks in the chromatogram to calculate the mutant allele as a percentage of the total alleles, or by calculating the ratio of the mutant peak to the WT peak. The sensitivity of the technique for detecting a mutant allele has been assessed in two reports by serially diluting a sample from a known *CEBPA*-mutant patient with a WT sample, before PCR amplification and fragment analysis. From these experiments it was found that mutations could be detected when they were approximately 5% of the total alleles, and that the observed mutant level was highly correlated with the expected level from the mixes (Lin *et al*, 2006;Benthaus *et al*, 2008). One disadvantage of this technique is that point mutations and indels not affecting overall length cannot be detected or quantified. This problem can be overcome when quantifying known point mutations, by digesting the labelled PCR product with a restriction enzyme that will digest either the WT or the mutant alleles, allowing them to be separated by size. However, this is time consuming as a new assay needs to be designed specifically for each individual mutation and may not be possible for all mutations due to the difficulty of designing a specific digestion for homopolymers of Cs and Gs, which are both common in the *CEBPA* sequence and frequently mutated.

Pyrosequencing (QIAGEN, Crawley, UK) is another technique enabling sensitive quantification of mutant level in patient samples. Unlike fragment analysis, pyrosequencing allows direct quantification of point mutations as well as insertions or deletions. However, the PCR products that can be analysed in a single assay are relatively small, at less than 100 bp. Therefore, to quantify mutations across *CEBPA*, more than 10 different overlapping PCR amplicons would need to be designed and optimised, which would be challenging given the GC-rich nature of this gene. In addition, the individual nature of the mutations present in different patients would require optimisation of the pyrosequencing assay for each one, making this technique unsuitable for this application. Quantitative real-time polymerase chain reaction (Q-PCR) based assays could also be utilised to quantify the relative levels of WT and mutant alleles in a DNA sample. However, as with pyrosequencing, the individual nature of the different mutations would require a significant amount of optimisation.

Therefore, fragment analysis was selected as the method to quantify mutant level in this study, as it has been shown to be both a sensitive and relatively straightforward technique for detecting *CEBPA* mutations. In addition, it is a quantification method that has been successfully used in the department for other genes, even for point mutations such as most *FLT3*/TKD mutations (Mead *et al*, 2007;Gale *et al*, 2008).

5.1.4 Allelic distribution of *CEBPA* mutations

As introduced above, it is possible that *CEBPA*-double patients with two different alterations have either monoallelic or biallelic mutations. There are a number of groups that have cloned the entire *CEBPA* coding sequence of such patients and sequenced clones to determine whether the mutations were present on the same allele. Of 72 cases with two different mutations that have been investigated, 52 (72%) were found to have different mutations on different alleles and, therefore, classified as biallelic *CEBPA* mutations (Preudhomme *et al*, 2002;Barjesteh van Waalwijk van Doorn-Khosrovani *et al*, 2003;Frohling *et al*, 2004;Lin *et al*, 2005;Shih *et al*, 2006;Dufour *et al*, 2010). The majority of clones from a further 17 (24%) patients were also found to bear mutations on different alleles, however one or two clones were additionally found with both mutations (Lin *et al*, 2005;Shih *et al*, 2006). In the remaining 3 (4%) patients, the mutations were carried on the same allele (Preudhomme *et al*, 2002;Frohling *et al*, 2004). Therefore, the general consensus is that the vast majority of *CEBPA*-double patients have biallelic disease. Indeed, competitive transplantation models of *CEBPA* mutations in mice have found that only mice who receive cells with two affected *CEBPA* alleles develop AML (Bereshchenko *et al*, 2009), suggesting a requirement for complete loss of the WT allele.

The data presented in chapter 4, in line with other published studies, showed that the benefit of a *CEBPA* mutation is restricted to *CEBPA*-double cases, with no impact on outcome for a single *CEBPA* mutation. If mutant status is to be used for the risk-stratification of patients it is, therefore, important to correctly classify those with *CEBPA* mutations as *CEBPA*-single or *CEBPA*-double. Standard screening techniques, such as dHPLC or direct nucleotide sequencing, are effective at detecting the presence of mutations, but if two heterozygous mutations are detected in the same patient then these methods give no information regarding the allelic distribution.

This chapter presents data on the quantification of mutant level by fragment analysis of fluorescently-labelled PCR products in the cohort of *CEBPA*-mutant cases described in chapter 3. The allelic distribution of mutations in *CEBPA*-double cases will be investigated by cloning of the entire coding sequence in order to determine whether mutations are biallelic and, by implication, whether the favourable impact on prognosis is related to the presence of two *CEBPA* mutations *per se*, or the presence of biallelic mutations.

5.2 Materials and Methods

5.2.1 Quantification of *CEBPA* mutant level by fragment analysis

As described in chapter 3, the entire coding sequence of *CEBPA* was divided into three overlapping fragments called fragments 1, 2 and 3 (see section 3.2.1). The primer pairs used to amplify each fragment were CEBPA/1F and CEBPA/1R3, CEBPA/2F and CEBPA/2R, and CEBPA/3F and CEBPA/3R, respectively (Appendix Table 1). For the quantification assay, the forward primer in each pair was labelled at the 5′ end with a fluorescent dye, D4 (WellRED oligos, Sigma-Aldrich Company Ltd., Poole, UK). Addition of the fluorescent label and the sensitivity of the technique necessitated adjustment of both the PCR reaction mix and the cycling conditions used in the standard BIOTAQ DNA Polymerase protocol given in section 2.1.2. The concentration of each primer was reduced from 0.5µM to 0.25µM and 5% DMSO was also included to aid in amplification of the GC-rich sequence. With regard to the cycling conditions, all 3 steps of each cycle were extended to 1 minute and the temperature ramping of the thermocycler block was set at 0.5°C/second. Only 28 cycles of amplification were performed, instead of 35, in order to limit the amount of PCR product generated and prevent saturation of the fluorescence detector during fragment analysis. The final extension step at 72°C was also extended from 5 minutes to 15 minutes to ensure all PCR products were fully extended. Table 5.1 summarises the primer pairs, annealing temperatures and expected amplicon sizes for each fragment.

Post-cycling, PCR products were diluted 1 in 20 in a formamide sample loading solution containing a DNA size standard ladder (DNA Size Standard Kit – 600, Beckman Coulter, High Wycombe, UK) and analysed by capillary electrophoresis using the CEQ 8000 Genetic Analysis System (Beckman Coulter). For reference, all runs included a PCR product amplified from a cell line with known WT *CEBPA*, such as HL60 or NB4, and this was used to identify the WT peak in patient samples. Mutant peaks representing insertion or deletion mutations were identified by their larger or smaller fragment sizes and the area under

Table 5.1 Summary of PCR protocols for fragment analysis

Fragment	Forward primer	Reverse primer	Annealing Temperature	Expected WT fragment size, bp	Observed WT fragment size* (bp ±2SD)	Observed – Expected (bp)
1	D4-CEBPA/1F	CEBPA/1R3	62°C	548	524 ±4.37	-24
2	D4-CEBPA/2F	CEBPA/2R	62°C	463	447 ±2.19	-16
3	D4-CEBPA/3F	CEBPA/3R	64°C	424	410 ±2.28	-14

* Observed WT fragment size is the mean fragment size of between 5 and 8 separate fragment analysis runs of a PCR product from NB4 cell line DNA, a known *CEBPA*-WT. Abbreviations: bp, base pairs; SD, standard deviation; D4, 5' fluorescent label; WT, wild-type.

WT and mutant peaks was determined by the instrument software. The mutant level was defined as the area under the mutant peak as a percentage of the total alleles, and was calculated using the following formula: area under mutant peak / (area under mutant peak + area under WT peak) x 100. The size of the insertion or deletion was calculated from the difference in fragment size between the WT and mutant peaks, with fragment size called by the instrument software.

5.2.2 Quantification of mutant level by restriction enzyme digestion

There were three patients where homozygous *CEBPA*-mutations were detected, however, none of these mutations, two missense and one indel, affected the overall length (patients 91, 101 and 102, Appendix Table 2). One case was selected for quantification, patient 102 containing the homozygous 962A>G change, so that the level of any remaining WT allele, which was not visible on the sequencing chromatogram, could be quantified. An assay was designed so that PCR amplification with a mismatch reverse primer created a restriction enzyme recognition site in WT, but not mutant, alleles, enabling separation of WT and mutant alleles by size. Amplicons of 243 bp were generated using primers D4-CEBPA/3F and CEBPA/G964T/R(MM), and 8µl of each PCR product was digested at 65°C for 3 hours in a reaction mix with 1µl *Tsp*509I and 1µl of manufacturer's buffer 1 (New England Biolabs, Hitchin, UK). After digestion, 2µl of the reaction was analysed as described above (section 5.2.1) on the CEQ 8000 Genetic Analysis System. WT alleles were digested to a 213 bp fragment and 962A>G mutant alleles were not digested giving a 243 bp fragment.

5.2.3 Analysis of full-length clones of the *CEBPA* coding sequence

Of the 59 *CEBPA*-double cases in this cohort, 13 were selected for analysis of the presence of biallelic mutations, and details of the mutations in these patients are given in Table 5.2.

5.2.3.1 Cloning of the *CEBPA* coding sequence

In each case, PCR products of 1201 bp containing the entire coding sequence of *CEBPA* were generated using genomic DNA with primers CEBPA/1F and CEBPA/3R and the standard BIOTAQ DNA polymerase reaction mix (see section 2.1.2), with 5% DMSO as an additive, in a total volume of 20µl. Cycling conditions were 35 cycles of 95°C for 60 seconds, 62°C for 60

Table 5.2 Details of 13 cloned *CEBPA*-double cases and technique used to screen clones

	Mutation 1[1]			Mutation 2			
Patient no.[2]	**DNA change**	**Predicted size change (bp)**	**Screening method**	**DNA change**	**Fragment**	**Predicted size change (bp)**	**Screening method**
12[3]	86_94delinsT	-8	Fragment analysis	943_945dup	3	+3	Fragment analysis
13[3]	98_116del	-19	Fragment analysis	977_978ins66nt	3	+66	Agarose gel
15	107_113del	-7	Fragment analysis	543C>A	2	None	RED
24[3]	126_132del	-7	Fragment analysis	934_936dup	3	+3	Fragment analysis
30	155dupT	+1	MM primer + RED	499dupG	2	+1	RED
37[3]	184_186delinsG	-2	RED	986_987ins105nt	3	+105	Agarose gel
39	191_194dup	+4	Fragment analysis	949_950insGTC	3	+3	Fragment analysis
44	230_233del	-4	Fragment analysis	917_934del	3	-18	Fragment analysis
47	230_231dup	+2	Fragment analysis spiked with WT	927_986dup	3	+60	Agarose gel
49	245_246insGTGTT	+5	Fragment analysis	909_923dup	3	+15	Fragment analysis
60	296_299dup	+4	Fragment analysis	934_936dup	3	+3	Fragment analysis
64	326_327insTA	+2	RED	934_936dup	3	+3	Fragment analysis
65	339_342del	-4	Fragment analysis	934_936dup	3	+3	Fragment analysis

[1]All mutations were in fragment 1. [2]Patient number relates to Appendix Table 2. [3]Cases were also heterozygous for one of the common polymorphisms: G/T SNP at nucleotide 690 (patients 12, 24 and 37) or H195_P196dup (patient 13). Abbreviations: ins, insertion; del, deletion; dup, duplication; nt, nucleotide; MM, mismatch; RED, restriction enzyme digestion; WT, wild-type.

seconds and 72°C for 90 seconds followed by a final extension at 72°C for 10 minutes, with a temperature ramping speed of 0.5°C/second. Products were checked on a 1.5% agarose gel stained with ethidium bromide. PCR products were cloned into the pCR2.1-TOPO vector using the TOPO TA Cloning Kit (Invitrogen Life Technologies, Paisley, UK), and the TOPO cloning reaction was used to transform One Shot MAX Efficiency DH5 α -T1 chemically competent *E. coli* (Invitrogen Life Technologies), as described in sections 2.1.8 and 2.1.9. Bacteria were spread on LB agar plates containing 100µg/ml carbenicillin, which were pre-coated with 40µl of a 40mg/ml X-gal solution to allow blue/white colony screening, as detailed in section 2.1.9, and incubated overnight at 37°C. At least 14 white colonies per patient were picked, seeded into 96-well plates and incubated overnight at 37°C, as described in section 2.1.10.

5.2.3.2 Detection of mutations

For each case, the two mutated *CEBPA* fragments were amplified by PCR from each clone and analysed for the presence or absence of the appropriate mutation. All PCRs, regardless of the downstream detection method, used BIOTAQ DNA Polymerase to amplify the required fragment from 1µl of the bacterial culture in a reaction mix supplemented with 5% DMSO, and were preceded with a hot start of 5 minutes at 95°C in order to ensure lysis of the bacteria. An appropriate technique was selected to detect each individual mutation and this information is given in Table 5.2.

Of the 23 different mutations in the 13 cases, 3 were large insertions of at least 60 bp in fragment 3. For these, amplification from primers CEBPA/3F and CEBPA/3R and agarose gel electrophoresis size separation of the PCR products easily identified the large insertions. A further 14 mutations caused size changes of between 3 and 19 bp, which were identifiable by fragment analysis of the appropriate fragment following the method given in section 5.2.1, apart from the initial hot start before cycling. One case had a 2 bp insertion, which was difficult to distinguish confidently from WT clones by fragment analysis, due to the variability in fragment size called by the instrument software (see section 5.3.2), an issue exacerbated by the very large amount of PCR product produced from bacterial clones versus that from genomic DNA. Therefore, PCR product from the bacterial clone was mixed 1:1 with PCR product from a known *CEBPA*-WT sample and the fragment analysis chromatograms of the "mixed" and "unmixed" clone were compared. A double peak in the mixed sample compared to the unmixed sample identified mutant clones. Of the remaining 5 mutations, 4 caused size changes of 1 or 2 bp and one was a point mutation, and for these,

PCR products were generated from the primers specified in Table 5.3. In 4 of these cases, direct digestion of PCR products with a restriction enzyme followed by agarose gel electrophoresis discriminated between WT and mutant clones. In the remaining case, a mismatch primer was designed to introduce a restriction site in PCR products from WT, but not mutant, clones. The restriction enzymes used and the resultant fragment sizes for WT or mutant clones are given in Table 5.3.

5.2.3.3 Detection of common polymorphisms

Of the 13 patients where full-length *CEBPA* products were cloned, 4 were also heterozygous for one of the two common polymorphisms, G/T at nt 690 or H195_P196dup. The polymorphism status of the clones from these patients was determined using either PCR with a mismatch primer and *Sgr*AI digest for the G/T at nt 690 or fragment analysis for the H195_P196dup. The methods used were as detailed in section 3.2.5, apart from the use of 1µl of bacterial culture as a template and an initial hot start step of 5 minutes at 95°C.

5.3 Results

By fragment analysis assays of all three *CEBPA* fragments, it was found that the size of a PCR product from a known *CEBPA* WT sample, as called by the instrument software, was smaller than the expected product size, with a reduction of 24 bp (4%), 16 bp (3%) and 14 bp (3%) for fragments 1, 2 and 3, respectively (Table 5.1). However, nucleotide sequencing of PCR products in both directions confirmed that the complete expected sequence was present. There are two other reports that have also found the observed WT fragment size to be smaller than the expected WT fragment size (Benthaus *et al*, 2008;Juhl-Christensen *et al*, 2008). In both of these studies, the *CEBPA* coding sequence was divided into four amplicons, with a reduction of between 12 bp (4%) and 18 bp (4%) in one study (Benthaus *et al*, 2008), and between 8 bp (2%) and 22 bp (4%) in the other (Juhl-Christensen *et al*, 2008). This phenomenon may be due to by the GC-rich nature of the *CEBPA* sequence causing secondary structure formation, which interferes with electrophoresis of fragments (Benthaus *et al*, 2008). However, the inclusion of a known WT control in all runs enabled easy identification of the WT fragment in fragment analysis of patient samples.

Table 5.3 Screening of full-length clones by restriction enzyme digestion

							Fragment sizes post-RED (bp)	
Patient No.[1]	**DNA Change**	**Fragment**	**Forward primer**	**Reverse primer**	**Amplicon size (bp)**	**Restriction enzyme**	**WT**	**MUT**
15	543C>A	2	CEBPA/2F2	CEBPA/2R	333	*Mse*I	333	70 + 263
30	155dupT	1	CEBPA/1F	CEBPA/G157A /R(MM)	215	*Dde*I	31 + 184	215
	499dupG	2	CEBPA/2F2	CEBPA/2R	333	*Bse*RI	31 + 302	333
37	184_186delinsG	1	CEBPA/1F	CEBPA/1R	443	*Bsp*EI	443	211 + 230
64	326_327insTA	1	CEBPA/1F	CEBPA/1R3	548	*Bfa*I	548	192 + 356

[1]Patient number relates to Appendix Table 2. Abbreviations: ins, insertion; del, deletion; dup, duplication; MM, mismatch; RED, restriction enzyme digestion; WT, wild-type; MUT, mutant; bp, base pairs.

Mutations in *CEBPA* causing length changes that were 2 bp or more were detected easily by fragment analysis of fluorescently labelled PCR products. However, single base pair insertions or deletions could not be distinguished from the WT peak. A number of changes to the protocol were tested in an attempt to solve this issue, for example new primers were designed to generate smaller PCR products, the fluorescent label was tested on the reverse rather than the forward primer, and products were electrophoresed at a lower voltage over a longer period on the CEQ 8000 Genetic Analysis System. However, the expected mutant peak could not be resolved from the WT fragment peak. Therefore, direct mutant level quantification by fragment analysis was not possible in 39 (23%) of the 166 *CEBPA* mutations detected in 107 patients.

5.3.1 Quantification of H195_P196dup polymorphism allele level by fragment analysis

To assess the use of fragment analysis for the quantification of the proportion of alleles with size changes in *CEBPA*, the level of the known polymorphism H195_P196dup was determined in 65 of the 68 (96%) patients in the cohort who were identified as heterozygotes for this change. PCR products could not be obtained in the remaining three cases. This 6 bp insertion was clearly identifiable by fragment analysis of fragment 2 (Figure 5.3 A), and although levels ranged between 39% and 64%, the median level of H195_P196dup in these cases was 50% ±8% (median ±2SD), with a very tight distribution around the median level (Figure 5.3 B). In addition, the size of the insertion was, to the nearest integer, the expected 6 bp for all samples tested, with a median difference in size between the WT and H195_P196dup peaks of 5.97 ±0.23 bp. These results were consistent with the level and insertion size expected for a heterozygous 6 bp insertion present in every cell, and gave confidence that fragment analysis was a suitable method for quantifying the level of size changes in *CEBPA*.

5.3.2 Quantification of *CEBPA* mutant level by fragment analysis

Quantification of the mutant level of both insertion and deletion mutations causing changes greater than or equal to 2 bp was straightforward for all *CEBPA* fragments. Examples of fragment analysis chromatograms for a 2 bp insertion in fragment 1 and a 12 bp deletion in fragment 2 are shown in Figure 5.4 A and B. These mutations each constituted nearly 50% of the total alleles (46% and 47%, respectively), however the technique also detected low level mutations, with one mutant quantified at 9% of the sample (Figure 5.4 C).

.

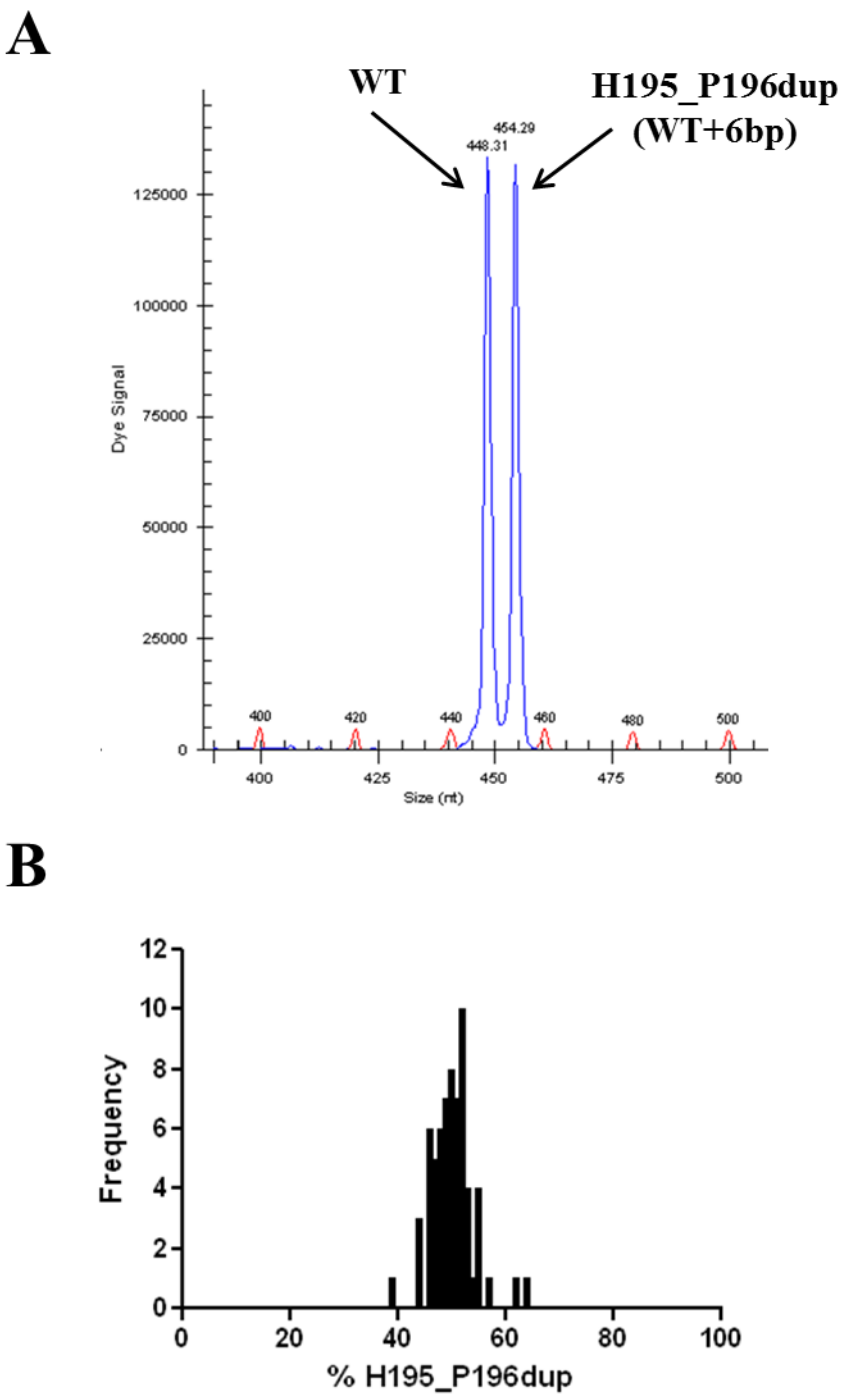


Figure 5.3 Quantification of H195_P196dup polymorphism allele level by fragment analysis. A. Fragment analysis chromatogram of *CEBPA* fragment 2 for an H195_P196dup heterozygote. Blue-labelled peaks represent PCR products and red peaks are a size standard ladder. The proportion of the total alleles with the H195_P196dup was 50% in this case. B. Distribution of the level of H195_P196dup alleles in 65 cases.

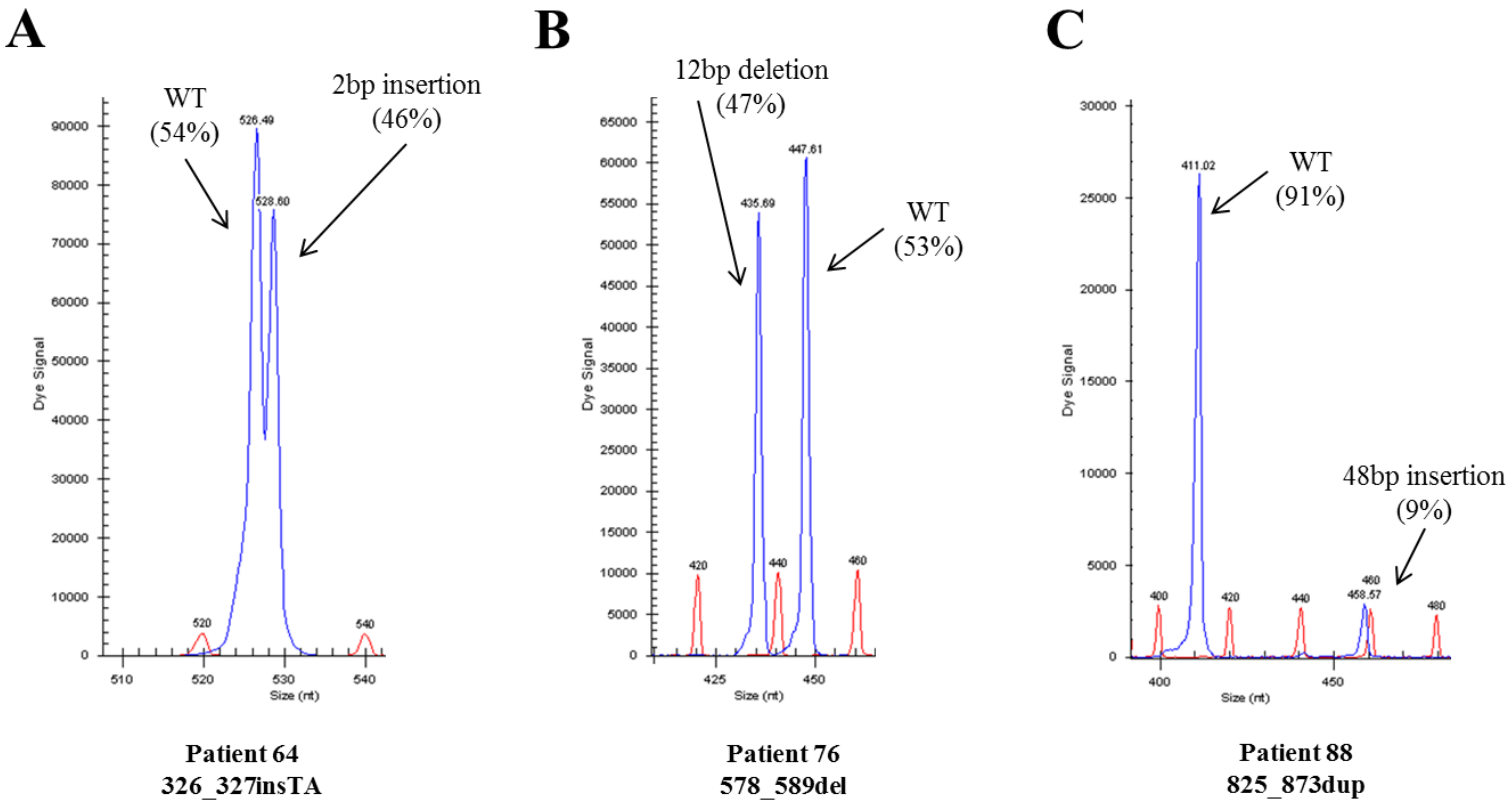


Figure 5.4 Detection and quantification of *CEBPA* mutant level by fragment analysis. Fragment analysis chromatograms of (A) Fragment 1 from patient 64 showing the expected 2 bp insertion, (B) Fragment 2 from patient 76 with a 12 bp deletion, (C) Fragment 3 from patient 88 showing a low level mutation. The size of the insertion by fragment analysis was 48 bp, the expected size from nucleotide sequencing was 49 bp, (D, E) CEBPA/3F* to CEBPA/G964T/R(MM) products from a WT control and patient 102 pre- (D) and post-*Tsp509*I digestion (E), which discriminated between alleles with or without the 962A>G mutation. Blue-labelled peaks represent PCR products and red peaks are a size standard ladder.

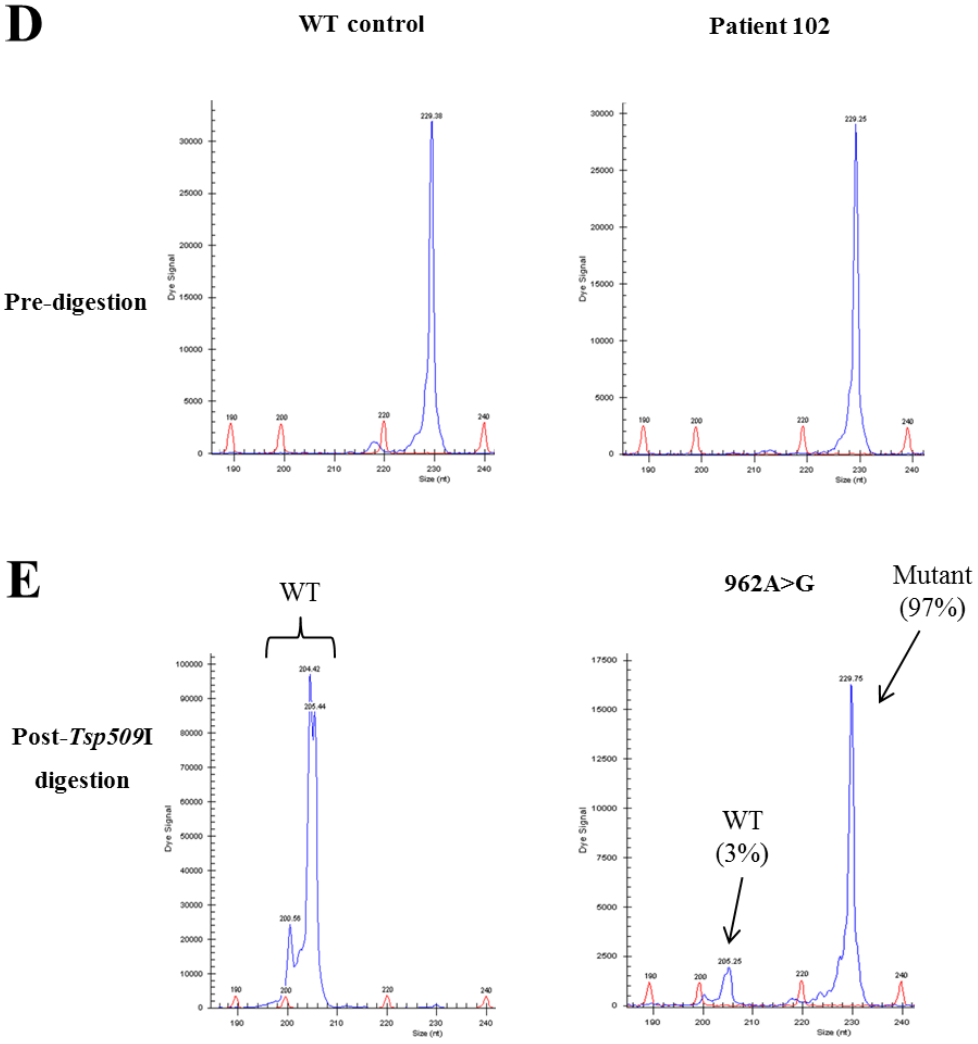


Figure 5.4 continued

Of the 100 mutations causing a 2 bp or more size change that were analysed, it was found in 19 (19%) that the size of the insertion or deletion was incorrectly called by the instrument software, by between 1 and 4 bp, when compared to the size change expected from nucleotide sequencing This is likely to be a mutation-specific effect due to the GC-rich sequence causing secondary structure, which may interfere with the capillary electrophoresis.

Of the 166 *CEBPA* mutations detected in 107 patients, 23 (14%) could not be directly detected by fragment analysis as they were either point mutations (n = 21) or indels that did not change the overall sequence length (n = 2). Each of the 20 different mutations would have required an individual assay to be designed using a specific digest or mismatch primers for amplification followed by a specific digest. Therefore a specific assay was designed for only one of these cases with a missense mutation (patient 102, 962A>G) only identified by dHPLC analysis of patient PCR product mixed with that from a known WT sample, and no WT allele was visible on the sequencing chromatogram. It was again found that the observed fragment sizes, for both digested and undigested alleles, were smaller than those expected. However, comparison of the digestion pattern for a known *CEBPA*-WT with that from patient 102 (Figure 5.4 D and E) allowed quantification of the mutation as 97%. This is consistent with either a homozygous or hemizygous mutation, and while UPD has not been confirmed in this case it is known to be a common cause of homozygous *CEBPA* mutations (Fitzgibbon *et al*, 2005;Wouters *et al*, 2007c), and therefore classification of this change as homozygous was appropriate.

5.3.3 *CEBPA* mutant level

In total, 101 mutations from 75 *CEBPA*-mutant cases (22 *CEBPA*-single, 53 *CEBPA*-double) were quantified and the median mutant level was 44% (range 9% to 97%). The distribution of *CEBPA* mutant level is shown in Figure 5.5 A. Quantification of the H195_P196dup polymorphism had allowed an approximation of the technical limits of fragment analysis when quantifying the level of a heterozygous size change known to be present in all cells. If the range of levels set by quantification of the H195_P196dup polymorphism was applied to the *CEBPA* mutations, 77 mutations (76%) from 58 patients had a level of between 39% and 64%, consistent with the presence of a heterozygous mutation in all, or at least most, cells. While nearly a quarter of mutations quantified (23%) were less than the 39% lower limit, the mutant level was less than 25% (i.e. consistent with a heterozygous mutation in less than half of cells) for just 5 mutations from 5 different patients (3 *CEBPA*-single, 2 *CEBPA*-double). One of these cases (*CEBPA*-mutant level 13%) also had an *NPM1* mutation

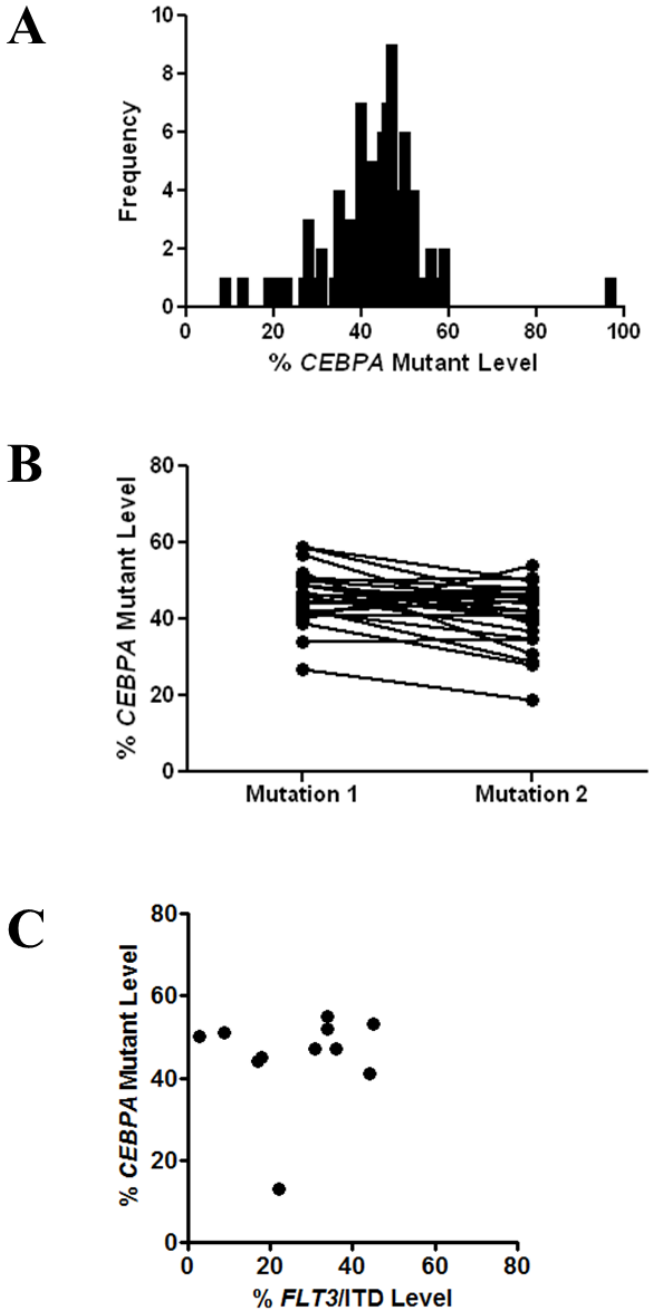


Figure 5.5 *CEBPA* mutant level

A. Distribution of *CEBPA* mutant level of 101 mutations from 75 patients.

B. Paired mutant levels in 26 *CEBPA*-double cases.

C. Comparison of *CEBPA* and *FLT3*/ITD mutant level in 11 cases positive for both mutations. For the 5 *CEBPA*-double cases with both mutations quantified, the mean of the two levels is plotted.

of 49%, suggesting that the *CEBPA* mutation was a secondary event. Additional markers were not available in the other cases.

In 26 *CEBPA*-double cases it was possible to quantify the level of both mutations (Figure 5.5 B). All patients had at least one mutation constituting more than 25% of alleles, and the paired levels were highly correlated (P = .0002, paired t-test), with an average difference of 7% (range 0 to 21%) in the level of the different mutations from the same patient. This suggested that both mutations were present in the same cell. At least one *CEBPA* mutation was quantified in 11 FLT3/ITD-mutant patients. Nearly all of the *CEBPA* mutant levels in these cases were consistent with the presence of a heterozygous mutation in most cells, with the level in 10 cases (91%) quantified as between 41% and 53%. However, the level of the *FLT3*/ITD varied widely (3% to 45%, Figure 5.5 C), suggesting that acquisition of a *CEBPA* mutation could precede a *FLT3*/ITD.

5.3.4 Confirmation of biallelic mutations by cloning

There were 13 *CEBPA*-double cases where the entire *CEBPA* coding sequence was cloned and between 14 and 33 clones per patient were screened for the presence of each mutation. Overall, most clones (87%) had one mutation, 9% were WT and 4% had both mutations (Table 5.4). This indicated that, in these patients, the two mutations predominantly occurred on different alleles.

Four of these cases were also heterozygous for one of the two common polymorphisms, and in these patients the clones with one mutation were additionally screened to determine the polymorphism status of the cloned allele. Results were obtained on 77 clones and this data is given in Table 5.5. In patients 12, 13 and 37, there was an association between the polymorphism allele and the *CEBPA* mutation, with the N-terminal and C-terminal mutations associated with the same polymorphism allele in every clone. Moreover, the different mutations were associated with different polymorphism alleles. For example, of the 15 clones analysed from patient 37, a heterozygote for the G/T SNP at nt 690, all 10 clones with the N-terminal mutation had a G at nt 690, while all 5 with the C-terminal mutation had a T at nt 690. In the remaining case (patient 24), 24 (92%) of the 26 clones with one mutation screened also followed this pattern, however 2 clones were found where the mutation was on the opposite polymorphism allele from the other clones with the same mutation. It was also noted that, for each of these 4 cases, there was a preponderance

Table 5.4 Screening of full-length clones in 13 *CEBPA*-double cases.

Patient No.[1]	Mutation 1	% Mutation 1	Mutation 2	% Mutation 2	Number of clones screened	Clones with one mutation, n (% of clones)	Clones with both mutations, n (% of clones)	Wild-type clones, n (% of clones)
12[2]	86_94delinsT	59	943_945dup	50	19	14 (74)	3 (15)	2 (11)
13[2]	98_116del	49	977_978ins66nt	40	32	26 (82)	2 (6)	4 (12)
15	107_113del	49	543C>A	N/A	27	27 (100)	0	0
24[2]	126_132del	59	934_936dup	45	33	26 (79)	2 (6)	5 (15)
30	155dupT	N/A	499dupG	N/A	19	15 (79)	1 (5)	3 (16)
37[2]	184_186delinsG	52	986_987ins105nt	31	17	15 (88)	0	2 (12)
39	191_194dup	43	949_950insGTC	29	30	27 (90)	2 (7)	1 (3)
44	230_233del	44	917_934del	47	19	15 (79)	0	4 (21)
47	230_231dup	39	927_986dup	28	14	14 (100)	0	0
49	245_246insGTGTT	50	909_923dup	46	19	19 (100)	0	0
60	296_299dup	45	934_936dup	42	20	17 (85)	0	3 (15)
64	326_327insTA	46	934_936dup	48	15	12 (80)	1 (7)	2 (13)
65	339_342del	51	934_936dup	48	21	20 (95)	1 (5)	0
				TOTAL	**285**	**247 (87)**	**12 (4)**	**26 (9)**

[1]Patient number relates to Appendix Table 2. [2]Cases were also heterozygous for one of the common polymorphisms: G/T SNP at nucleotide 690 (patients 12, 24 and 37) or H195_P196dup (patient 13). Abbreviations: ins, insertion; del, deletion; dup, duplication; nt, nucleotide; N/A, not available.

Table 5.5 Polymorphism analysis in full-length clones with only one mutation in four *CEBPA*-double cases.

		Mutant level, %			No. of clones with N-terminal mutation			No. of clones with C-terminal mutation		
Patient no.[1]	**Polymorphism**	**N-terminal mutation**	**C-terminal mutation**	**Total no. of clones screened**	**Total no.**	**Major allele[2]**	**Minor allele[2]**	**Total no.**	**Major allele[2]**	**Minor allele[2]**
12	G/T nt 690	59	50	14	**12**	12	0	**2**	0	2
13	H195_P196dup	49	40	22[3]	**19**	19	0	**3**	0	3
24	G/T nt 690	59	45	26	**17**	16	1	**9**	1	8
37	G/T nt 690	52	31	15	**10**	10	0	**5**	0	5
			TOTAL	**77**						

[1]Patient number relates to Appendix Table 2. [2]Major allele denotes G at nt 690 or non-H195_P196dup, minor allele denotes T at nt 690 or presence of H195_P196dup as appropriate. [3]The polymorphism status was not determined for 4 clones with one mutation from patient 13. Abbreviations: nt, nucleotide; WT, wild-type; dup, duplication.

of clones with N-terminal mutations (58 of 77, 75%) as compared to clones with C-terminal mutations (19 of 77, 25%) (Table 5.5). The cause of this imbalance is unknown but may, in part, be the result of both the relatively small number of alleles studied per patient and a higher mutant level of the N-terminal compared to the C-terminal mutation, of between 9% and 21%, in each case (Table 5.5).

5.4 Discussion

The data presented in this chapter shows quantification of *CEBPA* mutant level in the same cohort of *CEBPA*-mutant patients determined in chapter 3 and confirmation of biallelic mutations in selected *CEBPA*-double cases by cloning of the entire coding sequence. Together, this information can be used to examine the allelic composition of the population of leukaemic cells at disease presentation.

The technique selected to quantify mutant level was fragment analysis of fluorescently labelled PCR products as there have been several reports showing that this method could successfully detect *CEBPA* length mutations in a sensitive manner (Lin *et al*, 2006;Juhl-Christensen *et al*, 2008;Benthaus *et al*, 2008). Indeed, when used in this study, fragment analysis was found to be a straightforward technique for size changes of two base pairs or more, with detection and quantification of mutations of this type even when the mutant constituted less than 10% of the total alleles (Figure 5.4 C). Additionally, quantification of the known polymorphism H195_P196dup in 65 cases gave a median allele level of 50% (range 39% to 64%), with a tight distribution around this figure (Figure 5.3 B), which is the level expected for a heterozygous polymorphism present in all cells. These results indicated that the use of fragment analysis for quantification in this context was appropriate. However, unlike previous reports (Lin *et al*, 2006;Juhl-Christensen *et al*, 2008;Benthaus *et al*, 2008), insertions or deletions of just a single base pair could not be detected in this study, despite attempts at optimisation of the method. This difference may be because an Applied Biosystems Genetic Analyzer was used in each of those reports, rather than the CEQ 8000 Genetic Analysis System (Beckman Coulter) that was available for this study. However, both instruments had the same issue with regards to the reduced WT fragment sizes called by the instrument software versus those expected (Table 5.1) (Benthaus *et al*, 2008;Juhl-Christensen *et al*, 2008). In addition, in this study the size of the insertion or deletion, as calculated from the called fragment sizes, was not accurately determined in nearly one fifth of mutations, when compared with the nucleotide sequence. Both of these observations are

probably due to the GC-rich nature of the *CEBPA* gene, causing secondary structure, which affected how the PCR product electrophoresed. This secondary structure may not have been completely removed by the denaturation step performed by the instrument before the electrophoresis, which involved heating the PCR product in a sample loading solution containing de-ionised formamide, a stabiliser of single-stranded DNA molecules. However, neither issue appeared to significantly impact upon quantification of mutant level, as indicated by analysis of H195_P196dup polymorphism alleles.

Of the 166 *CEBPA* mutations detected in 107 patients by WAVE analysis, most (n = 104, 63%) caused a size change of 2 bp or more and could be quantified directly by the selected method. A further 39 mutations (23%) caused 1 bp length alterations and 23 (14%) were point mutations or indels that did not affect overall length. Therefore, 37% of the known mutations could not be directly quantified by fragment analysis in this study. In one particular case with a homozygous point mutation, a mismatch primer was used for amplification followed by allele-specific restriction enzyme digestion of WT, but not mutant, alleles allowing separation by size. Overall, 101 mutations from 75 patients were quantified, which was considered sufficient to explore the distribution of *CEBPA* mutant levels in AML patients.

The median mutant level of the 101 mutations quantified was 44% (range 9% to 97%) (Figure 5.5 A). Approximately three quarters of the mutations quantified were of a level consistent with that of the known H195_P196dup polymorphism and were suggestive of the presence of a heterozygous mutation in most cells in most cases. There is only one other published brief report on *CEBPA* mutant level, in which it was stated that the mutant level of 8 *CEBPA*-mutant cases ranged between 33% and 60% (Pabst *et al*, 2001b), which is consistent with the data from this, much larger, study. The *CEBPA* mutant levels in the present cohort are therefore in agreement with the concept of *CEBPA* mutations as early events in leukaemogenesis, as suggested by their relative stability in disease progression and occurrence as germline mutations in familial AML (chapter 3). In further support of this is the finding that in 10 of the 11 *CEBPA*-mutant cases with a *FLT3*/ITD mutation where quantification was performed, the *CEBPA* level was more than 40%, while the *FLT3*/ITD level was variable, even down to 3% (Figure 5.5 C), suggesting that, in at least some cases, the *FLT3*/ITD was acquired after the *CEBPA* mutation. The *CEBPA* mutant level was low (less than 25%) in a small number of cases, which could be due to the presence of either normal or leukaemic *CEBPA*-WT cells in the sample. Blast counts were not available for this study, however, one of these cases also had an *NPM1* mutation of 49%, indicating a *CEBPA* mutation is not always an early event.

In the cohort described in chapter 3, there were 56 patients where two different *CEBPA* mutations were detected, accounting for 95% of those classified as *CEBPA*-double. These patients had a favourable prognosis provided that a *FLT3*/ITD mutation was not also present (chapter 4). The level of both mutations could be quantified in 26 of these cases, and in all cases one, or both, of the mutations constituted at least a quarter of total alleles, and the paired levels were highly correlated (P = .0002, paired t-test, Figure 5.5B). There are at least three different allelic and cellular distributions possible for two intermediate-level mutations in the same patient, including different mutations in different cells and different mutations in the same cell, either on the same, or different, alleles (Figure 5.2), and the paired mutant levels were suggestive of the presence of both mutations in the same cells. To investigate the allelic distribution, the entire coding sequence was therefore cloned in 13 of the 56 (23%) *CEBPA*-double cases with two different mutations, and clones were screened for the presence of each mutation. The presence of biallelic mutations was confirmed in all 13 cases. This is in line with evidence from other reports which confirmed the presence of biallelic *CEBPA* mutations in 69 (96%) of 72 cases with two different mutations (Preudhomme *et al*, 2002;Barjesteh van Waalwijk van Doorn-Khosrovani *et al*, 2003;Frohling *et al*, 2004;Lin *et al*, 2005;Shih *et al*, 2006;Dufour *et al*, 2010). The vast majority of clones (87%) had just one mutation, consistent with the mutations being on different alleles, although a small number of clones (4%) from 7 different cases were found with both mutations. This has been described in other reports (Lin *et al*, 2005;Shih *et al*, 2006), but it is not clear whether this is a true result or a technical error. It is possible that these double-mutant clones are the result of a PCR artefact caused by primers bound to partially annealed, mismatched templates leading to strand displacement and the formation of chimeric PCR products, as has been proposed for the *JAK2* gene (Beer *et al*, 2010). However, it is also possible that a recombination event had occurred within the *CEBPA* gene in a small subclone of the leukaemic cells so that both mutations were on the same allele. Evidence from mouse models suggests that one affected *CEBPA* allele is not sufficient for leukaemogenesis (Kirstetter *et al*, 2008;Bereshchenko *et al*, 2009), and therefore cells with both mutations on the same allele may not have a selection advantage and would remain a low proportion of the population. Despite this uncertainty regarding the double-mutant clones, it is clear that, in all 13 cases analysed, single-mutant clones predominated. This indicates that the different mutations are on different alleles, but does not rule out the possibility that the mutations could be in different cells (Figure 5.2, option A). Therefore, the polymorphism status of clones from 4 of the cloned *CEBPA*-double cases with an informative polymorphism was investigated. This found an association between the polymorphism allele and the *CEBPA* mutation, with the N-terminal and C-terminal

mutations associated with different polymorphism alleles in clones from the same patient (Table 5.5). Of the 77 clones examined, only 2 clones from one patient did not follow this pattern, and the mutation was on the opposite polymorphism allele from other clones with the same mutation. Similarly to the clones with both mutations, it is possible that this was caused by a recombination event in a cell or is a PCR artefact. In either case, this was clearly a rare event (3%) and the dominant pattern found was that the different mutations were each on a different polymorphism allele. Given that mutation levels in these 4 patients were all between 31% and 59%, the polymorphism status of the clones showed it was most likely that the different mutations were in the same cell (Figure 5.2, option B.ii).

The data presented in this chapter has shown that *CEBPA* mutant level is consistent with a heterozygous mutation present in most cells and that in *CEBPA*-double cases both alleles are mutated, with loss of WT *CEBPA*. Acquisition of a *CEBPA* mutation(s) is, therefore, probably an early event in leukaemogenesis in most, but not all, *CEBPA*-mutant cases. In addition, the data indicates that biallelic *CEBPA* mutations, and therefore loss of normal *CEBPA*, are associated with a favourable clinical outcome in *FLT3*/ITD-negative patients. Unlike *CEBPA*-double patients, the prognosis of *CEBPA*-single patients was not found to be significantly different from *CEBPA*-WT cases in the present cohort (chapter 4), and it is, therefore, important to correctly classify *CEBPA*-mutant patients as either *CEBPA*-single or *CEBPA*-double. This can be challenging if the type of mutation(s) detected does not fall into either the "classic" N-terminal or C-terminal mutation categories, and exploring these non-classical mutations is the subject of chapter 6.

CHAPTER 6: TRANSACTIVATION POTENTIAL OF NON-CLASSICAL CEBPA MUTATIONS

6.1 Introduction

Mutations in *CEBPA* in AML patients can be detected across the whole gene and a variety of alterations have been identified, including missense, nonsense and in-frame or out-of-frame size changes. The data presented in chapter 4 shows that *CEBPA*-double patients lacking a *FLT3*/ITD have a favourable outcome compared to *CEBPA*-single or *CEBPA*-WT cases. However, if this information is to be used to aid clinical management decisions, it will be necessary to correctly classify patients as *CEBPA*-WT, *CEBPA*-single or *CEBPA*-double, requiring decisions as to whether individual DNA sequence alterations are functionally relevant mutations or non-pathogenic changes, whether germline or acquired. Whilst a role in leukaemogenesis has been established for the classical *CEBPA* mutations, this is less clear for the non-classical alterations and further investigation is required. This is of particular importance for *CEBPA*-double patients with a classical mutation in combination with a non-classical sequence change, who may be more correctly classified as *CEBPA*-single if the latter alteration was non-pathogenic.

6.1.1 Impact of *CEBPA* mutations on protein function

As introduced in chapter 3 (see section 3.1.2), mutations in *CEBPA* in AML patients were first reported by Pabst and colleagues in 2001, and they performed *in vitro* assays to explore the impact of these patient-derived mutations on C/EBPα protein function (Pabst *et al*, 2001b). Since that report, there have been several comprehensive functional studies of *CEBPA* mutations, both *in vitro* and *in vivo*, and these data are introduced briefly below in sections 6.1.1.1 and 6.1.1.2. The focus has been on modelling the effects of out-of-frame insertions/deletions or nonsense mutations in the N-terminus (classical N-terminal mutations) and/or in-frame insertions/deletions in the DNA binding (DBD) or leucine zipper (LZD) domains (classical C-terminal mutations), as these are the most common types of mutations identified in patients. Indeed, 121 (74%) of 163 mutations in the present cohort were heterozygous classical N- or C-terminal mutations. However, there remains a significant minority of mutations that do not fall into either of these categories.

6.1.1.1 *In vitro* studies of classical N- and C-terminal *CEBPA* mutations

The major observation made with regard to classical N-terminal mutations has been that they cause loss of p42 isoform expression and a concomitant increase in p30 C/EBPα (Pabst *et al*, 2001b), an isoform that lacks the first transactivation domain (TAD1), but retains the DBD and LZD. Porse *et al* (2001) found that full-length p42 isoform expression was associated with an arrest in mitotic growth, thought to be mediated by repression of E2F-dependent transcription, which required the amino terminus of C/EBPα (Porse *et al*, 2001). In line with this, the p30 isoform has been found to lack both the ability to inhibit cell growth and to induce terminal differentiation in adipocyte and haemopoietic cell lines (Lin *et al*, 1993;D'Alo' *et al*, 2003;Kato *et al*, 2011). However, these findings have not been confirmed in all studies, which may be related to the experimental system used. For instance, expression of the p30 isoform in a murine cell line had an anti-mitotic effect (Cleaves *et al*, 2004), but expression of an N-terminal mutant in human haemopoietic progenitors did not reduce cell proliferation (Quintana-Bustamante *et al*, 2012).

Whilst both p30 and p42 C/EBPα regulate a number of common genes, there are several targets restricted to only one isoform, for example MPP11, p84N5 and SMYD2 are all downregulated by p30 isoform expression but not p42 (Wang *et al*, 2007) and p42, but not p30, upregulates transcription of the G-CSF receptor (D'Alo' *et al*, 2003). Analysis of isoform-specific targets has also revealed mechanisms of cross-talk between p30 and p42. For instance, the expression of the p30 but not p42 isoform in K562 and primary AML cells was associated with upregulated Ubc9, leading to sumoylation of p42 C/EBPα, which impaired the transcriptional activity of the p42 protein (Geletu *et al*, 2007).

The p30 isoform is expressed together with p42 C/EBPα in normal cells and changes in the p42/p30 isoform ratio have been observed in cell processes, such as adipocyte differentiation (Lin *et al*, 1993;Calkhoven *et al*, 2000). It has been hypothesised that regulation of the p42/p30 ratio enables cells to respond to extracellular conditions. In nutrient- or growth factor-rich conditions, p30 isoform expression is increased, probably mediated by increased eukaryotic translation initiation factors such as eIF2α and eIF4E, thereby promoting cell proliferation and delaying terminal differentiation (Calkhoven *et al*, 2000;Nerlov, 2004). In AML with N-terminal *CEBPA* mutations, the increase in p30 and concomitant decrease in p42 isoform expression is hypothesised to cause a loss of regulation of the p42/p30 ratio, thereby increasing cell proliferation and reducing differentiation (Nerlov, 2004).

Wild-type p42, hereafter called WT, can bind to target DNA sequences and activate transcription of genes, such as those coding for the G-CSF receptor and neutrophil elastase. Therefore, the functional impact of classical mutations on this transactivation (TA) activity has been explored by several groups using *in vitro* luciferase reporter assays, where human or murine cell lines have been transiently transfected to exogenously express C/EBPα (Pabst *et al*, 2001b;Gombart *et al*, 2002;Cleaves *et al*, 2004;Pabst *et al*, 2009;Kato *et al*, 2011). In assays transfecting a single C/EBPα construct into cells, C/EBPα with either a classical N-terminal mutant or the p30 isoform had lower TA activity than the WT control (Pabst *et al*, 2001b;Gombart *et al*, 2002;Cleaves *et al*, 2004;Kato *et al*, 2011). This reduction was significant in all reports, and activity was quantified at between 21% and 16% of the WT activity in one study (Pabst *et al*, 2001b) and reduced from 40-fold activation for WT over vector alone to 5-fold for the p30 isoform in another (Cleaves *et al*, 2004). N-terminal mutants or the p30 isoform have also been co-expressed with a WT C/EBPα construct in luciferase reporter assays and reduced TA activity was still observed. In one such study, TA activity was 28% of the WT activity (Pabst *et al*, 2009), and in another TA activity was similar for an N-terminal mutant construct or a p30 isoform construct when expressed either singly or co-expressed with WT (Pabst *et al*, 2001b). Indeed, the p30 isoform has been described as having a dominant negative effect over WT in several reports (Pabst *et al*, 2001b;Gombart *et al*, 2002;Cleaves *et al*, 2004;Pabst *et al*, 2009). The mechanism(s) for the reduction in TA activity seen for the p30 isoform are unclear and may be related to a lack of TAD1 and/or a loss of DNA binding. When the ability to bind a target DNA sequence has been assessed by electrophoretic mobility shift assays (EMSAs), classical N-terminal mutants or the p30 isoform have variously been reported to have slightly reduced (Kato *et al*, 2011), significantly reduced (7-fold reduction compared with WT) (Pabst *et al*, 2001b) or nearly complete absence of DNA binding ability (D'Alo' *et al*, 2003). There is also some evidence that the p30 isoform may differentially bind to specific promoters, losing affinity for the G-CSF receptor promoter but retaining it for the neutrophil elastase and PU.1 gene promoters (Cleaves *et al*, 2004).

Wild-type C/EBPα promotes myeloid differentiation and cell-cycle arrest but, in *in vitro* studies, expression of classical N-terminal mutant or the p30 isoform in human and murine cell lines has been shown to cause a loss of myeloid differentiation (Pabst *et al*, 2001b;D'Alo' *et al*, 2003;Cleaves *et al*, 2004;Kato *et al*, 2011). However, exogenous overexpression of classical N-terminal mutant C/EBPα in primary human haemopoietic progenitor cells has been reported to cause a loss of myeloid differentiation (Schwieger *et al*, 2004), as well as to favour it at the expense of erythroid differentiation (Quintana-Bustamante *et al*, 2012). Similarly, a consensus has not been reached on the impact of a

classical N-terminal mutant or the p30 isoform on cell growth, cell cycling and self-renewal potential, as tested by colony formation assays, re-plating efficiency and cell-cycle analysis of transduced human or murine cell lines and primary haemopoietic cells. This variation may be due to different experimental conditions, in particular differences between human and murine systems (D'Alo' *et al*, 2003;Cleaves *et al*, 2004;Schwieger *et al*, 2004;Kato *et al*, 2011;Quintana-Bustamante *et al*, 2012).

Classical C-terminal mutations have also been shown to strikingly reduce TA activity as compared to WT C/EBPα in luciferase reporter assays when individually transfected into cell lines, with 2% and less than 20% of the WT activity reported for several different mutations (Pabst *et al*, 2001b;Gombart *et al*, 2002;Kato *et al*, 2011). In assays co-expressing a C-terminal mutant with WT C/EBPα, 36% of the WT TA activity was reported in one study (Pabst *et al*, 2009), but no impact on activity was observed in two others (Gombart *et al*, 2002;Kato *et al*, 2011), and a dominant negative activity for C-terminal mutant over WT C/EBPα has not been identified (Pabst *et al*, 2001b). The lack of TA activity in the single construct transfection assays is associated with nearly complete loss of DNA binding for all classical C-terminal mutants, as tested by EMSA (Pabst *et al*, 2001b;Gombart *et al*, 2002;Kato *et al*, 2011). This is consistent with the finding that expression of a C-terminal mutant in both a murine cell line and human haemopoietic progenitor cells caused a loss of myeloid differentiation (Kato *et al*, 2011;Quintana-Bustamante *et al*, 2012). Both classical N- and C-terminal mutants are reported to retain their nuclear subcellular location (Pabst *et al*, 2001b;Gombart *et al*, 2002), however the p30 isoform localised on the chromosome during the mitotic phase, whilst a C-terminal mutant C/EBPα did not (Kato *et al*, 2011).

6.1.1.2 *In vivo* studies of classical N- and C-terminal *CEBPA* mutations

A number of different mouse models have been developed to explore the impact of classical N- and C-terminal C/EBPα mutants *in vivo*. The first was a knock-in model that combined the p30 isoform, called the L allele, and a classical C-terminal mutation (K313dup, the K allele), which were expressed from the endogenous *cebpa* promoter (Bereshchenko *et al*, 2009). In this study, due to the lethality associated with the mutation of both *cebpa* alleles, foetal liver cells from K/K, K/L and L/L mice were competitively transplanted into lethally irradiated syngeneic mice (Kirstetter *et al*, 2008). All mice transplanted with K/K, K/L or L/L cells developed leukaemia, however those receiving K/L cells had significantly shorter survival than those receiving K/K or L/L cells. Moreover, K/L and L/L tumours were nearly always granulocytic, and fractionation of the leukaemic cells by immunophenotype and transplantation into secondary recipients demonstrated that the leukaemia initiating cell(s)

was found in the myeloid progenitor compartment (Kirstetter *et al*, 2008;Bereshchenko *et al*, 2009). By contrast, approximately only one quarter of K/K tumours could be classified as myeloid, while the remainder were characterised by an immature erythroid phenotype. However, the frequency of phenotypic HSCs was increased in mice that received K/K or K/L cells, but not L/L cells. The authors therefore hypothesised that the p30 isoform and the C-terminal mutant had different roles in leukaemogenesis, with the C-terminal mutation causing pre-malignant haemopoietic expansion whilst the p30 isoform maintained residual myeloid commitment (Bereshchenko *et al*, 2009).

In a second model, Kato and colleagues transduced murine BM mononuclear cells with retroviral C/EBPα expression construct(s) to induce overexpression of a C-terminal mutant (K304_R323dup), an N-terminal mutant (T60fs), or both, and these cells were then transplanted into irradiated syngeneic mice (Kato *et al*, 2011). Nearly all mice receiving cells transduced with the C-terminal mutant construct developed AML within 4 to 12 months after transplantation, however most mice receiving an N-terminal mutant remained healthy for at least a year, with a minority developing a B-cell acute lymphoid leukaemia. As with the knock-in model, co-expression of the N- and C-terminal mutants shortened the latency of AML development to 3 to 5 months (Kato *et al*, 2011). However, contrasting results were reported by Quintana-Bustamante *et al* (2012), who transduced human haemopoietic progenitor cells with a lentiviral C/EBPα expression construct for WT, an N-terminal mutant, a C-terminal mutant or both the N- and C-terminal mutant constructs (Quintana-Bustamante *et al*, 2012). Whilst initial engraftment was observed for all constructs, by 8 weeks all or almost all of the graft had been lost. These data indicated that all constructs caused loss of re-populating capacity and suggested that further events were required for the transformation of normal human haemopoietic stem/progenitors.

6.1.1.3 Functional impact of non-classical *CEBPA* mutations

Predicting the impact on protein function of a specific DNA sequence alteration can be difficult. As indicated above, functional studies of mutated C/EBPα have largely focussed on the two most common types of mutations, and the evidence suggests that all p30-causing *CEBPA* mutations and all classical C-terminal mutations so far evaluated generate a protein lacking some of the functional activities of normal p42 protein. However, a significant proportion of the mutations in the present cohort (42 of 163, 26%) were identified as neither of the classical mutations, and only a few individual non-classical alterations identified in patients have been tested in *in vitro* assays. Pabst *et al* (2001b) assayed three non-classical mutations for their impact on TA activity in a luciferase reporter assay, an N-terminal

missense mutation that did not cause an increase in the p30 isoform expression (H84L), a frame-shift in the middle of the gene that caused a complete loss of the C-terminus for both the p30 and the p42 isoforms (R165fs), and a frame shift in the C-terminus that slightly truncated both isoforms (V351fs). These mutants varied widely in their TA activity as compared with WT, with no difference seen for H84L, a reduction in activity for V351fs C/EBPα, and a large increase in activity for the R165fs mutant (Pabst *et al*, 2001b). The findings that H84L mutant C/EBPα did not impair TA activity and the V351fs mutant decreased activity are predictable given the known functional consequences of the classical mutations. However, the observation that R165fs mutant C/EBPα strongly increased TA activity is unexpected and was not explained.

In the patient cohort described in chapter 3, 22 (13%) of the 163 mutations detected were nonsense mutations or frame-shift/truncating insertions/deletions after the p30 isoform translation initiation site (15 in *CEBPA*-single and 7 in *CEBPA*-double cases). These are predicted to cause loss of some or all of the C-terminal DBD and LZD and could be expected to cause a loss of C/EBPα TA activity similar to that of classical C-terminal mutants. In addition, in the middle region of *CEBPA* there were six missense mutations, of which five were in *CEBPA*-single cases, and three in-frame insertions/deletions, all in *CEBPA*-single cases. The functional significance of these mutations is not clear. Unless this region is a "hotspot" for random mutations it is likely that their presence causes a selection advantage. In theory they could, by unknown mechanisms, lead to reduced TA activity and may act as a classical N-terminal mutation, with dominant-negative activity over the WT protein. Alternatively, they might not affect TA activity, as determined in *in vitro* assays, and have a different role in leukaemia pathogenesis. Therefore, to further explore the role of these atypical mutations an assay to determine the TA activity of C/EBPα mutants was developed.

6.1.2 Quantifying C/EBPα transactivation activity by luciferase reporter assay

As described above, luciferase reporter assays have been employed in several different reports as a technique for determining the TA activity of C/EBPα. In these assays, a luciferase reporter construct is generated that contains copies of a C/EBPα target sequence, such as the binding sequence in the *CSF3R* gene promoter (G-CSF receptor), upstream of a luciferase gene, and this is then co-transfected into cells together with a construct containing the required *CEBPA* sequence (Figure 6.1). C/EBPα that can recognise the target sequence in the reporter construct and bind to the DNA as a dimer activates transcription of the luciferase gene (Figure 6.1 A). C/EBPα variants that are deficient in DNA binding and/or

transcriptional activation activity will result in reduced levels of luciferase transcription (Figure 6.1 B and C). The amount of luciferase generated can then be assessed by cell lysis, addition of the luciferase substrate luciferin, and measurement of the luminescence in a luminometer. This system can also be utilised to explore the interaction between different C/EBPα proteins by co-transfection of more than one *CEBPA* construct. This is important because patients with *CEBPA* mutations may co-express either mutant and WT C/EBPα (*CEBPA*-single) or two different mutant proteins (*CEBPA*-double).

This chapter presents the results of an investigation into the impact of non-classical *CEBPA* mutations on C/EBPα TA activity using a luciferase reporter assay.

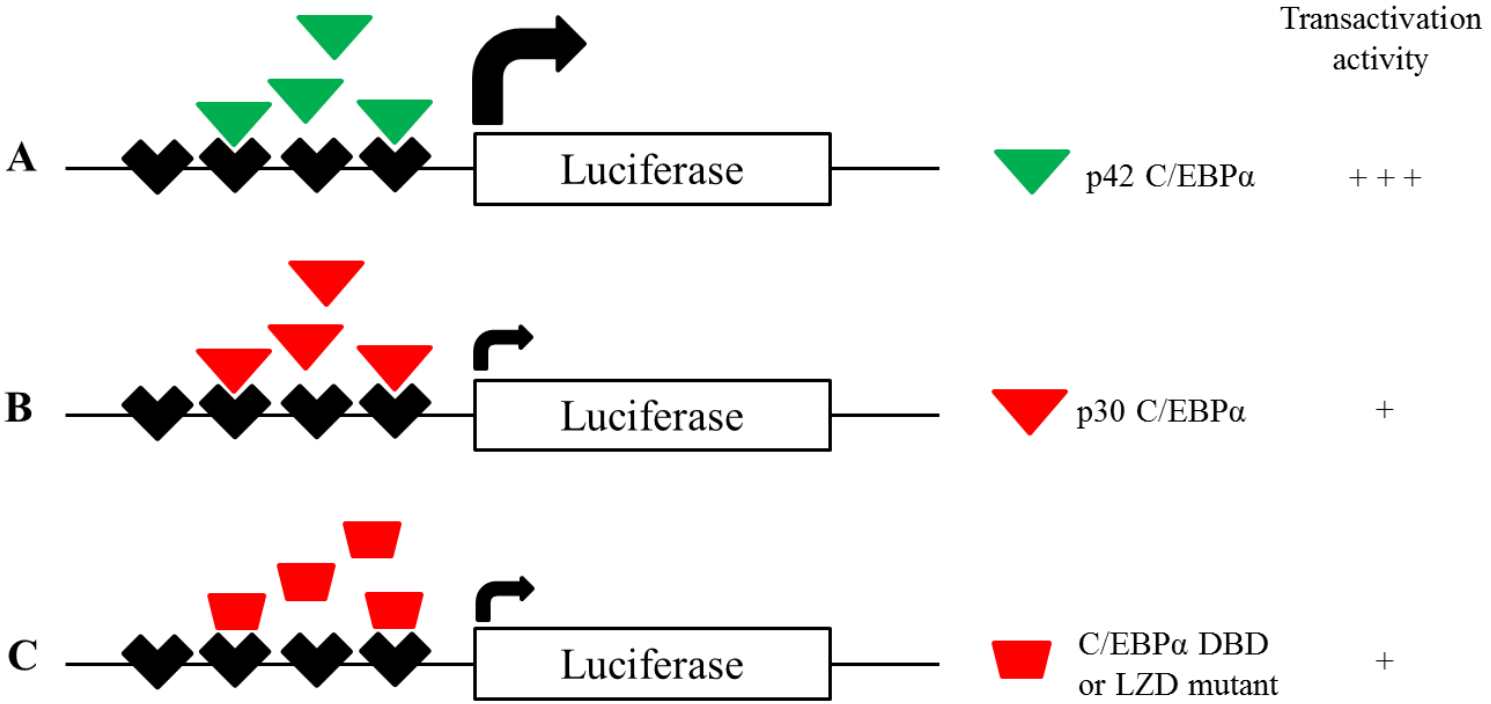


Figure 6.1 Principles of testing C/EBPα transactivation activity using a luciferase reporter assay. A luciferase reporter construct containing C/EBPα binding sites is co-transfected into cells with one or more *CEBPA* constructs. Wild-type p42 C/EBPα (A) can recognise its target sequence, bind to DNA, dimerise and activate transcription of the luciferase gene. The p30 isoform (B), retains the DNA binding and leucine zipper domains (DBD and LZD), but lacks the first transactivation domain, reducing transcription of the luciferase gene. C/EBPα with a DBD or LZD mutation (C) may be deficient in binding to DNA or dimerisation, thereby reducing luciferase transcription. The relative amounts of active firefly luciferase are determined by addition of the substrate, luciferin, and measuring luminescence. For simplification, C/EBPα dimers and mixtures are not shown here.

6.2 Materials and Methods

To study the TA activity of non-classical *CEBPA* mutations, a number of C/EBPα expression constructs (section 6.2.1) were made by PCR amplification of the required *CEBPA* sequence from genomic DNA (section 6.2.2) that was cloned into murine stem cell virus (MSCV) vectors containing a fluorescent marker gene and used to transform bacteria (section 6.2.3). Bacterial clones were analysed to identify constructs containing the desired *CEBPA* insert (section 6.2.4). C/EBPα expression vectors were then co-transfected into 293T cells with a *Renilla* luciferase expression plasmid supplied by Promega and a luciferase reporter construct, pTK81-CSF3R-luc, which was kindly donated by Thomas Pabst of the University of Bern, Switzerland with the permission of Daniel Tenen of the Beth Israel Deaconess Medical Center, Harvard, USA, and has been described previously (Pabst *et al*, 2001b) (section 6.2.5). The TA activity was then determined using the Dual-Luciferase Reporter Assay System from Promega (section 6.2.6).

6.2.1 Selection of C/EBPα controls and mutants

Overall, 16 different *CEBPA* sequences were cloned into MSCV vectors (Table 6.1). All constructs contained the entire *CEBPA* coding sequence except one, p30, which only included the sequence for the p30 isoform and was utilised as a control for immunoblotting. Three were control constructs containing WT *CEBPA*, a classical N-terminal or a classical C-terminal mutation. The N-terminal mutant control (hereafter called N), Q83fs, was detected in three cases in the cohort presented in chapter 3, two *CEBPA*-double and one *CEBPA*-single (Table 6.2), and was predicted to cause loss of full-length p42 isoform and an increase of p30 translation. The C-terminal mutant control (hereafter called C) was K313dup, a duplication in the LZD that has been investigated previously in both *in vitro* and *in vivo* studies (Carnicer *et al*, 2008;Bereshchenko *et al*, 2009) and was detected in six cases in the cohort (4 *CEBPA*-double, 2 *CEBPA*-single, Table 6.2). A second classical C-terminal mutation that was in the DBD (K304_Q305insL) identified in three *CEBPA*-double cases and the known 6 bp polymorphism (H195_P196dup, see section 3.1.2) were investigated as additional controls.

The remaining 10 *CEBPA* sequences were selected from the non-classical mutations identified in the patient cohort. Of the 115 mutations detected in 59 *CEBPA*-double patients, 14 (12%) were neither classical N-terminal nor classical C-terminal mutations. For the purposes of this study homozygous mutations were considered non-classical. They included four out-of-frame insertions/deletions, one nonsense and one missense mutation in the

Table 6.1 *CEBPA* constructs generated

Construct Type	Nucleotide change	Amino acid change	Method used to screen clones[1]
Controls:			
WT / Kozak-WT	None	None	Sequencing
Kozak-p30	1_357del	M1_V119del	Sequencing
6 bp polymorphism	584_589dup	H195_P196dup	Fragment analysis D4-CEBPA/2F + CEBPA/2R Δ = +6 bp
Classic N-terminal	247delC	Q83fs	Sequencing
Classic C-terminal (LZD)	937_939dup	K313dup	Sequencing
Classic C-terminal (DBD)	912_913insTTG	K304_Q305insL	Fragment analysis D4-CEBPA/3F + MSCV/R2 Δ = +3 bp
Atypical mutations from *CEBPA*-doubles cases:			
Out-of-frame del in middle	625delC	Q209fs	Sequencing
Out-of-frame ins in middle	707_708insT	A238fs	CEBPA/2F2 + CEBPA/2R *Fok*I digestion WT = 35+298 bp MUT = 35+93+205 bp
Out-of-frame ins in LZD	938_939insTA	K313fs	Sequencing
Missense in DBD	883G>C	A295P	CEBPA/3F + MSCV/R2 *Fok*I digestion WT = 400 bp MUT = 136+264 bp
Missense in middle	827A>G	K276R	CEBPA/3F + MSCV/R2 *Mnl*I digestion WT = 42+358 bp MUT = 39+42+319 bp
Atypical mutations from *CEBPA*-singles cases:			
Deletion in middle	558_566del	P187_P189del	Fragment analysis D4-CEBPA/2F + CEBPA/2R Δ = -9 bp
Deletion in middle	578_589del	H193_P196del	Fragment analysis D4-CEBPA/2F + CEBPA/2R Δ = -12 bp
Missense in middle	548C>A	P183Q	CEBPA/2F2 + CEBPA/2R *Bbv*I digestion WT = 47+50+89+147 bp MUT = 33+47+50+ 80+114 bp

Table 6.1 Continued

Construct Type	Nucleotide change	Amino acid change	Method used to screen clones[1]
Atypical mutations from *CEBPA*-single cases continued:			
Missense in middle	698C>G	P233R	CEBPA/2F2 + CEBPA/2R *Fsp*I digestion WT = 333 bp MUT = 106+227 bp
Missense in middle	724G>A	G242S	CEBPA/2F2 + CEBPA/2R *Dde*I digestion WT = 333 bp MUT = 82+251 bp

[1] Clones containing the desired *CEBPA* sequence were identified by either sequencing plasmid DNA from MSCV/F2 and MSCV/R2 or by PCR amplification with the indicated primers and subsequent fragment analysis or specific restriction enzyme digestion.
All primer sequences are given in Appendix Table 1. Abbreviations: del, deletion; ins, insertion; dup, duplication; DBD, DNA-binding domain; LZD, leucine zipper domain; bp, base pairs; Δ, difference in fragment size between WT and mutant clones; WT, wild-type clones; MUT, mutant clones.

Table 6.2 Patients identified with the mutations used in the constructs

CEBPA construct		Patient(s) identified with mutation		
Type	**Amino acid change**	**Single or Double[1]**	**Second mutation (location in gene[2])**	**No.[3]**
Controls:				
WT / Kozak-WT	None			
Kozak-p30	M1_V119del			
6 bp polymorphism	H195_P196dup			
Classic N-terminal	Q83fs	Single Double Double	NA R300_Q312dup (C) K313dup (C)	50 51 52[4]
Classic C-terminal (LZD)	K313dup	Double Double Double Double Single Single	G53fs (N) I68fs (N) L78fs (N) Q83fs (N) NA NA	29 43 48 52[4] 96 97
Classic C-terminal (DBD)	K304_Q305insL	Double Double Double	P23fs (N) H24fs (N) G99fs (N)	5 8 59
Atypical mutations from Doubles:				
Out-of-frame del in middle	Q209fs	Double	H24fs (N)	7
Out-of-frame ins in middle	A238fs	Double	V95fs (N)	56
Out-of-frame ins in LZD	K313fs	Double	G38fs (N)	21
Missense in DBD	A295P	Double	A44fs (N)	25
Missense in middle	K276R	Double Single	L178fs (Middle) NA	71 86
Atypical mutations from Singles:				
Deletion in middle	P187_P189del	Single Single	NA NA	74 75
Deletion in middle	H193_P196del	Single	NA	76
Missense in middle	P183Q	Single	NA	73
Missense in middle	P233R	Single	NA	79
Missense in middle	G242S	Single Single	NA NA	80 81

[1]Patient classified as *CEBPA*-single (Single) or *CEBPA*-double (Double). [2]Location in gene refers to either the N-terminus (N, from p42 ATG to p30 ATG), middle (between the p30 ATG and the DBD) or C-terminus (C, DBD or LZD). [3]Patient number relates to Appendix Table 2. [4]Mutations not present in CR sample. Abbreviations: NA, not applicable; del, deletion; ins, insertion; dup, duplication; DBD, DNA-binding domain; LZD, leucine zipper domain.

middle region, and three missense, three homozygous and two out-of-frame insertions/deletions in the C-terminus (Table 3.3). Of these atypical mutations, five representative alterations were chosen: an out-of-frame deletion in the middle of the gene (Q209fs), an out-of-frame insertion in the middle (A238fs) and one in the LZD (K313fs), all predicted to truncate both the p30 and the p42 isoforms, and two missense mutations, one just upstream of and one within the DBD (K276R and A295P, respectively). One of these, K276R, occurred in both a *CEBPA*-single and a *CEBPA*-double case (Table 6.2) and it was unknown if either missense mutation would impact on the TA activity of C/EBPα.

Of the 48 *CEBPA*-single cases identified in the cohort, 28 (58%) had atypical *CEBPA* mutations (Table 3.3), including 11 cases with out-of-frame insertions/deletions and one with a nonsense mutation in the middle of the gene, and three with out-of-frame insertions/deletions in the C-terminus, all predicted to cause a truncated protein for both isoforms. In addition, ten patients had missense mutations (five in the middle and five in the C-terminus) and three cases had in-frame deletions in the middle. Five representative mutations were selected for further analysis: two in-frame deletions (H193_P196del and P187_P189del) that were detected in one and two cases respectively and affected similar residues to the 6 bp polymorphism (H195_P196dup), and three missense mutations in the middle (P183Q, P233R, G242S) of unknown significance (Table 6.1 and Table 6.2).

6.2.2 *CEBPA* amplicons for insertion into MSCV vectors

There were three different primer pairs used in PCRs to amplify the required *CEBPA* sequences for insertion into MSCV vectors (Figure 6.2). All primer sequences are given in Appendix Table 1. First, amplicons of the entire *CEBPA* coding sequence were amplified from known *CEBPA*-WT genomic DNA (gDNA) using primers CEBPA/Clone/F and CEBPA/Clone/R to generate a 1125 bp product (Figure 6.2 A). For both primers, the 22 nucleotides at the 3′ end were complementary to the indicated *CEBPA* sequence, and the sequence towards the 5′ end contained a restriction enzyme recognition site for cloning purposes, *Eco*RI in the forward and *Xho*I in the reverse primer, with an additional 12 nucleotides at the 5′end to enable efficient digestion. In addition, the forward primer sequence included a Kozak consensus sequence directly upstream of the translation initiation codon for the p42 isoform, and therefore this amplicon was termed Kozak-WT (Figure 6.2 A).

Second, primers CEBPA/Clone-p30/F and CEBPA/Clone/R were used to generate a 768 bp PCR product containing only the p30 isoform sequence from a known *CEBPA*-WT gDNA

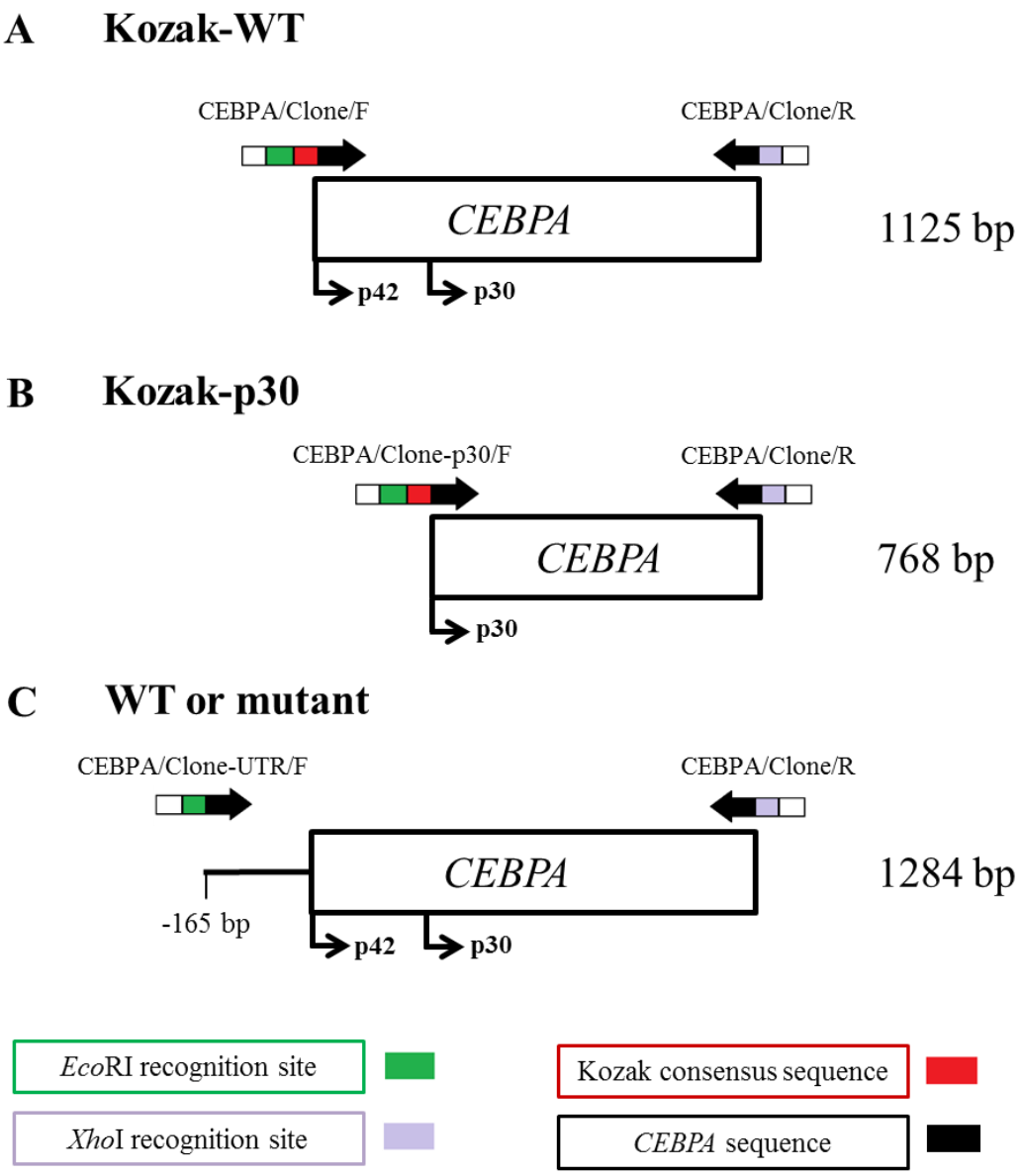


Figure 6.2 Generation of *CEBPA* constructs. The required *CEBPA* coding sequence was amplified from genomic DNA using primers that included a 3′ end complementary to the *CEBPA* sequence and a 5′ end that incorporated a restriction enzyme recognition site (EcoRI or XhoI) to enable cloning into the vector. Three different primer pairs were used: A. CEBPA/Clone/F and CEBPA/Clone/R were used to amplify the entire WT *CEBPA* coding sequence. The forward primer incorporated a Kozak consensus sequence directly upstream of the p42 translation initiation site (Kozak-WT). B. CEBPA/Clone-p30/F and CEBPA/Clone/R amplicons included only the p30 isoform *CEBPA* coding sequence and a Kozak consensus sequence was incorporated directly upstream of the p30 translation initiation site (Kozak-p30). C. CEBPA/Clone-UTR/F and CEBPA/Clone/R amplicons contained the entire *CEBPA* coding sequence and 165 base pairs (bp) of the 5′ untranslated region (WT or mutant) that included the endogenous Kozak consensus sequence.

(Figure 6.2 B). As above, the forward and reverse primer sequences included *Eco*RI and *Xho*I recognition sites and the forward primer contained a Kozak consensus sequence directly upstream of the translation initiation codon of the p30 isoform. This amplicon was termed Kozak-p30 (Figure 6.2 B). For reasons explained below (section 6.3), the Kozak-WT and Kozak-p30 C/EBPα constructs were only used as controls to produce protein for immunoblotting.

Primers CEBPA/Clone-UTR/F and CEBPA/Clone/R were used to amplify a 1284 bp PCR product that contained the entire *CEBPA* coding sequence and 165 bp of the 5′ untranslated region (UTR) from gDNA that included the endogenous Kozak consensus sequence (Figure 6.2 C). These primer sequences included the restriction enzyme recognition sites as described above. These primers were used to produce PCR products for creating each of the *CEBPA* constructs given in Table 6.1 except for the two Kozak constructs, and a gDNA sample containing the desired *CEBPA* sequence (WT or mutant) was used as template (Table 6.2).

The proof-reading enzyme Phusion Hot Start High-Fidelity DNA Polymerase was used for amplification with all three primer pairs to minimise the addition of incorrect nucleotides. The standard reaction mix was used (see section 2.1.2), except this was scaled up to 50µl and 5% DMSO was added. The annealing temperature was 62°C and the standard cycling conditions were used, except the extension time was increased from 42 seconds to 1 minute and the final extension step from 5 minutes to 10 minutes. The entire PCR product was electrophoresed in a 1% agarose gel along with a DNA ladder (Hyperladder I, Bioline, London, UK), and the correctly sized band was excised from the gel and purified using the QIAquick Gel Extraction kit (QIAGEN, Crawley, UK).

6.2.3 Cloning of *CEBPA* amplicons into MSCV vectors

Two different MSCV retroviral expression vectors, MSCV.I.GFP (6564 bp) and MSCV.I.eBFP2 (6541 bp), were used to create the *CEBPA* constructs. They were kindly provided by Dr. Martin Pule of the Department of Haematology, UCL Cancer Institute. Both plasmids contained an ampicillin resistance gene, an internal ribosome entry site (IRES) and a fluorescent marker gene, either green fluorescent protein (GFP) or enhanced blue fluorescent protein 2 (eBFP2), and the sequences were identical except for the GFP or eBFP2 sequence. The fluorescent marker genes were used to assess the efficiency of transfections. When two *CEBPA* constructs were co-transfected, each construct had a different marker gene and the different markers could be used to check that both *CEBPA*

constructs were in the same cells. All 16 *CEBPA* sequences were cloned into the MSCV.I.GFP vector, and the WT, N and C mutant controls were also cloned into the MSCV.I.eBFP2 vector. An overview of the cloning strategy is shown in Figure 6.3.

The cloning site of the MSCV vectors contained an *Eco*RI and an *Xho*I restriction enzyme recognition sequence, which were also present at the 5′ and the 3′ end of the *CEBPA* amplicons (section 6.2.2), and digestion of vector and PCR products with both *Eco*RI and *Xho*I created complementary over-hanging ends enabling ligation of the insert into the vector in both the correct location and orientation (Figure 6.3). For digestion of either vector, 5µg of vector was incubated with 5µl of *Eco*RI, 5µl of *Xho*I and 10µl of *Eco*RI 10x reaction buffer (New England Biolabs Ltd., Hitchin, UK) in a total volume of 100µl at 37°C for 5 hours. For digestion of the PCR amplicons, the digestion mix was identical except the entire purified product (section 6.2.2) was used rather than the vector. The entire digestion reaction for vectors and PCR amplicons was electrophoresed in a 1% agarose gel together with Hyperladder I and correctly sized bands were excised and purified using the QIAquick Gel Extraction kit.

The DNA concentrations of the purified digested products and vectors were determined using a NanoDrop 1000 Spectrophotometer and the values used to calculate appropriate volumes of both vector and insert to use in the ligation reaction according to the formula:

$$\text{ng of insert} = \frac{\text{length of insert, kilobase (kb)} \times \text{ng of vector}}{\text{length of vector, kb} \times \text{insert:vector ratio}}$$

The amount of vector used was always 1000ng and the insert:vector ratio 3:1. The ligation mix contained 3µl T4 DNA ligase (New England Biolabs UK Ltd.), 5µl T4 DNA ligase buffer, 1000ng of the digested vector, the calculated amount of digested insert and water to a total volume of 50µl, and this was incubated at 16°C for 15 minutes. The ligase reaction mix, 2µl, was then used to transform One Shot Max Efficiency DH5α-T1 *E. coli* (see section 2.1.9).

6.2.4 Analysis of bacterial clones

For all constructs where sequencing was used to screen colonies (Table 6.1), between 2 and 5 bacterial colonies per transformation were picked and individually seeded into 4ml of LB containing 100µg/ml carbenicillin and incubated with shaking at 37°C overnight.

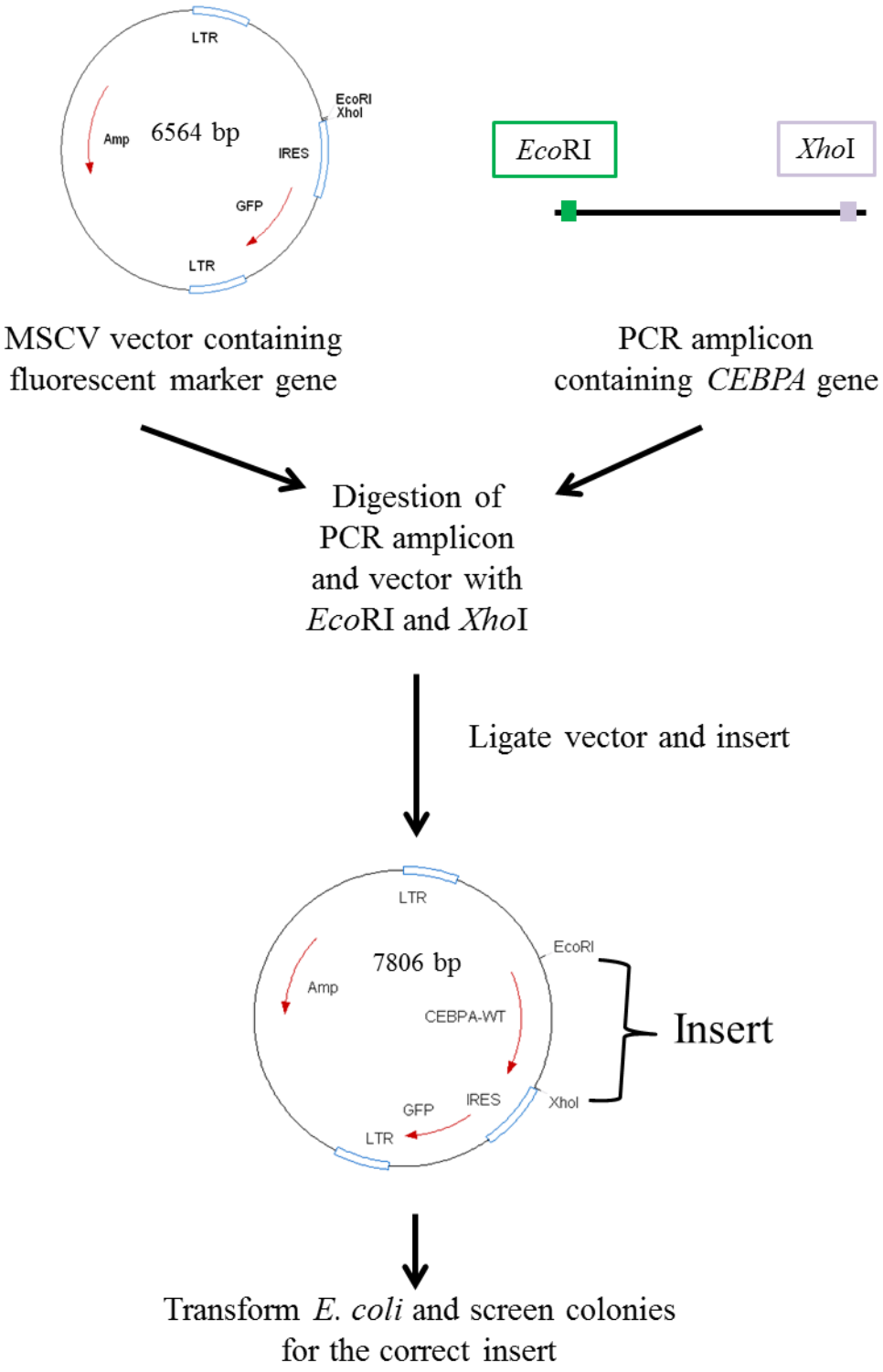


Figure 6.3 Overview of *CEBPA* cloning. PCR amplicons of the *CEBPA* coding sequence were obtained from genomic DNA with primers that introduced an *Eco*RI restriction site at the 5′ end and an *Xho*I restriction site at the 3′ end of the PCR product to enable cloning into the MSCV vector. The insert and vector were digested with *Eco*RI and *Xho*I before ligation and transformation of *E. coli*.

Glycerol stocks of each clone were made and the plasmid DNA was extracted from the remaining culture using the QIAprep Spin Miniprep Kit (QIAGEN), as described in section 2.1.10. The entire *CEBPA* insert was sequenced from primers that annealed in the vector, MSCV/F2 and MSCV/R2.

For the remaining constructs, between 10 and 20 colonies per transformation were picked and individually seeded into 200µl of LB containing 100µg/ml carbenicillin in a well of a 96-well plate and incubated overnight at 37°C (section 2.1.10). The region of the *CEBPA* sequence containing the location of the desired sequence alteration was then amplified using the bacterial culture as template with the primers given in Table 6.1 (see chapter 5 for PCR protocols). In four cases, fragment analysis with a fluorescently labelled primer was performed as described in chapter 5, with clones containing the required mutation selected by size. In six cases, PCR products were digested with a restriction enzyme that gave a different digestion pattern for WT and mutant clones, which could be detected by agarose gel electrophoresis. All restriction enzymes and digested fragment sizes for WT and mutants are given in Table 6.1. Two or three colonies per construct that were positive for the desired size change by fragment analysis or digestion pattern by restriction enzyme digestion were then selected for sequencing. The culture was used to inoculate 4ml of LB containing 100µg/ml carbenicillin and incubated with shaking at 37°C overnight. Glycerol stocks of each clone were made and the plasmid DNA was extracted from the remaining culture using the QIAprep Spin Miniprep Kit (see section 2.1.10). The entire *CEBPA* insert was sequenced from primers MSCV/F2 and MSCV/R2.

A single colony with the desired *CEBPA* sequence, either WT or mutant, was selected for each construct and the glycerol stock used to grow a 200ml culture from which the plasmid DNA was extracted using the HiSpeed Plasmid Midi Kit (QIAGEN), as described in section 2.1.10.

6.2.5 Co-transfection of 293T cells

293T cells were transiently co-transfected with a mixture of plasmids using GeneJuice as a transfection reagent as described in section 2.2.4. Optimisation of the amounts of each construct per well in the transfection was performed and is described in section 6.3.1. Following optimisation, cells were co-transfected for dual-luciferase assays, immunoblotting and FACS analysis as detailed below.

For dual-luciferase assays, each well of a 24-well plate received 80ng pTK81-CSF3R-luc reporter construct, 1ng *Renilla* luciferase construct (Promega) and a total of 100ng of empty MSCV vector (V) or *CEBPA* construct, either a single construct or a 1:1 mixture of two constructs. For immunoblotting and FACS analysis, 293T cells were transfected in 6-well plates. The amount of DNA was increased to 320ng pTK81-CSF3R-luc, 4ng *Renilla* luciferase and 400ng of V or *CEBPA* construct per well, either a single construct or a 1:1 mixture of two constructs. Cells were left for 24 hours at 37°C in a humidified atmosphere with 5% CO_2, and the efficiency of the transfection checked by detection of GFP and/or eBFP2 expression with a fluorescence microscope before further analysis.

6.2.6 Immunoblotting for C/EBPα protein

Cell lysates from transfected and non-transfected 293T cells were prepared and between 20µg and 30µg protein subjected to SDS-PAGE, as described in sections 2.3.4 to 2.3.6. The separated proteins were transferred to a nitrocellulose membrane that was then blocked (2.3.6 and 2.3.7). Two C/EBPα antibodies were used in this study, #2295 and #2843 (both rabbit anti-human from Cell Signaling Technologies). The epitope for #2295 is downstream of the p30 translation initiation site and therefore both p42 and p30 isoforms could be detected, while the epitope for #2843 is near the N-terminus and therefore detects p42, but not p30. C/EBPα antibody was incubated with the membrane over-night on a rocking platform at 4°C at a concentration of 1:1000 in 5% non-fat milk powder in TBST for #2295 and in 5% BSA in PBS for #2843. Incubation with secondary antibody and detection were performed as described in section 2.3.7. If both C/EBPα antibodies were to be used, the membrane was stripped after detection of the first antibody and re-blocked before incubating with the second antibody. After detection of the primary antibody/antibodies, the membrane was re-blocked, without stripping, and incubated for 1-2 hours at room temperature with anti-tubulin antibody (mouse anti-human, Sigma-Aldrich Company Ltd.) as a loading control at a concentration of 1:5000 in 1% BSA in PBS. Incubation with the secondary antibody and detection were then performed as described in section 2.3.7.

6.2.7 Flow cytometric analysis

Growth medium was removed from cells in 6-well plates and the cells washed once in PBS, then incubated with 300µl trypsin-EDTA solution per well until they dissociated from the plate surface. The cell suspension was transferred to a FACS tube and centrifuged for 5 minutes at 450g, the supernatant removed and the cells resuspended in 500µl PBS before analysis of GFP and BFP expression on a CyAn flow cytometer (Beckman Coulter).

6.2.8 Dual-luciferase reporter assays

The Dual-Luciferase Reporter Assay System from Promega was used to measure the TA activity of C/EBPα variants (Figure 6.4). In this system, variation in transfection efficiency is controlled for by co-transfection of the C/EBPα and pTK81-CSF3R-luc constructs into 293T cells together with a *Renilla* luciferase expression plasmid, supplied by Promega, as an internal control. The firefly luciferase reporter construct, pTK81-CSF3R-luc, contains four copies of the C/EBPα binding site from the *CSF3R* gene promoter upstream of a firefly luciferase gene, and transcription of the luciferase gene is caused by active C/EBPα (Pabst *et al*, 2001b). It was assumed that the *Renilla* luciferase expression was independent from the C/EBPα and firefly luciferase expression. The firefly and *Renilla* luciferases require different substrates (luciferin and coelenterazine, respectively) and this is exploited in the assay.

As per the manufacturer's instructions, co-transfected 293T cells were lysed in the supplied passive lysis buffer and the firefly luciferase reaction initiated by addition of the lysate to LARII, which contains luciferin (Figure 6.4). The luminescence generated, reflecting the TA activity of the C/EBPα, was measured using a TD-20/20 luminometer (Turner Designs) that was programmed for a 2 second pre-read delay followed by a 10 second measurement period. Addition of Stop & Glo Reagent simultaneously quenched the firefly luciferase reaction and initiated the *Renilla* luciferase reaction by supplying coelenterazine. The luminescence caused by the *Renilla* luciferase reaction was then measured and reflected the transfection efficiency of that sample. The mean luminescence over the measurement period was calculated for both the firefly (L) and the *Renilla* (R) luciferase for each sample. Results were normalised for transfection efficiency using the *Renilla* activity by calculating L/R for each sample. Unless indicated otherwise, within each experiment three separate wells were transfected per condition and the average L/R of the three measurements calculated before normalising all results in the same assay to the WT control, which was set at 100%, so that all results were expressed as a percentage of WT C/EBPα activity. Each condition was repeated in three or more separate experiments and the mean plotted with standard error of the mean as error bars. A t-test with two-tails was used to compare results and *P*-values < .05 were considered significant.

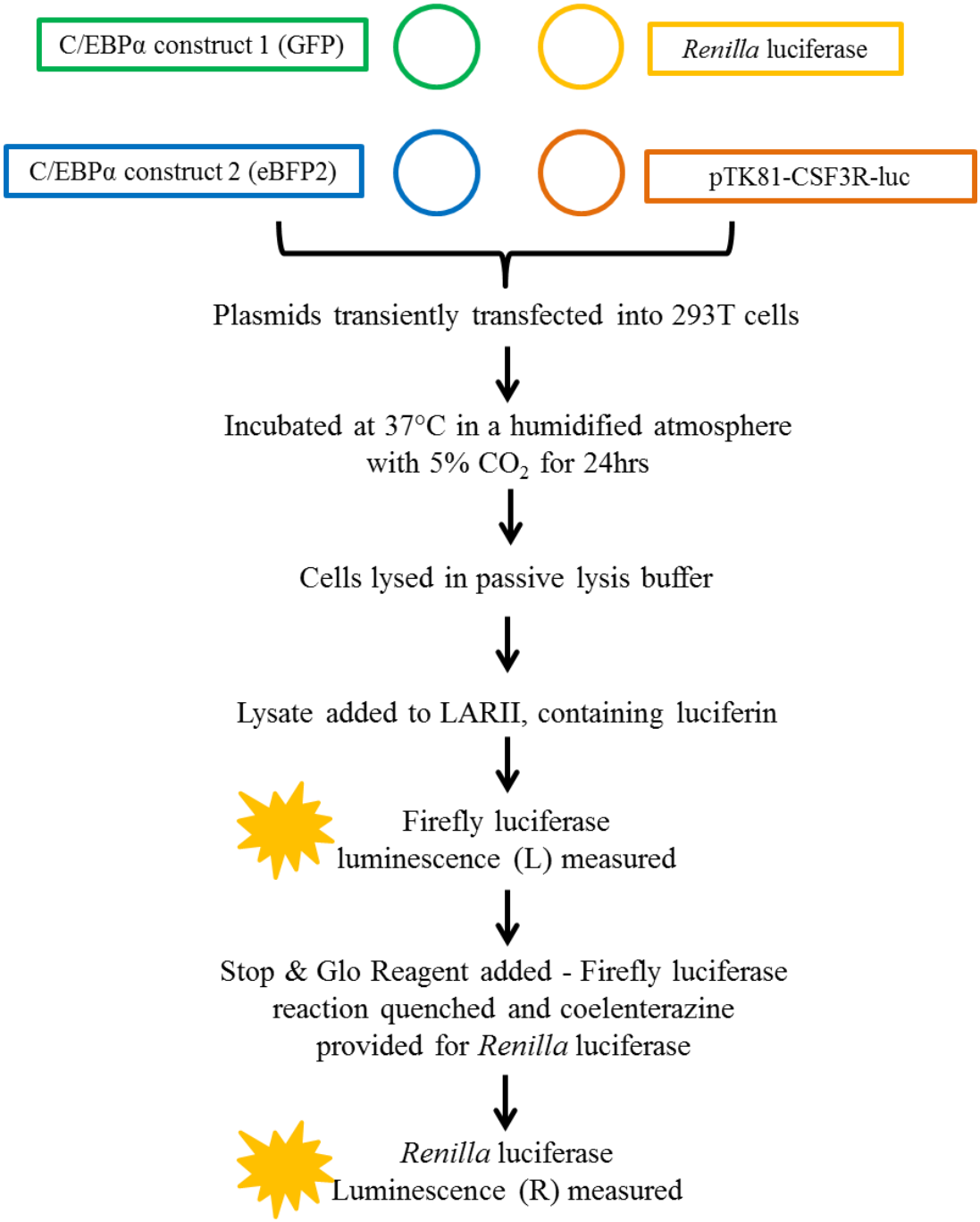


Figure 6.4 Overview of the dual-luciferase reporter assay. C/EBPα constructs with fluorescent marker genes (green fluorescent protein, GFP, or enhanced blue fluorescent protein 2, eBFP2) were co-transfected into 293T cells with the firefly luciferase reporter construct (pTK81-CSF3R-luc) and a Renilla luciferase construct as an internal control for the transfection. Using the Dual-Luciferase Reporter Assay System (Promega), firefly luciferase and Renilla luciferase activity were sequentially measured in cell lysates.

6.3 Results

Three control *CEBPA* constructs, WT, an N-terminal mutant (Q83fs, N) and a C-terminal mutant (K313dup, C), were made by cloning PCR products of the entire *CEBPA* coding sequence and 165 bp of the 5′ UTR from a gDNA sample containing the desired *CEBPA* sequence into both MSCV.I.GFP and MSCV.I.eBPF2 (Figure 6.2 C). The conditions for the dual-luciferase assay were optimised using these control constructs. Two other control constructs, Kozak-WT and Kozak-p30, were made by cloning PCR products of the *CEBPA* coding sequence from known *CEBPA*-WT gDNA into MSCV.I.GFP, with Kozak-WT containing the entire p42 sequence and Kozak-p30 only the p30 isoform sequence. Both were amplified with a forward primer that contained a Kozak consensus sequence directly upstream of the translation initiation sites for the p42 and p30 isoforms, respectively (Figure 6.2 A and B). The Kozak-WT construct was prepared first but it was found to be unsuitable for the luciferase-reporter assay as only the p42 isoform, not the p30, was expressed from this construct. Endogenous expression of the p30 isoform requires a conserved open reading frame (ORF) that initiates at "site D", 25 bp upstream of the most commonly used p42 isoform translation initiation site (Calkhoven *et al*, 2000). This short upstream ORF is in a different reading frame from the *CEBPA* coding sequence and terminates only five codons after initiation. With, presumably, insufficient space for re-initiation at the p42 ATG, this is predicted to lead to translation re-initiation by the ribosome further downstream at the p30 isoform translation initiation site, with mutation of site D associated with a loss of p30 isoform expression (Calkhoven *et al*, 2000). The Kozak-WT and Kozak-p30 constructs were, therefore, only used as controls to produce protein for immunoblotting.

6.3.1 Optimising assay conditions

Optimisation was performed to determine the amount of each construct that should be used in the co-transfection of 293T cells for the dual-luciferase assays. For the *Renilla* luciferase expression construct, transfection of 293T cells with 1ng per well of a 24-well plate resulted in strong luminescence from the *Renilla* luciferase reaction (R). For instance, the mean R across 24 transfected wells in one experiment was 1803 relative light units (range 1325 to 2701). Therefore, this amount was used in all experiments.

In order to investigate the impact of the amount of C/EBPα expression construct on the transfection efficiency, cells were transfected with 100ng, 200ng or 460ng of construct per well of a 24-well plate for each WT, N or C control MSCV.I.GFP construct. The proportion of cells expressing the fluorescent marker gene after 24 hours was determined by flow

cytometry (Figure 6.5). The average transfection efficiency overall was 58% (range 47% to 74%), and there were no striking differences according to either the amount of construct per well or which construct was used.

To examine the impact of the amount of *CEBPA* construct in the dual-luciferase assay, 50ng, 100ng or 200ng of MSCV.I.GFP vector alone and each MSCV.I.GFP WT, N or C construct was co-transfected with 1ng of the *Renilla* luciferase expression construct and 80ng of the luciferase reporter construct, pTK81-CSF3R-luc. This amount of pTK81-CSF3R-luc was selected as it had been used in a previous report (Pabst *et al*, 2009). The firefly luciferase activity (L) and R were measured by dual-luciferase assay after 24 hours and the results normalised for variation in transfection efficiency by calculating L/R. The mean and standard error of the mean (SEM) of three separate transfections is shown in Figure 6.6 A. The background luminescence from the cells with the vector alone was very low (mean L/R 0.006, range 0.004 to 0.007). Increasing the amount of construct per well from 50ng to 200ng caused a slight increase in L/R but the differences were not significant for a specific construct (Figure 6.6 A). As expected, the WT construct strongly activated transcription, with an average 44-fold increase in L/R over the vector alone, and there were significant differences in the TA activity of the different C/EBPα control constructs, which will be discussed in section 6.3.2.3.

The impact of different amounts of pTK81-CSF3R-luc on the dual-luciferase assay was also tested. 293T cells were transfected with 50ng, 80ng, 100ng or 150ng of the luciferase reporter construct and 50ng or 100ng of the V alone, WT, N or C MSCV.I.GFP C/EBPα expression construct together with 1ng of the *Renilla* luciferase expression construct per well, and TA activity measured after 24 hours (Figure 6.6 B). The experiment was only performed once. As before, the background L/R for the MSCV.I.GFP transfected cells was very low for all conditions (average 0.012, range 0.006 to 0.023). Overall, increasing the amount of pTK81-CSF3R-luc was associated with an increase in the L/R, regardless of the C/EBPα construct in the co-transfection, with the exception of a slight reduction in L/R from 50ng to 80ng of the luciferase reporter construct. However, the amount of pTK81-CSF3R-luc did not impact on the relative TA activity pattern of the different C/EBPα expression constructs within the same set of conditions, with the highest L/R values for the WT construct conditions and a reduction in L/R for the N and C mutant constructs (Figure 6.6 B).

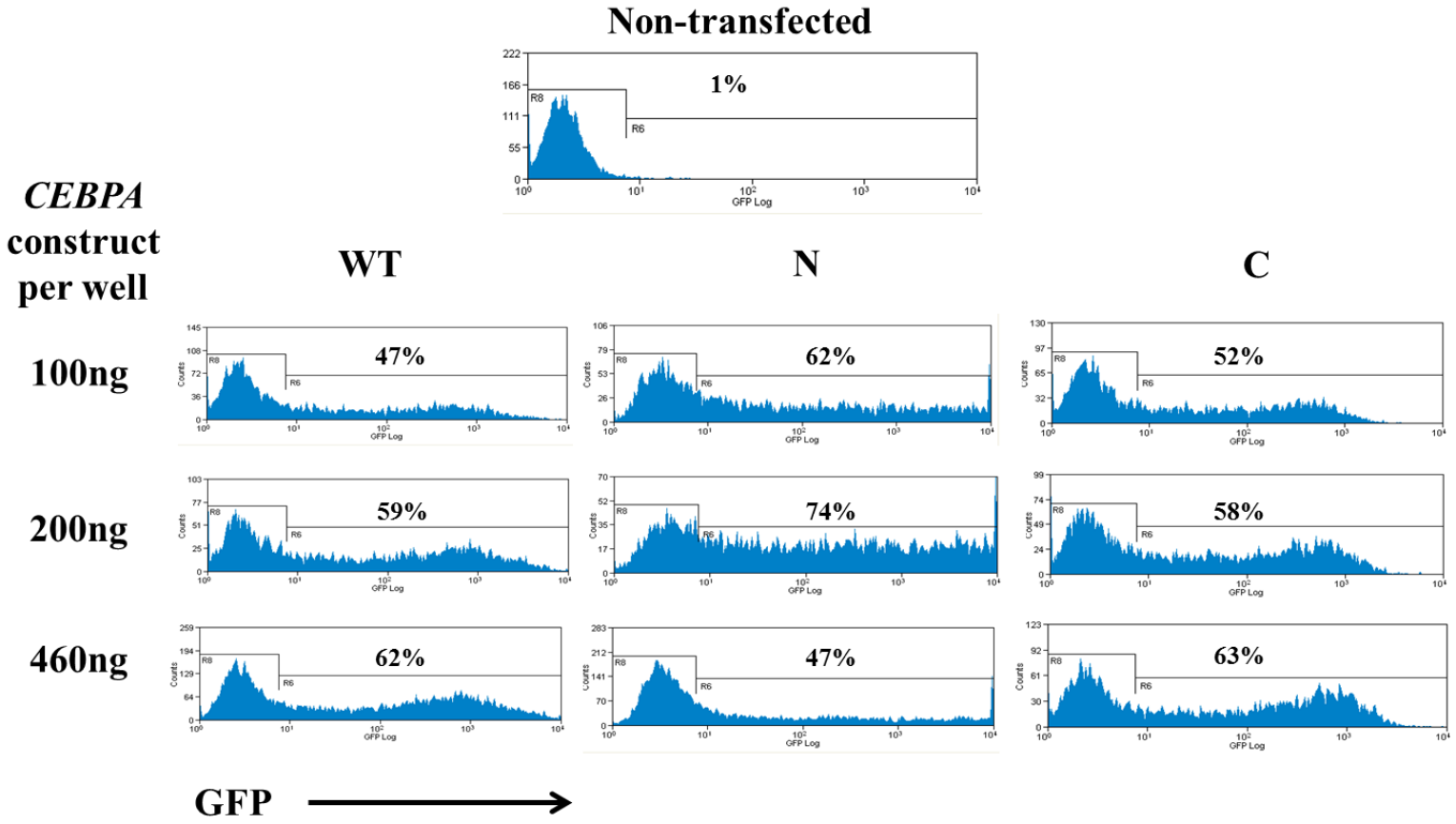


Figure 6.5 Impact of *CEBPA* construct quantity on transfection efficiency. Representative FACS plots of 293T cells 24 hours after transfection with varying amounts (100ng, 200ng or 460ng) of WT, an N-terminal mutant (N) or a C-terminal mutant (C) C/EBPα MSCV.I.GFP construct per well of a 24-well plate. The percentage indicates the proportion of GFP positive cells.

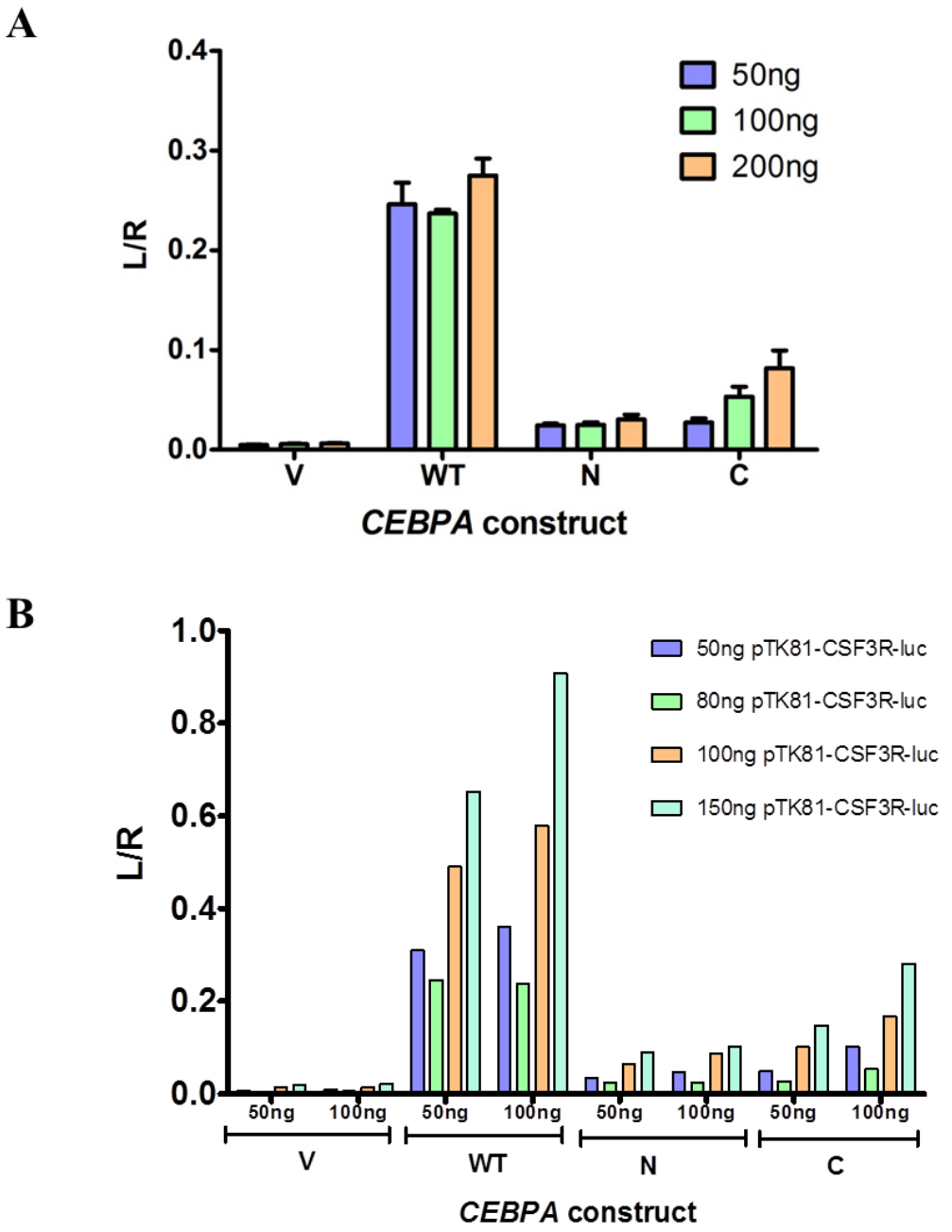


Figure 6.6 Optimisation of transient transfections for dual-luciferase assays. Results of dual-luciferase assays (Promega) performed on 293T cells 24 hours after co-transfection with 1ng of the *Renilla* luciferase expression construct and varying amounts of the MSCV.I.GFP *CEBPA* construct and the luciferase reporter construct (pTK81-CSF3R-luc). Firefly luminescence (L) was normalised for variation in transfection efficiency by calculating L/ *Renilla* luciferase luminescence (R). A. Cells co-transfected with 80ng of pTK81-CSF3R-luc and 50ng, 100ng or 200ng of the MSCV.I.GFP vector alone (V) or a *CEBPA* construct (WT, N or C) per well. The mean and SEM of three separate experiments is shown. B. Cells co-transfected with 50ng, 80ng, 100ng or 150ng of pTK81-CSF3R-luc, and either 50ng or 100ng of a *CEBPA* construct or V alone per well. The results of a single experiment are shown.

For all subsequent dual-luciferase assays cells were therefore co-transfected with 100ng of the C/EBPα expression construct, 80ng of pTK81-CSF3R-luc and 1ng of the *Renilla* luciferase construct per well of a 24-well plate

6.3.2 Transactivation activity of control *CEBPA* constructs

Using the co-transfection conditions determined above, C/EBPα protein expression, marker gene expression and the TA activity of the control constructs (WT, N and C) in both MSCV.I.GFP and MSCV.I.eBFP2 were investigated.

6.3.2.1 C/EBPα expression in cells transfected with single constructs

To confirm that the control *CEBPA* constructs expressed C/EBPα protein, each control construct (WT, N and C in both MSCV.I.GFP and MSCV.I.eBFP2) was individually transfected into 293T cells and C/EBPα protein expression determined by Western blotting of lysates prepared 24 hours after transfection using an antibody that detects both the p42 and p30 isoforms (Figure 6.7). The Kozak-WT construct lacked the upstream ORF and 293T cells transfected with this construct expressed the p42 protein, but not the p30 isoform (Figure 6.7, lane 2). The Kozak-p30 construct contained only the p30 isoform sequence and Western blotting detected the p30 isoform, but not the p42 isoform (Figure 6.7, lane 3). The WT, N and C constructs all contained 165 bp of the 5′ UTR, including the endogenous Kozak consensus sequence and the short upstream ORF. Accordingly, both the p42 and p30 isoforms were expressed from the WT and C constructs (Figure 6.7, lanes 4, 5, 8 and 9). Only the p30 isoform was detected in cells transfected with the N-terminal frame-shift mutant construct (Figure 6.7, lanes 6 and 7). As determined visually, there were no striking differences in C/EBPα expression between the MSCV.I.GFP and MSCV.I.eBFP2 constructs containing the same *CEBPA* sequence (Figure 6.7).

6.3.2.2 Marker gene and C/EBPα expression in cells co-transfected with two constructs

For co-transfection of two constructs, the marker gene was different for each construct. The control C/EBPα expression constructs were co-transfected in a 1:1 mix to a total of 100ng of DNA per well of a 24-well plate, in order to create a mix equivalent to the two *CEBPA* alleles in patients.

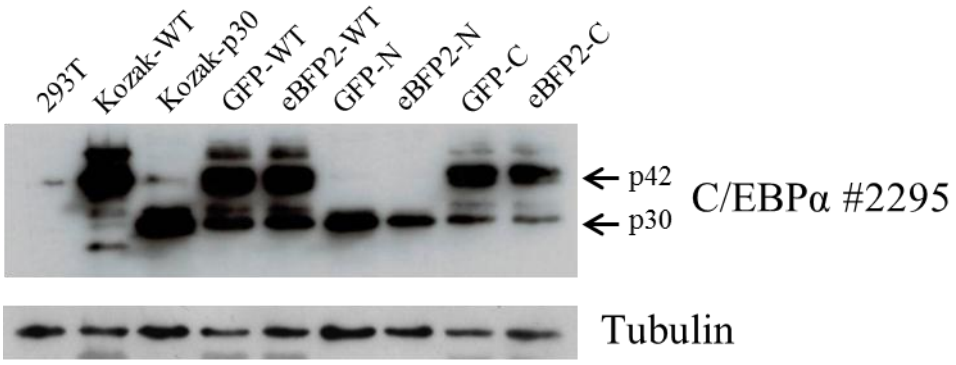


Figure 6.7 C/EBP*α* protein expression from *CEBPA* control constructs. Western blots of 293T cell lysates made 24 hours after transfection with control C/EBPα expression constructs (WT, N or C) with one of two fluorescent marker genes, GFP or eBFP2. All control constructs were individually transfected into cells. The Kozak-WT and Kozak-p30 constructs were used to produce p42 and p30 protein, respectively, as controls. The epitope for the anti-C/EBPα antibody #2295 is downstream of the p30 translation initiation site and it therefore detects both the p30 and the p42 isoforms.

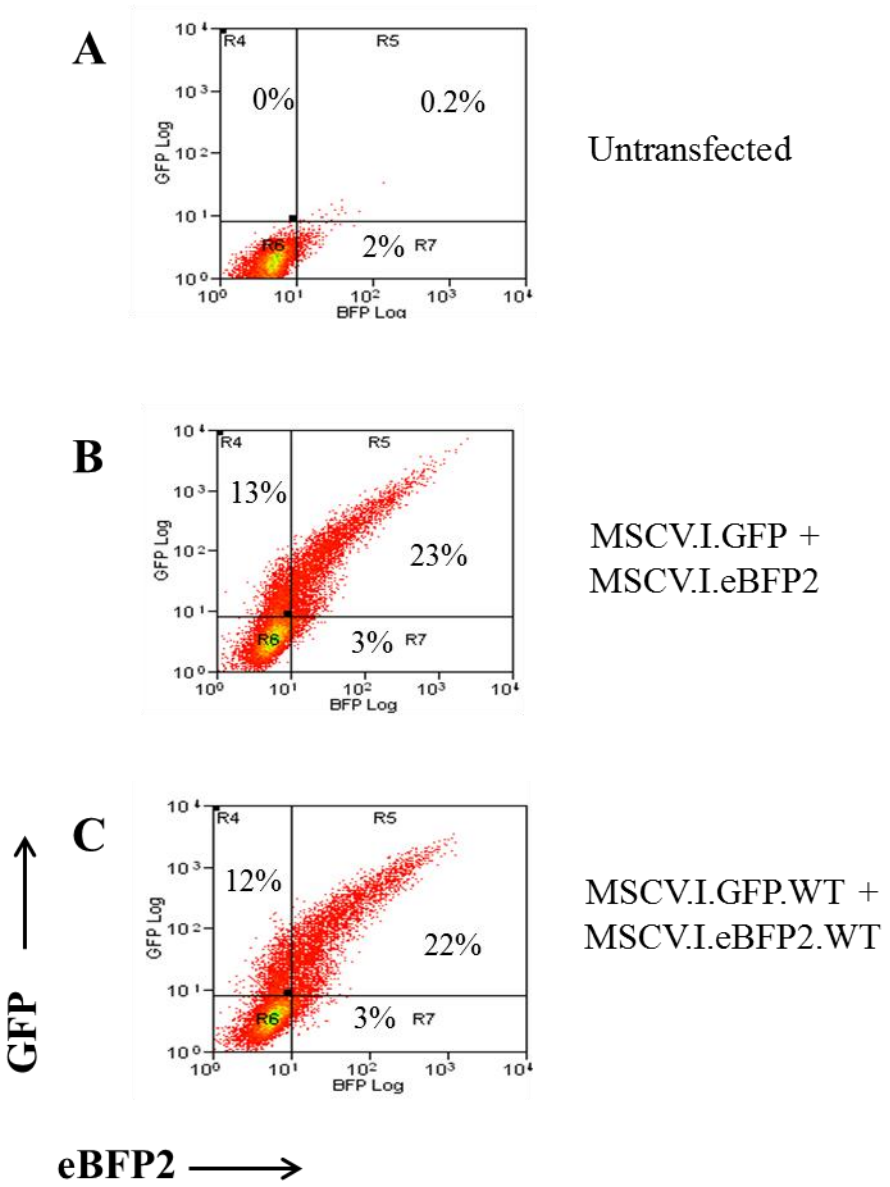


Figure 6.8 Marker gene expression in co-transfections of vector alone or *CEBPA* constructs. Representative FACS plots of 293T cells (A) untransfected or 24 hours post-transfection with a 1:1 mix of MSCV.I.GFP and MSCV.I.eBFP2, either (B) vector alone or (C) *CEBPA* WT constructs.

The relative transfection efficiency of constructs with the different marker genes was tested by transfecting cells with either the MSCV.I.GFP and MSCV.I.GFP vectors alone (V), or the equivalent WT constructs. The transfection efficiency was assessed after 24 hours by flow cytometric analysis. Each condition was tested in four separate wells across two different transfections. Overall, the average proportion of cells expressing one or both of the fluorescent marker genes was 32% (range 22% to 39%). An example is shown in Figure 6.8. Using gates from the untransfected cells (Figure 6.8 A), 36% and 34% of cells expressed GFP and/or eBFP2 for the V alone and the WT constructs, respectively (R4 and R5, Figure 6.8 B and C). Of these positive cells, approximately two thirds were both GFP and eBFP2 positive. It was consistently observed that the remaining transfected cells expressed GFP, but not eBFP2. The reasons for this were not clear.

To test the impact of the marker gene on C/EBPα protein expression in these co-transfections, mix and match co-transfections of MSCV.I.GFP and MSCV.I.eBFP2 constructs were performed with all six control constructs, WT, N and C with both GFP and eBFP2 marker genes in all possible combinations. The C/EBPα protein expression resulting from these was examined by Western blotting of lysates prepared after 24 hours (Figure 6.9). Both the p30 and the p42 isoforms were detected for all control C/EBPα construct combinations except the N-terminal mutant construct only condition, consistent with the loss of the p42 isoform expected for this mutant. For a particular construct combination no striking differences in protein expression were observed with respect to which marker gene was associated with which C/EBPα construct (Figure 6.9).

6.3.2.3 Transactivation activity of co-transfected control constructs

Dual-luciferase assays were performed on cells co-transfected with all the possible combinations of two control C/EBPα expression constructs (Table 6.3). The L/R was normalised to the WT activity within each assay, which was set at 100% so that TA activity was expressed as a percentage of the WT activity. Each condition was tested in between four and 11 separate experiments, and the means and SEMs are given in Table 6.4. The background TA activity of the V alone was very low at only 3% ± 0.4% (mean ± SEM) of the WT construct (Figure 6.10).

The impact of the marker gene, GFP or eBFP2, on construct activity was explored for the three *CEBPA* construct combinations: the N mutant plus WT constructs, the C mutant plus WT constructs and the N plus C mutant constructs. Each combination was tested with both possible marker gene-C/EBPα configurations, i.e. GFP-N with eBFP2-WT and eBFP2-N

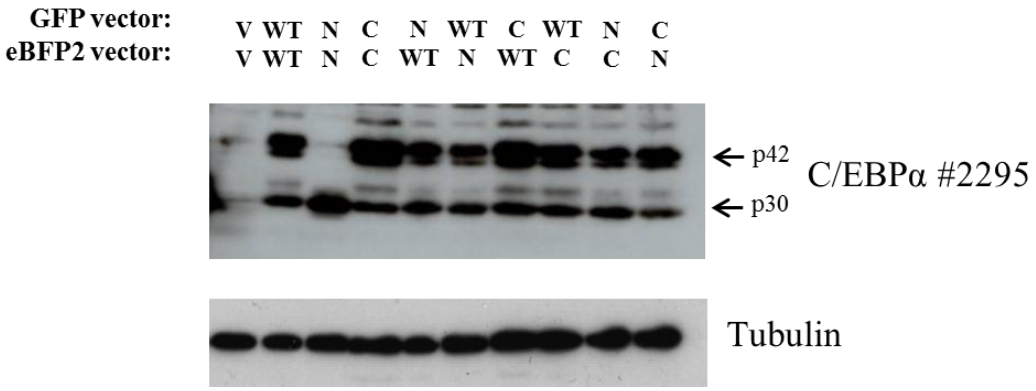


Figure 6.9 C/EBP*α* protein expression from co-transfected *CEBPA* control constructs
Western blots of 293T cell lysates made 24 hours after co-transfection with a 1:1 mix of MSCV.I.GFP and MSCV.I.eBFP2 constructs either vector alone (V) or containing control *CEBPA* sequences: WT, N-terminal mutant (N) or a C-terminal mutant (C).

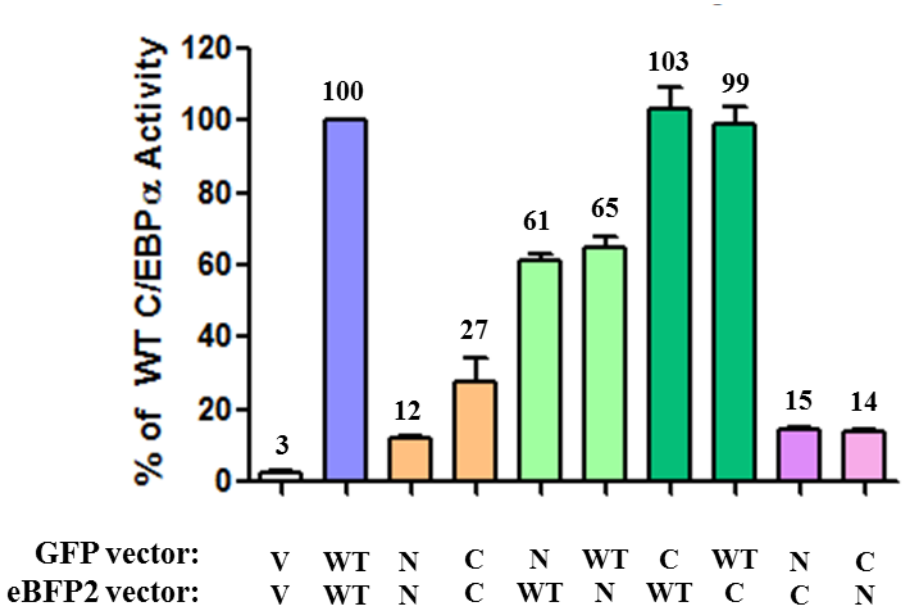


Figure 6.10 Transactivation activity of C/EBPα control constructs.
Transactivation activity was determined in 293T cells co-transfected with a 1:1 mix of vector alone (V) or C/EBPα control constructs (WT, N or C), one an MSCV.I.GFP construct and the other an MSCV.I.eBFP2 construct, by dual-luciferase assay (Promega). The mean and standard error of the mean of between 4 and 11 experiments is shown.

Table 6.3 Combinations of *CEBPA* constructs tested in dual-luciferase assays

Construct 2	None	WT (eBFP2)	N (eBFP2)	C (eBFP2)	Q209fs (GFP)
Construct 1 (GFP)					
Control constructs					
WT		✓	✓	✓	
N		✓	✓	✓	
C		✓	✓	✓	
6bp polymorphism[1]	✓	✓	✓	✓	
K304_Q305insL[2]	✓	✓	✓		
Test mutations in *CEBPA*-doubles					
Q209fs			✓		
A238fs			✓		
K313fs			✓		
A295P			✓		
K276R[3]		✓			✓
Test mutations in *CEBPA*-singles					
P187_P189del		✓	✓	✓	
H193_P196del		✓	✓	✓	
P183Q		✓	✓	✓	
P233R		✓	✓	✓	
G242S		✓	✓	✓	

[1]6bp polymorphism is the H195_P196dup polymorphism. [2]Classical C-terminal mutation in the DBD rather than the LZD. [3]Mutation in a *CEBPA*-single (patient 86) and *CEBPA*-double case (patient 71). The fluorescent marker gene in the MSCV *CEBPA* construct, either green fluorescent protein (GFP) or enhanced blue fluorescent protein (eBFP2) is indicated in each case. Abbreviations: WT, wild-type; N, Q83fs; C, K313dup.

with GFP-WT. Importantly, no significant differences were observed in the C/EBPα TA activity between the same *CEBPA* construct combinations with the alternate marker gene configuration (Table 6.4, Figure 6.10). For N+WT and WT+N the difference was only 4%, with a TA activity of 61% ± 2% (mean ± SEM) and 65% ± 3% of the WT C/EBPα TA activity, where the marker genes were GFP+eBFP2 respectively ($P = .3$). Similarly, for C+WT and WT+C a difference of 4% was also observed (103% ± 6% and 99% ± 5%, respectively) and for N+C and C+WT there was a difference of only 1% between the two marker gene configurations (15% ± 1% and 14% ± 1%). With an average difference of 3%, there was, therefore, no evidence that the choice of marker gene had an impact on the assessment of C/EBPα TA activity in the assay. Therefore, additional controls and test mutant constructs were all made with only the MSCV.I.GFP vector.

Whilst the WT construct strongly increased the firefly luminescence, either mutant construct alone, equivalent to a homozygous mutation, had significantly reduced TA activity of 12% ± 0.4% of the WT activity for the N mutant and 27% ± 7% for the C mutant construct ($P < .0001$ and $P = .002$, respectively; Figure 6.10). As given above, the combination of the N and C mutant constructs, equivalent to the most common situation observed in *CEBPA*-double patients, also had significantly reduced TA activity of 15% ± 1% and 14% ± 1% of the WT activity for the different marker gene combinations ($P < .0001$ and $P = .002$, respectively; Figure 6.10). These results are consistent with loss-of-function mutations leading to almost complete loss of WT TA activity.

To create a situation equivalent to a heterozygous mutation in a *CEBPA*-single case, the N and C mutant constructs were also each mixed 1:1 with the WT construct. In this context, the TA activity of the N-terminal mutant was 61% ± 2% and 65% ± 3% of the WT activity, depending on the marker gene combination ($P = .0004$ and $P = .001$, respectively; Figure 6.10 and Table 6.4). However, this activity was still significantly higher than the N construct alone ($P < .0001$), with no evidence for dominant-negative activity over WT protein (Figure 6.10). By contrast, the TA activity of the C mutant construct mixed with the WT construct was not different from the WT construct alone, with 103% ± 6% and 99% ± 5% of the WT activity, depending on the marker gene combination ($P = .6$ and $P = .9$, respectively; Figure 6.10 and Table 6.4).

The results from the control constructs indicated that the dual-luciferase assay was appropriate for assessing C/EBPα TA activity. The control construct combinations equivalent to a homozygous N-terminal mutation and to biallelic classical N- and C-terminal mutations had the lowest TA activity compared to WT C/EBPα, with each quantified as

Table 6.4 Transactivation activity of *CEBPA* control constructs

***CEBPA* construct 1 (GFP)**	***CEBPA* construct 2 (eBFP2)**	Mean % of WT activity	SEM, %	*P* vs. WT	*P* vs. WT+N	*P* vs. WT+C
Control constructs:						
WT	WT	100	0	N.A.		
N	N	12	0.4	<0.0001*		
C	C	27	7	0.002*		
N	WT	61	2	0.0004*		
WT	N	65	3	0.001*		
C	WT	103	6	0.6		
WT	C	99	5	0.9		
N	C	15	1	<0.0001*		<0.0001*
C	N	14	1	<0.0001*	<0.0001*	
Additional controls:						
6 bp polymorphism	None	91	11	0.5		
	WT	87	4	0.1		
	N	56	4	0.007*	0.1	
	C	89	8	0.3		0.3
K304_Q305insL	None	36	7	0.01*		
	WT	100	6	0.99		
	N	14	2	0.0006*	<0.0001*	

**P*-value significant (< 0.05). Abbreviations: N.A., not applicable; WT, wild-type; SEM, standard error of the mean; N, Q83fs; C, K313dup.

between 12% and 15% of the WT activity (Figure 6.10). The combination equivalent to a homozygous C-terminal mutation also strongly reduced the TA activity by nearly three quarters compared to the WT activity, although not to the same extent as the N alone or N+C combinations. By contrast, the combinations equivalent to *CEBPA* alleles in *CEBPA*-single patients did not cause such striking reductions in TA activity. The C-terminal mutant did not impact on the C/EBPα TA activity when mixed 1:1 with WT construct and, whilst a significant reduction in activity was observed for the N-terminal mutant construct co-transfected with WT as compared to WT activity, this was not to the same extent as the N mutant construct alone (Figure 6.10).

6.3.3 Additional control constructs

Two additional control *CEBPA* constructs were investigated in this study. The first was the known 6 bp polymorphism (H195_P196dup). 293T cells transfected with this construct expressed both the p30 and p42 isoform proteins by immunoblotting (Figure 6.11, last lane). The polymorphism was tested in dual-luciferase assays on its own, as well as in combination with the WT, N and C constructs, because as a polymorphism it could occur in all of these allelic contexts. No significant difference in TA activity was observed between the polymorphism alone (91% ± 11% of the WT activity [mean ± SEM of three experiments], $P = .5$) nor the combination with either the WT (87% ± 5%, $P = .1$) or C (89% ± 8%, $P = .3$) constructs when compared with the WT control (Table 6.4 and Figure 6.12). Whilst the polymorphism mixed with the N construct had significantly reduced TA activity versus the WT control, this could simply be attributed to the N-terminal mutation, with no significant difference between the polymorphism+N and the WT+N construct combinations (56% ± 4% and 65% ± 3%, respectively, $P = .1$; Table 6.4). These results indicate that the 6 bp duplication polymorphism does not have a significant impact on C/EBPα TA activity.

The second additional control construct was a C-terminal in-frame insertion that affected the DBD, rather than the LZD (K304_Q305insL; Table 6.1). Both the p30 and the p42 isoforms were expressed from this construct by Western blotting (Figure 6.11). As with the LZD C-terminal control mutation (C), the DBD mutation K304_Q305insL was tested alone and mixed with the WT or N construct (Table 6.3). The K304_Q305insL mutant C/EBPα had a very similar impact on TA activity to the C mutant construct, with 36% ± 7% of the WT activity in the homozygous state ($P = .01$ versus WT activity) and 14% ± 2% in combination with the N construct ($P = .0006$) (Table 6.4 and Figure 6.12). No impact on TA activity was observed when the K304_Q305insL construct was mixed with the WT construct (100% ± 6.4% of the WT activity, $P = .99$; Figure 6.12). Given the similarity in TA activity in the

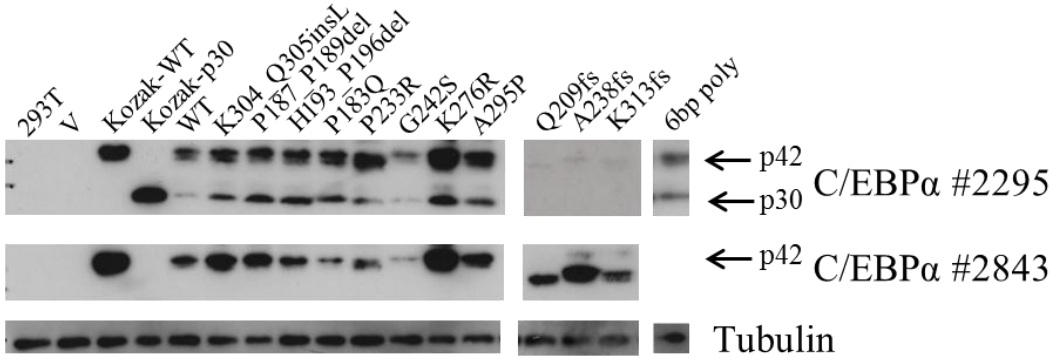


Figure 6.11 C/EBP*α* protein expression from test *CEBPA* constructs.
Western blots of lysates made from 293T cells either untransfected or 24 hours post-transfection with MSCV.I.GFP vector alone (V) or MSCV.I.GFP *CEBPA* constructs. Last lane 6 bp poly is the H195_P196dup polymorphism.

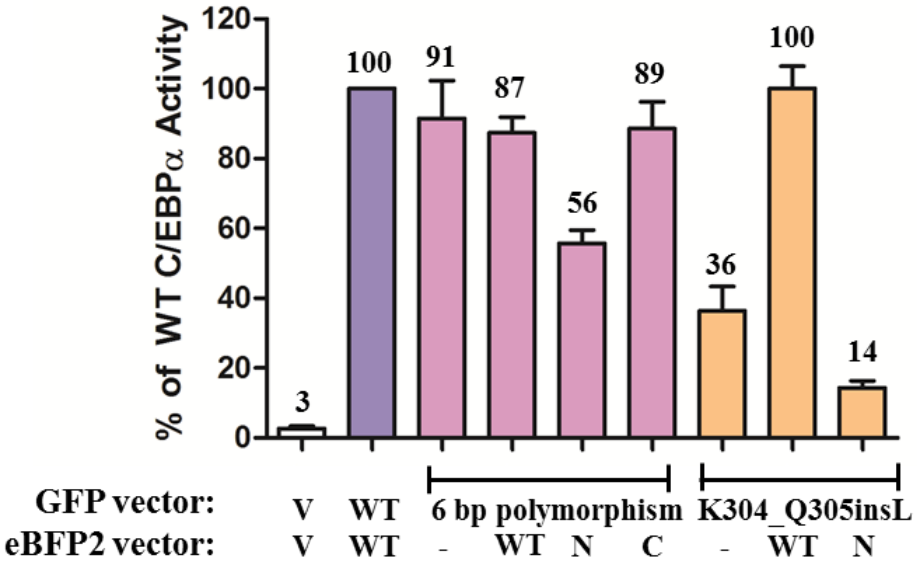


Figure 6.12 Transactivation activity of additional control C/EBPα expression constructs. Dual-luciferase assay results from 293T cells 24 hours after co-transfection with a 1:1 mix of MSCV.I.GFP and MSCV.I.eBFP2 either vector alone (V) or constructs containing control and additional control *CEBPA* sequences. The mean and standard error of the mean from 3 experiments is shown. Abbreviations: WT, wild-type *CEBPA*; N, Q83fs; C, K313dup.

different construct combinations between the K304_Q305insL mutant in the DBD and the K313dup (C) mutant in the LZD, only the LZD mutant (C) was used for the investigation of the atypical mutations.

6.3.4 Atypical mutations from *CEBPA*-double cases

In the cohort of patients presented in chapter 3, there were 14 mutations identified in *CEBPA*-double cases that were neither heterozygous out-of-frame insertions/deletions in the N-terminus nor in-frame insertions/deletions in the C-terminus. Nine of these atypical mutations occurred in patients who also had a classical N- or C-terminal mutation, three were homozygous mutations in the DBD or LZD, and the remaining two mutations were identified in the same patient. As described in section 6.2.1, five mutations were selected for functional analysis to investigate whether *CEBPA*-double patients with such atypical mutations had similar C/EBPα TA activity to those with two classical mutations. The selected mutations and the patients they were identified in are given in Table 6.1 and Table 6.2. Four of these mutations were an out-of-frame deletion in the middle of the gene (Q209fs), an out-of-frame insertion in the middle (A238fs), an out-of-frame insertion in the LZD (K313fs) and a missense mutation in the DBD (A295P). Each of them occurred in just one *CEBPA*-double case in the cohort studied and were combined with an out-of-frame insertion/deletion in the N-terminus (Table 6.3). The final mutation, K276R, was located just 5′ of the DBD and was identified in both a *CEBPA*-single case, and a *CEBPA*-double case together with an out-of-frame deletion in the middle of the gene.

Strong p42 and p30 C/EBPα isoform protein expression was seen from both the K276R and A295P constructs (Figure 6.11). For the Q209fs, A238fs and K313fs constructs, probing with the #2843 C/EBPα antibody that has an epitope within the N-terminus upstream of the p30 translation initiation site, proteins of a slightly smaller size than the p42 control band were identified, but no signal was detected when probing with the #2295 C/EBPα antibody that has an unknown epitope but detects both the p30 and p42 isoforms (Figure 6.11). These results are consistent with frame-shift/truncating mutations if the antibody epitope is after amino acid K313. *In silico*, all three constructs were predicted to produce a peptide of approximately 33kDa, in line with the immunoblotting results.

In order to test the TA activity of these atypical mutations in a context equivalent to that found in the patients, the Q209fs, A238fs and K313fs constructs were all tested in combination with the N construct, and the K276R construct was mixed with both the WT construct, as in patient 86, and with the Q209fs to substitute for L178fs, as in patient 71

(Table 6.1, Table 6.2). Four of these construct combinations (Q209fs, A238fs, K313fs and A295P with the N construct), had very low TA activity of 9% ± 1%, 10% ± 1%, 12% ± 1% and 15% ± 3% of the WT C/EBPα activity, respectively (mean ± SEM of three experiments; Figure 6.13). These TA activities were highly significantly reduced versus both the WT activity ($P \leq .002$ for all) and versus the 65% ± 3% observed for the WT+N construct combination ($P \leq .0001$ for all; Table 6.5). These results indicated that they were behaving as classical pathogenic mutations in this assay. By contrast, the combination of K276R+WT constructs increased C/EBPα TA activity to 129% ± 14% and of the K276R+Q209fs constructs to 155% ± 25% of the WT construct activity (Figure 6.13). These were not significantly different from the WT activity ($P = .2$ for each), and indicated that K276R mutant C/EBPα can still dimerise, bind to DNA and activate transcription as least as well as WT C/EBPα. These results have implications for the classification of patient 71 as a *CEBPA*-double case.

6.3.5 Atypical mutations from *CEBPA*-single cases

Of the 48 *CEBPA*-single patients in the present cohort, 28 (58%) had non-classical mutations. Of these, five were selected for functional analysis in order to further explore the functional impact of a range of *CEBPA* sequence alterations (Table 6.1). The mutations selected were two in-frame deletions in TAD2 (P187_P189del and H193_P196del) and three missense mutations, one in TAD2 (P183Q) and two between TAD2 and the DBD (P233R and G242S). The P187_P189del mutation was identified in two patients and H193_P196del in a single case (Table 6.2). These were of interest because the common in-frame H195_P196dup polymorphism is in the same region, demonstrating that amino acid changes in that location do not necessarily affect protein function. P183Q and P233R were each detected in a single case, while G242S was identified in two *CEBPA*-single patients, and these missense mutations all occurred outside of the C-terminal DBD and LZD (Table 6.2).

Western blotting showed that both the p30 and p42 isoforms were expressed from all of the constructs, although G242S was expressed less strongly than the other mutant constructs (Figure 6.11). To explore the functional impact of these mutations in a range of allelic contexts, each was tested by co-transfection with either the WT, N or C control constructs (Table 6.3). The results of these assays are shown in Figure 6.14 and Table 6.5. Of the five mutations, three appeared to have no significant impact on C/EBPα TA function when combined with the WT or the C control constructs: H193_P189del, P183Q and G242S. For each of these constructs in combination with the WT construct the TA activity was 104% ± 8% ($P = .7$ vs. WT construct alone), 98% ± 11% ($P = .9$) and 99% ± 5% ($P = .9$) of the WT

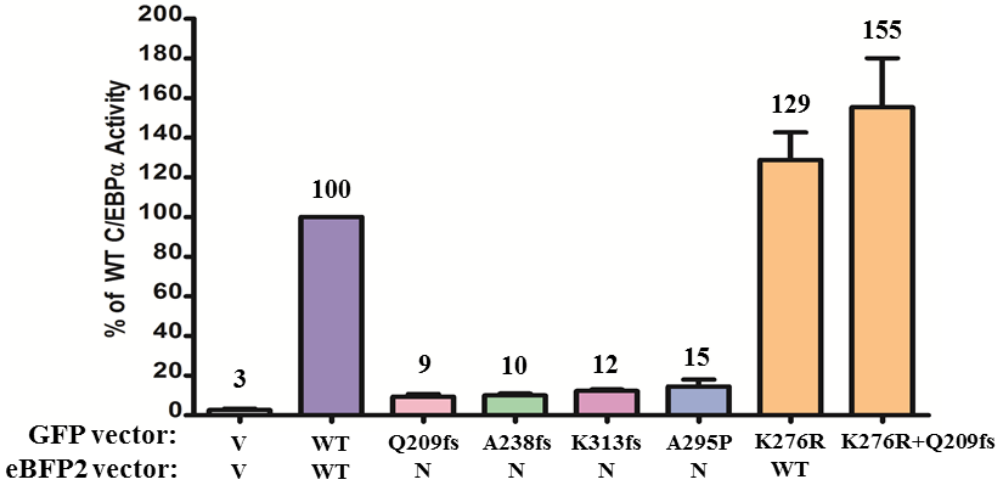


Figure 6.13 Transactivation activity of test mutants from *CEBPA*-double patients
Dual-luciferase assay results from 293T cells 24 hours post-transfection with a 1:1 mix of vector alone (V) or *CEBPA* constructs. The mean and standard error of the mean from 3 experiments is shown. Abbreviations: WT, wild-type *CEBPA*; N, Q83fs; C, K313dup.

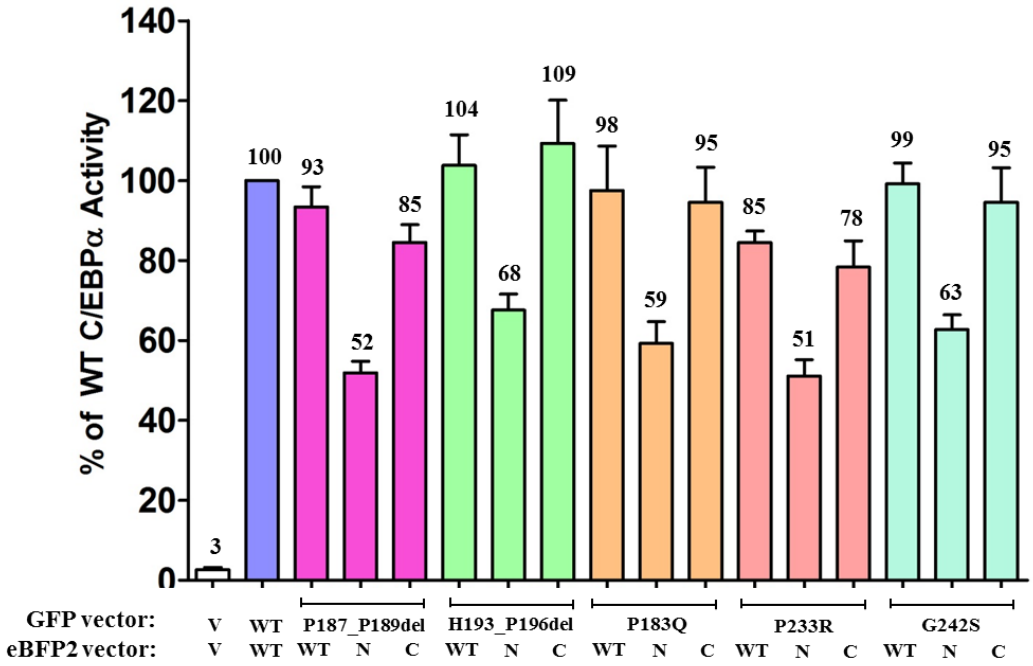


Figure 6.14 Transactivation activity of test mutants from *CEBPA*-single patients.
Dual-luciferase assay results from 293T cells 24 hours post-transfection with either a 1:1 mix of vector alone (V) or *CEBPA* constructs. The mean and standard error of the mean of 3 experiments is shown. Abbreviations: WT, wild-type *CEBPA*; N, Q83fs; C, K313dup.

Table 6.5 Transactivation activity of *CEBPA* constructs with atypical mutations

***CEBPA* construct 1 (GFP)**	***CEBPA* construct 2 (eBFP2)**	Mean % of WT activity	SEM, %	*P* vs. WT	*P* vs. WT+N	*P* vs. WT+C
Atypical mutations in *CEBPA*-doubles:						
Q209fs	N	9	1	0.0002*	<0.0001*	
A238fs	N	10	1	0.0001*	<0.0001*	
K313fs	N	12	1	<0.0001*	<0.0001*	
A295P	N	15	3	0.002*	0.0001*	
K276R	WT	129	14	0.2		
	Q209fs (GFP)	155	25	0.2		
Atypical mutations in *CEBPA*-singles:						
P187_P189del	WT	93	5	0.3		
	N	52	3	0.004*	0.03*	
	C	85	4	0.07		0.09
H193_P196del	WT	104	8	0.7		
	N	68	4	0.02*	0.6	
	C	109	11	0.5		0.4
P183Q	WT	98	11	0.9		
	N	59	6	0.02*	0.4	
	C	95	9	0.6		0.7
P233R	WT	85	3	0.03*		
	N	51	4	0.007*	0.04*	
	C	78	7	0.08		0.04*
G242S	WT	99	5	0.9		
	N	63	4	0.01*	0.6	
	C	95	9	0.6		0.6

**P*-value significant (< 0.05). Abbreviations: WT, wild-type; SEM, standard error of the mean; N, Q83fs; C, K313dup.

C/EBPα activity, and in combination with the C construct the TA activity was 109% ± 11% (*P* = .5 vs. WT construct alone), 95% ± 9% (*P* = .6) and 95% ± 9% (*P* = .6), for H193_P189del, P183Q and G242S, respectively. Although a significant reduction in TA activity was seen for each of these mutants co-transfected with the N construct as compared with WT (68% ± 4% [*P* = .02], 59% ± 5% [*P* = .02] and 63% ± 3% [*P* = .01], respectively), this is attributable to the N mutant construct, as these differences were lost in comparison to the WT+N construct combination (*P* = .6 for all).

There was some evidence that the remaining atypical mutants, P187_P189del and P233R, caused a slight reduction in C/EBPα TA activity. The TA activity of the P233R missense mutant combined with the WT construct was 85% ± 3% of the WT C/EBPα activity (*P* = .03 vs. WT alone), 51% ± 4% in combination with the N construct (*P* = .04 vs. WT+N) and 78% ± 7% in combination with the C construct (*P* = .04 vs. WT+C). The P187_P189del did not significantly impact on TA activity in combination with the WT construct, with 93% ± 5% of the WT C/EBPα activity (*P* = .3 vs. WT alone), but did cause a significant reduction in TA activity when co-transfected with the N construct with 52% ± 3% activity (*P* = .03 vs. WT+N) and a trend to reduced activity in combination with the C construct (85% ± 5%, *P* = .09 vs. WT+C). However, the TA activity of the P233R and P187_P189del mutants when combined with an N or C control mutant was much closer to the activity of the control mutant mixed with WT than the control mutants either alone or in combination with each other, and therefore these mutations did not appear to impact on TA activity as strongly as the control mutants and could be the result of technical variation.

The data on the impact of all the *CEBPA* variants investigated in this study on C/EBPα TA activity is summarised in Table 6.6. The reduction in TA activity for a particular combination versus the WT construct in that context was calculated and the condition assigned to one of four groups: no difference, a reduction of less than 20% that was nonetheless still significant, a significant reduction of between 20% and 40%, or greater than 40% of the WT activity.

Table 6.6 Summary of the impact of *CEBPA* sequence alterations on C/EBPα transactivation activity

	***CEBPA* status modelled**		
***CEBPA* construct**	***CEBPA*-single +WT**	***CEBPA*-double +N**	***CEBPA*-double +C**
Impact versus:	**WT alone**	**WT + N**	**WT + C**
Controls:			
N	↓↓	↓↓↓	↓↓↓
C	↔	↓↓↓	↓↓↓
6 bp polymorphism	↔	↔	↔
K304_Q305insL	↔	↓↓↓	
Atypical mutations: *CEBPA*-doubles			
Q209fs		↓↓↓	
A238fs		↓↓↓	
K313fs		↓↓↓	
A295P		↓↓↓	
K276R	↔	↔	
Atypical mutations: *CEBPA*-singles			
P187_P189del	↔	↓	↓
H193_P196del	↔	↔	↔
P183Q	↔	↔	↔
P233R	↓	↓	↓↓
G242S	↔	↔	↔

Abbreviations: WT, wild-type; N, Q83fs; C, K313dup.
Arrows indicates the impact on C/EBPα activity in the luciferase reporter assay of the test vs. a WT construct in that particular context, as indicated. No significant impact is indicated by "↔" and significantly reduced activity by "↓", with the number of arrows demonstrating the strength of this reduction (↓, < 20%; ↓↓, 20% - 40%; ↓↓↓, > 40%). *P*-values < .05 were considered significant.

6.4 Discussion

The impact of classical N- and C-terminal *CEBPA* mutations on C/EBPα protein function has been explored in both *in vitro* and *in vivo* models. By contrast, the consequence of non-classical mutations is less well understood. Non-classical sequence alterations constituted over one quarter of the mutations detected in the patient cohort described in chapter 4, and therefore the impact of selected nucleotide changes on C/EBPα TA activity was explored using a luciferase reporter assay.

The entire *CEBPA* coding sequence and 165 bp of the endogenous 5′ UTR was amplified from gDNA samples containing WT, a classical N-terminal mutation (Q83fs) or a classical C-terminal mutation (K313dup) and these amplicons were cloned into both MSCV.I.GFP and MSCV.I.eBFP2. These control constructs were used to optimise a luciferase reporter assay to assess the ability of C/EBPα variants to activate transcription of a luciferase gene from a target sequence (Figure 6.1). A luciferase reporter assay was selected for this study as this has been used successfully for this purpose in a number of previous reports (Pabst *et al*, 2001b;Gombart *et al*, 2002;Cleaves *et al*, 2004;Pabst *et al*, 2009;Kato *et al*, 2011).

The results of luciferase reporter assays need to be normalised across different samples for the variation in transfection efficiency. A *Renilla* expression construct was selected as an internal control for this purpose as both the firefly and *Renilla* luciferase activities could be measured simply in a single assay using the Dual-Luciferase Reporter Assay System (Promega) (Figure 6.4). There was also a requirement for the equal transfection of cells with two C/EBPα expression constructs, equivalent to the two different *CEBPA* alleles present in patients. To facilitate the evaluation of this, 293T cells were transfected with equal amounts of *CEBPA* constructs containing different fluorescent marker genes, either GFP or eBFP2. Transfected cells could be analysed by fluorescent microscopy or flow cytometry to determine the presence of both constructs in the same cells by detection of both green and blue fluorescence. An alternative strategy for ensuring 1:1 expression of two C/EBPα constructs could have been generation of a bicistronic expression vector encoding a 2A peptide sequence between two different *CEBPA* gene sequences, which would undergo cleavage to produce equal expression. However, having separate constructs allowed several different combinations of vectors to be tested without the requirement for further cloning.

Optimisation of transfections for the dual-luciferase assay was performed using the three C/EBPα control constructs (WT, N and C) transfected individually. Varying the amount of *CEBPA* construct in the transfection between 100ng and 460ng per well of a 24-well plate

did not have a striking impact on transfection efficiency, as determined by flow cytometric analysis for fluorescent marker gene expression (Figure 6.5). In dual-luciferase assays, increasing the amount of C/EBPα expression construct from 50ng to 200ng did cause a slight, but not significant, increase in the firefly luciferase activity normalised for the transfection efficiency (L/R) for each control (P = .16; Figure 6.6 A). Increasing the amount of the luciferase reporter construct also increased the L/R for a given C/EBPα expression construct, although this experiment was only performed once (Figure 6.6 B). The conditions selected for the transfections were: 1ng of the *Renilla* expression construct, 80ng of the luciferase reporter construct and a total of 100ng of the *CEBPA* construct(s) per well of a 24-well plate. Using these conditions, C/EBPα protein was strongly expressed in cells 24 hours after transfection and, whilst the WT and the C constructs produced both the p42 and the p30 isoforms, as expected only the p30 isoform could be detected in cells transfected with the N construct (Figure 6.7). In dual-luciferase assays, the background L/R of the vector alone was very low, whilst the WT MSCV.I.GFP caused a 41-fold increase in L/R over the vector alone, by contrast the N and C mutant constructs caused 4-fold and 9-fold increases, respectively (Figure 6.6 A). This reduction in TA activity for the control mutant constructs was in line with the expected results for loss-of-function mutations.

The selected transfection conditions were explored further in co-transfection experiments of equal amounts of MSCV.I.GFP and MSCV.I.eBFP2 control C/EBPα constructs to check for equal expression of the two constructs. By FACS analysis, an average of 32% of cells (range 22% to 39%) were GFP and/or eBFP2 positive 24 hours after transfection (Figure 6.8). Whilst approximately two thirds of the positive cells were both GFP and eBFP2 positive, the remaining transfected cells expressed GFP, but not eBFP2. The reasons for this were unknown, but there was no apparent effect on C/EBPα protein expression with no striking differences observed by Western blotting between cells transfected with the same combination of control C/EBPα constructs in the two possible marker gene configurations (Figure 6.9). Importantly, the marker gene did not impact on the results obtained in luciferase reporter assays (Figure 6.10). It was, therefore, assumed that marker gene selection did not affect the assay results.

In these co-transfection experiments, when the results were normalised to the WT C/EBPα activity, the N-terminal mutant alone, the C-terminal mutant alone or both together had strongly reduced TA activity compared to the WT activity (Table 6.4, Figure 6.10). These data are in line with several reports that classical N-terminal or C-terminal mutations cause substantial loss of TA activity either alone (Pabst *et al*, 2001b;Gombart *et al*, 2002;Kato *et al*, 2011), or in combination with each other (Pabst *et al*, 2009). These combinations were

equivalent to a homozygous N-terminal mutant, a homozygous C-terminal mutant, and biallelic N- and C-terminal mutations, respectively, and the TA activity was consistent with a substantial loss of normal WT function. Of the 59 *CEBPA*-double cases identified in chapter 3, the majority (83%) had both a classical N- and C-terminal mutation, only three (5%) had a homozygous C-terminal mutation and none had a homozygous N-terminal mutation. Across nine published cohorts studying a total of 2641 patients, only one (0.04%) patient with a homozygous N-terminal mutation has been reported (Pabst *et al*, 2001b;Snaddon *et al*, 2003;Frohling *et al*, 2004;Bienz *et al*, 2005;Marcucci *et al*, 2008;Wouters *et al*, 2009;Pabst *et al*, 2009;Renneville *et al*, 2009a;Dufour *et al*, 2010). Therefore, whilst similar TA activities were observed for the N alone, C alone and N plus C constructs, the classical N- plus classical C-terminal mutation is much more likely to be selected for during leukaemogenesis in patients, suggestive of roles for the different C/EBPα mutants beyond their impact on TA activity.

When the C-terminal mutant was mixed equally with WT, equivalent to a heterozygous C-terminal mutation in a *CEBPA*-single patient, there was no difference in activity as compared with the WT control alone (Figure 6.10). Similar results were obtained from a second classical C-terminal mutant, K304_Q305insL, located in the DBD rather than the LZD (Figure 6.12). This is in line with two other studies that also observed no reduction in TA activity for classical C-terminal mutants in combination with WT C/EBPα (Gombart *et al*, 2002;Kato *et al*, 2011), although a reduction in TA activity to 36% of the WT activity was reported in one study (Pabst *et al*, 2001b). By contrast, the N-terminal mutant mixed equally with WT, equivalent to a *CEBPA*-single case with an N-terminal mutation, reduced TA activity to approximately two thirds of the WT activity (Figure 6.10). This is equivalent to other studies in which TA activity has been reported for an N-terminal mutant, or the p30 isoform, in combination with WT (Pabst *et al*, 2001b;Gombart *et al*, 2002;Cleaves *et al*, 2004;Pabst *et al*, 2009;Kato *et al*, 2011). The magnitude of reduction in the present study was similar to that in one of these reports, although actual values were not available (Gombart *et al*, 2002). However, others have reported larger reductions to 16% of the WT activity in one study (Pabst *et al*, 2001b) and 18% in another (Pabst *et al*, 2009), and it has been suggested that the p30 isoform has a dominant-negative effect over the p42 isoform (Pabst *et al*, 2001b;Cleaves *et al*, 2004). However, there was no evidence for this effect in the present study. Comparisons between different studies can be problematic due to the use of different cell lines, expression vectors and ratios of WT to N-terminal mutant or p30 isoform constructs. This is particularly the case when the aim of the reported study was very different to that of the present work. For instance, for the purpose of exploring the relationship between the WT and the p30 isoform, the WT:p30 construct ratio was titrated

from 1:3 to 1:7 in one study (Cleaves *et al*, 2004) and 1:2 to 1:6 in another (Gombart *et al*, 2002), both observing a decrease in TA activity with increasing p30 construct. Overall, in the present studies, the TA activities of the control C/EBPα expression construct combinations were broadly in line with those reported in the literature, with the greatest loss of activity seen in combinations equivalent to homozygous or biallelic *CEBPA* mutations. These results indicated that the assay developed for these studies was appropriate for determining the impact of atypical *CEBPA* mutations on TA activity.

Five of the 14 non-classical mutations identified in the 59 *CEBPA*-double patients were selected for analysis, including two frame-shift/truncating mutations in the middle of the gene (Q209fs and A238fs) and one in the LZD (K313fs), all predicted to truncate both the p30 and the p42 isoforms, and two missense mutations, one just upstream of and one within the DBD (K276R and A295P, respectively; Table 6.1). Five representative atypical mutations were selected from the 28 *CEBPA*-single cases with non-classical alterations. These included two in-frame deletions (P187_P189del and H193_P196del) and three missense mutations in the middle of the gene (P183Q, P233R and G242S; Table 6.1). Mutations from *CEBPA*-double cases were all tested in combination with a *CEBPA* construct equivalent to the second mutation present in the patient they were identified in, whilst those from *CEBPA*-single cases were tested in combination with the WT, N and C constructs (Table 6.1 and Table 6.2).

All three frame-shift/truncating mutations (Q209fs, A238fs and K313fs) expressed a truncated C/EBPα peptide of a size consistent with that predicted *in silico* (Figure 6.11), showing that, *in vitro*, these mutations caused loss of both the normal p42 and the p30 isoforms. In combination with the N construct they all resulted in levels of TA activity similar to that of the classical N+C construct combination (Figure 6.13). These observations are not unexpected, given that each lacks either the entire DBD/LZD (Q209fs and A238fs) or a substantial portion of the LZD (K313dup). A similar mutation, R165fs, was investigated in one other report, where a strong increase in TA activity was observed (Pabst *et al*, 2001b). This result is unexpected and no explanation was given.

The two in-frame deletions (P187_P189del and H193_P196del) affected similar residues to the 6 bp duplication polymorphism (H195_P196dup), which did not appear to have a detrimental effect on activity (Figure 6.12). In line with this, H193_P196del had no impact on activity in this assay (Figure 6.14). P187_P189del did cause a slight, but significant, reduction in combination with either an N- or C-terminal mutation (Table 6.5), but the activity was much closer to that of the N or C construct in combination with the WT than the

activity of the N+C construct combination (Table 6.4). These data suggest that in-frame size changes in this region of *CEBPA* do not impact on TA activity.

In total, five missense mutations were tested, four located in the middle of the gene (K276R, P183Q, P233R and G242S) and one located in the C-terminal DBD (A295P).Using PolyPhen-2 prediction software (http://genetics.bwh.harvard.edu/pph2/), three were assessed as probably damaging (A295P, K276R and P233R), P183Q as possibly damaging and G242S as benign. In line with the prediction, the G242S mutation did not have any impact on TA activity in combination with any of the three control constructs (Figure 6.14 and Table 6.5). This was not unexpected given that this residue is not located in either the DBD or LZD. The A295P mutation, detected in a *CEBPA*-double case, was only tested in combination with the N mutant construct, which resulted in TA activity equivalent to that of the N+C combination, compatible with a prediction of probably damaging to C/EBPα (Figure 6.13). This missense mutation changes a residue in the DBD functional domain of C/EBPα and it is, therefore, not surprising that a loss of TA activity was observed. Although there was some evidence that the P233R mutation, assessed as probably damaging, slightly reduced TA activity when in combination with the control constructs (Figure 6.14 and Table 6.5), the TA activity of the P233R+N or P233R+C construct combinations remained significantly higher than that of the classical N+C combination ($P = .01$ and $P = .01$, respectively). These observations are not unexpected given that the amino acid substitution is located in a region between known functional domains, TAD2 and the DBD. Despite the predictions of probably and possibly damaging, K276R and P183Q did not negatively affect TA activity when combined with the control constructs (Table 6.5). The P183Q substitution is located in TAD2, the same domain as the 6 bp duplication polymorphism, rather than the DBD/LZD and it is not unexpected that no impact on TA activity was observed. The K276R mutation is located only two residues upstream of the DBD, according to one definition of this region (Figure 3.1). However, there was no evidence that K276R mutant C/EBPα had impaired TA activity as compared to the WT construct, in line with its location outside of the DBD.

The evidence presented above indicates that in-frame insertions/deletions or missense mutations in the middle region of *CEBPA* do not prevent C/EBPα binding to a target DNA sequence as a dimer and activating transcription. These types of atypical mutations were detected in nine cases (0.6%) in the cohort described in chapter 3 and they could either have no functional consequence or they could affect functions of C/EBPα that were not assayed here, and so be selected for during leukaemogenesis. For instance, the middle portion of *CEBPA* contains a region that mediates the interaction between C/EBPα and the cyclin-

dependent kinases cdk2 and cdk4, which are important for cell cycle progression (Wang *et al*, 2001). In addition, another region in TAD2 is required for C/EBPα to interact with the SWI/SNF chromatin remodelling complex, an interaction which was required for C/EBPα to suppress cell proliferation (Muller *et al*, 2004). In theory, mutations in these regions could disrupt these interactions and lead to an increase in cell proliferative capacity and further experiments are required to determine if this is the case.

In conclusion, whilst there have been several *in vitro* and *in vivo* studies of the impact of classical N- and C-terminal *CEBPA* mutations on C/EBPα function, similar data for non-classical mutations is very limited. Therefore, the impact of ten atypical mutations on TA activity was investigated in this chapter. In general, it was found that mutations affecting the DBD or LZD caused loss of TA function equivalent to the classical mutations, whether by an amino acid substitution or by truncation of the C-terminus. By contrast, mutations affecting the middle of the gene, between the translation initiation site for the p30 isoform and the DBD, did not strongly impact on TA activity. Remission samples were not available for the majority of *CEBPA*-mutant cases in this study, but may be helpful in order to determine whether sequence changes are germline or acquired only in the leukaemic cells. However, as pathological germline *CEBPA* mutations have been reported in cases of familial AML (Smith *et al*, 2004;Pabst *et al*, 2008;Renneville *et al*, 2009b), a decision would still need to be made regarding whether a particular sequence change was pathogenic. Constitutional *CEBPA* mutations in pedigrees with AML have all been classical N-terminal alterations. A small number of cases without a family history of myeloid malignancies have been reported as having germline mutations located outside of the N-terminus, including three missense mutations in the C-terminus (Taskesen *et al*, 2011). One of these cases was *CEBPA*-single and two were *CEBPA*-double, with an additional classical N-terminal mutation in one and a classical C-terminal mutation in the other. Moreover, the two *CEBPA*-double cases clustered with other *CEBPA*-double cases by GEP (Taskesen *et al*, 2011). Reports of this kind underline the importance of further functional study of non-classical mutations.

This data has implications for the classification of several patients included as *CEBPA*-mutant in this study. For instance, if K276R is a non-pathogenic passenger alteration, then patient 71 may more appropriately be classified as *CEBPA*-single, rather than *CEBPA*-double, and patient 86 as *CEBPA*-WT (Table 6.2). It should be noted that outcome analyses were performed excluding patients with missense mutations and in-frame size changes in the middle of the gene, and the favourable prognosis of *CEBPA*-double patients was maintained (Figure 4.2 D and Appendix Table 2).

CHAPTER 7: CONCLUSIONS AND FUTURE DIRECTIONS

The clinical outcome of younger adult patients with AML is highly heterogeneous and there are a number of clinical and molecular features that can be utilised to define groups of patients with distinct prognoses (chapter 1). These prognostic factors are useful for developing risk-adapted treatment strategies for personalised therapy, where the most intensive and/or experimental therapies can be targeted at patients with high-risk, rather than those with low-risk, features. This is particularly pertinent for the debate regarding which patients should receive an allo-SCT, for instance there is evidence that the benefit of this treatment outweighs the risk of TRM in patients with adverse-risk, but not favourable-risk, cytogenetics (Cornelissen *et al*, 2007). However, there is also heterogeneity within established risk-groups and questions remain regarding a number of the currently identified molecular markers, particularly concerning the interaction between different markers, and greater understanding is required. Knowledge of these markers can also be important for understanding the underlying biology of the leukaemia and thereby aid in the development of novel treatment strategies, for instance as with FLT3 inhibitors in patients with activating *FLT3* mutations.

This thesis presents an investigation of *CEBPA* mutations, which are known to be a prognostic factor in AML. The incidence of mutations was determined in a large cohort of young adult patients treated in the UK MRC AML 10 and AML 12 trials (chapter 3), and the biological characteristics and clinical outcome of the *CEBPA*-mutated patients were explored in chapter 4, along with factors which modify this impact. The mutations identified in patient samples were then further investigated with regard to their mutant level and allelic distribution (chapter 5), and the impact of non-classical mutations on C/EBPα function in a luciferase reporter assay (chapter 6).

A method was developed to detect *CEBPA* mutations across the whole coding sequence using dHPLC analysis of three overlapping PCR products. This technique was shown to be effective at detecting both point mutations and size changes in all three fragments, and both heterozygous and homozygous mutations were identified in a large cohort of 1427 patients, and at a comparable frequency to other studies (chapter 3). Overall, 107 (7%) patients were classified as *CEBPA*-mutant, including 48 cases with one heterozygous mutation, *CEBPA*-single, and 59 with two heterozygous or one homozygous mutation, *CEBPA*-double. The types and locations of mutations differed between *CEBPA*-single and *CEBPA*-double

patients. Mutations in *CEBPA*-single cases were distributed across the entire gene, whilst those in *CEBPA*-double cases clustered quite tightly in the N- and C-terminal hotspots, with most cases having one of each of these types (Figure 3.7 and Table 3.3).

There were a number of characteristics in common between *CEBPA*-single and *CEBPA*-double cases as compared with the *CEBPA*-WT group, such as an association with FAB types M1 or M2 and with intermediate-risk cytogenetics, in particular a normal karyotype (Table 4.3). However, there were several striking differences including younger age and a lack of *NPM1* and *FLT3*/ITD mutations in *CEBPA*-double cases, but not *CEBPA*-single patients. Importantly, there were significant differences in clinical outcome at 8 years when patients were stratified by *CEBPA* status. The presence of *CEBPA*-double mutations was an independent favourable factor for OS, RFS and relapse in multivariate analysis (Table 4.5), while the outcome of *CEBPA*-single cases was similar to that of *CEBPA*-WT patients (Table 4.4). At the time this work was completed, only three other reports, all in smaller cohorts, had analysed patients by *CEBPA*-WT, *CEBPA*-single and *CEBPA*-double status, with two groups reporting that the benefit for *CEBPA* mutations was restricted to *CEBPA*-double cases (Wouters *et al*, 2009;Pabst *et al*, 2009), and one finding little difference in outcome between *CEBPA*-single and *CEBPA*-double cases (Table 4.2) (Renneville *et al*, 2009a). However, the data from the present study together with three subsequent reports on different patient groups (Hou *et al*, 2009;Dufour *et al*, 2010;Taskesen *et al*, 2011) all confirm a favourable prognosis in *CEBPA*-double, but not *CEBPA*-single, patients, compared with those lacking *CEBPA* mutations.

A further important consideration was the potential interaction between *CEBPA* mutations and other known characteristics. Whilst the favourable outcome of *CEBPA*-double patients could be observed in both intermediate-risk AK and NK patients, it was lost in the presence of a *FLT3*/ITD (Figure 4.4 A and B). In the whole cohort just seven (0.5%) patients were both *CEBPA*-double and *FLT3*/ITD-positive, however these had a very poor outcome with only one case alive 3 years after trial entry. The strong negative impact of a *FLT3*/ITD observed in these *CEBPA*-double cases is in line with recent data from mouse models where a *FLT3*/ITD has been co-expressed with mutant *CEBPA*. Reckzeh *et al* (2012) generated a knock-in model where biallelic *CEBPA* mutations (p30 isoform and K313dup) and a *FLT3*/ITD were co-expressed from their endogenous promoters and found that the presence of a *FLT3*/ITD shortened the latency of mutant *CEBPA*-driven leukaemogenesis, with median survival reduced from 33.8 to 22.5 weeks. In addition, Kato *et al* (2011) transduced murine BM mononuclear cells with retroviruses to co-express a C-terminal C/EBPα mutant together with a *FLT3*/ITD and transplanted them into syngeneic mice. The recipient mice

developed leukaemia in a much shorter period of time than mice transplanted with cells expressing the mutant C/EBPα alone (2 to 3 weeks vs. 4 to 12 months).

The potential interactions between single *CEBPA* mutations and other characteristics were also investigated in the studies presented here, and there was no evidence that the presence of a *CEBPA*-single mutation further improved the favourable outcome of an *NPM1* mutation (Figure 4.4 C). Taken together, these data suggest screening for *CEBPA* mutations at disease presentation should only be performed in patients with intermediate-risk karyotypes without a *FLT3*/ITD or *NPM1* mutation, which would have limited *CEBPA* mutation analysis to only a third of patients in the current cohort. There are an ever-increasing number of recurrently mutated genes identified in AML and identifying which markers should be screened in the molecular work-up of new AML patients will be increasingly challenging. Therefore, studies of this kind are important to aid in rationalisation of which markers should be analysed at diagnosis.

The mutant level of 101 mutations from 75 patients was quantified by fragment analysis of fluorescently labelled PCR products (chapter 5). The level of the majority of mutations was consistent with the level expected for a heterozygous mutation present in nearly all cells, suggesting that the *CEBPA* mutation was acquired early in leukaemogenesis in these patients. A low *CEBPA* mutant level (< 25%) was observed for only five mutations from five different patients, and in one of these patients an *NPM1* mutant level of 49% had been identified, indicating that acquisition of a *CEBPA* mutation is not always an early event. In *CEBPA*-double cases, paired mutant levels were highly correlated in the 26 patients where both mutant levels were available, and cloning of *CEBPA* alleles in 13 patients confirmed the presence of biallelic mutations. Taken together, these data indicated that in all probability the different mutations were on different alleles in the same cells in *CEBPA*-double cases.

The majority of mutations identified in the current cohort were classical N-terminal (out-of-frame insertions/deletions or nonsense mutations in the N-terminus) or classical C-terminal (in-frame insertions/deletions in the DBD or LZD), and there have been several comprehensive studies investigating the *in vitro* and *in vivo* consequences of these types of mutations. However, non-classical mutations accounted for over a quarter of the non-synonymous sequence alterations identified, and the impact of non-classical mutations on protein function remains unclear. Understanding whether particular sequence alterations are pathogenic changes or non-functional bystanders is important if *CEBPA* status is to be used to aid clinical decisions. Therefore a luciferase reporter assay was developed to test the

transactivation activity of selected non-classical mutations, where two *CEBPA* constructs were co-transfected into 293T cells with the reporter construct (chapter 6). Three control C/EBPα expression constructs were used to test the assay: WT, a classical N-terminal mutation (Q83fs, N) and a classical C-terminal mutation (K313dup, C).Whilst the WT construct alone strongly activated transcription of the reporter luciferase construct, control mutant construct combinations equivalent to biallelic or homozygous *CEBPA* mutations in *CEBPA*-double patients (N alone, C alone or both together) all had a transactivation activity less than one third that of the WT activity. In co-transfections with an equal amount of the WT construct, equivalent to a heterozygous mutation in a *CEBPA*-single case, the N construct had approximately two thirds of the WT activity, however the C construct combination did not impact on transactivation activity in this assay. Five atypical mutations identified in *CEBPA*-double cases and five from *CEBPA*-single patients were selected for analysis. Four mutations from *CEBPA*-double cases in combination with the N construct had transactivation activities equivalent to the N+C control construct combination. These included two frame-shift/truncating insertions/deletions in the middle of the gene and one in the LZD, and a missense mutation in the DBD. By contrast, three missense mutations (two from *CEBPA*-single cases and one detected in both a *CEBPA*-single and *CEBPA*-double case) and an in-frame deletion (from a *CEBPA*-single case), all in the middle of the gene, had no impact on transactivation activity in this assay. There was some evidence that the remaining two atypical mutations, an in-frame deletion and a missense mutation in the middle of the gene (both from *CEBPA*-single cases), had a slight negative impact on C/EBPα transactivation activity, but not to the same extent as the control mutants. These data, whilst preliminary, highlight the challenges present in predicting the functional consequences of sequence alterations.

7.1 Future directions

If *CEBPA* mutant status is to be included in the molecular investigation of newly diagnosed AML patients, particularly of intermediate-risk karyotype patients without a *FLT3*/ITD or *NPM1* mutation, a robust method of mutation detection is required. There are a number of different techniques that have been used to identify *CEBPA* mutations, including Sanger sequencing and fragment analysis for size changes (chapter 3). However, there are caveats to these techniques. For instance, fragment analysis does not detect mutations that do not alter amplicon length, which accounted for 14% of the mutations in this study. Sanger sequencing is relatively labour intensive and can lack sensitivity for low level mutations. In addition, the high GC content of the *CEBPA* coding sequence (75%) makes optimisation of PCR and

sequencing protocols challenging. For the large retrospective cohort studied here, dHPLC analysis was an effective method for identifying *CEBPA* mutations. However, it should be noted that the WAVE platform is not widely available, can be technically demanding, and the sensitivity of detection depends on the particular sequence involved. Next generation sequencing techniques are being explored for targeted mutation analysis, particularly in multiplexed assays of a number of genes. As with conventional PCR amplification followed by Sanger sequencing, the high GC content of *CEBPA* can make amplification steps challenging, reducing the number of reads obtained. For instance, in a study of whole genome amplification by multiple displacement amplification, the median GC content of regions where amplification failed was 60% (Hou *et al*, 2012), whilst the *CEBPA* coding sequence has a GC content of 75%. More specifically, 454 deep sequencing of *CEBPA* in patient samples has been explored in one study (Grossmann *et al*, 2011). Using a standard protocol, amplification failed for two of four overlapping fragments across the entire *CEBPA* coding sequence of two test samples, although these difficulties could be overcome after optimisation of the amplification protocols (Grossmann *et al*, 2011). This highlights the difficulties of developing novel assays for the detection of *CEBPA* mutations. However, streamlining the molecular work-up of samples in the future is desirable, and the inclusion of *CEBPA* in targeted multiplexed NGS-based assays is important due to the impact of mutations on patient prognosis.

The data presented in this thesis, together with observations from several other studies (Wouters *et al*, 2009;Pabst *et al*, 2009;Hou *et al*, 2009;Dufour *et al*, 2010;Taskesen *et al*, 2011), has indicated a favourable prognosis for *CEBPA*-double patients. In the present cohort the outcome of *CEBPA*-double cases with a *FLT3*/ITD was very poor (chapter 4), however this is based on a small number of patients and requires further confirmation. It has not yet been explored in other cohorts. This is likely due to the low co-incidence of these two types of mutations, such that in smaller cohorts either no cases (Hou *et al*, 2009) or a single case (Dufour *et al*, 2010) were identified within this subgroup, and confirmation of this issue may require a meta-analysis. Given that *CEBPA*-double cases with a *FLT3*/ITD are infrequent, constituting only 0.5% of the present cohort and 0.6% of a large cohort of 1182 NK patients (Taskesen *et al*, 2011), it could be argued that understanding the impact of *FLT3*/ITDs in this subgroup is relatively unimportant. However, if *FLT3*/ITDs are associated with a poor outcome regardless of *CEBPA* status, there is a rationale for limiting screening for *CEBPA* mutations at patient presentation to those lacking a *FLT3/*ITD.

For the identification of patients who would not benefit from an allo-SCT, a working threshold of 35% relapse rate at 4 years has been suggested, below which the transplant

related mortality outweighs any benefit due to a reduced tendency to relapse (Cornelissen *et al*, 2007). In the 52 *CEBPA*-double patients lacking a *FLT3*/ITD and identified as having a relatively favourable prognosis, the relapse rate was approximately 39% at 4 years, similar to the working cut-off of 35%. Across the patients stratified as *CEBPA*-WT, -single and –double, there were no differences in treatment allocation, nor did censoring patients at transplant affect the positive impact of *CEBPA*-double status on clinical outcome (chapter 4). However, further investigation is clearly required to determine the role of *CEBPA* status in determining therapy with regard to whether an alloSCT is beneficial in *CEBPA*-double cases without a *FLT3*/ITD in first remission, or should perhaps be delayed to second remission. In addition, the impact of age within this context is of interest given that the intensive conditioning given before transplantation is better tolerated in patients less than 40 years of age, and the benefit of *CEBPA*-double status in this study was predominantly observed in this age group. These types of studies remain challenging due to the relatively small numbers of patients in the *CEBPA*-double and *FLT3*/ITD-negative subgroup (52 of 1427, 4%), indeed even a large study of 509 patients with an NK lacked sufficient *CEBPA*-mutant cases to investigate the impact of related-donor transplantation in first remission (Schlenk *et al*, 2008), and a meta-analysis may be required.

Alternative ways in which *CEBPA* mutations may aid clinical decisions have been explored, including as a marker for MRD detection. However, whilst RQ-PCR assays have been developed for the sensitive detection of particular *CEBPA* mutations (Smith *et al*, 2006), the individual nature of the mutations so far identified means that it would be challenging to develop a specific assay for MRD analysis of every *CEBPA*-mutant patient.

The reduction of C/EBPα activity has been reported to occur by a variety of mechanisms (Koschmieder *et al*, 2009), including suppression of C/EBPα translation by the CBFB-MYH11 fusion in AML with inv(16) (Helbling *et al*, 2005), reduction of activity as a consequence of phosphorylation by FLT3/ITD (Radomska *et al*, 2006), and the silencing of *CEBPA* transcription by the RUNX1-RUNX1T1 in t(8;21) AML (Pabst *et al*, 2001a) or by promoter hypermethylation (Chim *et al*, 2002;Wouters *et al*, 2007a). Whilst *CEBPA* promoter hypermethylation is not common (Chim *et al*, 2002), a subset of patients with silenced *CEBPA*, associated with proximal promoter hypermethylation, activating *NOTCH1* mutations and a mixed myeloid/lymphoid phenotype were identified by GEP (Wouters *et al*, 2007a). In view of this, it would be interesting to explore whether silencing of the normal allele occurs in some patients with a single heterozygous *CEBPA* mutation, thereby creating a *de facto CEBPA*-double mutant.

Evidence from mouse models and from study of familial AML suggests that mutation of *CEBPA* is a founding event in some leukaemias, however there can be a long latency before the on-set of disease. In 16 cases with germline N-terminal mutations, the median age at AML presentation was 25.5 years, range 4 to 46 years (Smith *et al*, 2004;Sellick *et al*, 2005;Nanri *et al*, 2006;Pabst *et al*, 2008;Renneville *et al*, 2009b;Taskesen *et al*, 2011). The acquisition of further genetic "hits" is likely required for the leukaemic phenotype, although the number of these driver mutations may be small, for instance recent whole genome sequencing of 12 NK AMLs with FAB-type M1 found approximately three mutations per genome with translational consequences that were expressed in the patient and recurrent in other AML samples (Welch *et al*, 2012). The evidence presented in chapter 4 demonstrates that other characteristics may impact on outcome in *CEBPA*-double patients, such as *FLT3*/ITD, and whilst *CEBPA*-double status is associated with a relatively favourable outcome it is clear that there are a significant number of patients within this group that have a poor clinical prognosis, for instance, 10% did not achieve a CR and 46% were not alive 8 years after trial entry. Therefore, exploration of what the cooperating events are is important. Clarification of the role of more recently identified gene mutations in *CEBPA*-mutant AML is required, for instance *DNMT3A* mutations are both frequently identified in AML and associated with poor patient outcome (Table 1.4). Subsequent to the work presented in this thesis, the *DNMT3A* mutant status has been determined in 916 patients with intermediate-risk cytogenetics from AML 10 and AML 12, including the 856 cases in the cohort presented in chapter 4. *DNMT3A* mutations were less common in patients with *CEBPA* mutations, with 5 of 52 (10%) *CEBPA*-double, 5 of 34 (15%) *CEBPA*-single and 231 of 934 (25%) *CEBPA*-WT cases also positive for a *DNMT3A* mutation, although there were no significant differences in pair-wise comparisons (*CEBPA*-WT vs. *CEBPA*-single, $P = .5$; *CEBPA*-WT vs. *CEBPA*-double, $P = .07$; *CEBPA*-single vs. *CEBPA*-double, $P = .5$, Pearson's Chi-square), and the impact of *DNMT3A* mutations on patient outcome in this context has not yet been investigated (personal communication from Professor Rosemary Gale).

A recent study investigated novel markers in *CEBPA*-double/NK AML by exome sequencing of five cases and identified a strong association between biallelic *CEBPA* mutations and mutations in *GATA2*, a transcription factor important for megakaryocytic differentiation and the proliferation and maintenance of haemopoietic stem/progenitor cells (Greif *et al*, 2012). Mutations affecting the first zinc finger domain of *GATA2* were detected in 13 (39%) of 33 *CEBPA*-double cases, but not in 38 *CEBPA*-single or 89 *CEBPA*-WT patients. A subsequent report identified *GATA2* mutations in 18 (18%) of 98 *CEBPA*-double, but not in 22 *CEBPA*-single, cases (Fasan *et al*, 2012). In the context of *CEBPA*-double

mutated AML, no significant difference was found in clinical outcome between *GATA2*-mutant and *GATA2*-WT patients by Greif *et al*, whilst Fasan *et al* found a trend to improved OS for *GATA2*-mutant cases, but saw no significant difference for event-free survival. In light of this, 98 of the 107 *CEBPA*-mutant patients identified in chapter 3 were screened for *GATA2* mutations, including 55 *CEBPA*-double and 43 *CEBPA*-single cases. *GATA2* mutations were detected in 15 of 55 (27%) *CEBPA*-double and in 7 of 43 (16%) *CEBPA*-single cases, this difference not being significant ($P = .2$), and the *GATA2* mutations were mutually exclusive with *FLT3*/ITDs. Overall, there was no evidence that a *GATA2* mutation significantly affected the clinical outcome of either *CEBPA*-single or *CEBPA*-double cases, with the latter retaining an improved prognosis over the former. However, with small numbers in subgroups for this and other reported cohorts, a meta-analysis of the available data is desirable. In addition, there are several avenues of investigation regarding the functional implications of the association between *CEBPA* and *GATA2* mutations, such as the interaction between the two proteins, with GATA2 variously reported as a negative (Tong *et al*, 2005) and positive (Greif *et al*, 2012) regulator of C/EBPα activity on the basis of *in vitro* assays.

As described above, exome sequencing of five *CEBPA*-double cases with an NK enabled the discovery of the association between *CEBPA* and *GATA2* mutations, and a similar approach could be taken to determine other cooperating events, particularly those predicting *CEBPA*-double cases with a poor prognosis. One approach would be to obtain exome sequencing on a small cohort of *CEBPA*-double cases with poor response to therapy or early relapse and compare this to a group with long-term disease-free survival.

The functional consequence of missense mutations and in-frame insertions/deletions in the middle of the *CEBPA* gene is still unclear. Whilst they did not appear to impair the TA activity of C/EBPα, it is possible that they may prevent the interaction of C/EBPα with other molecules, such as cdk2, cdk4 and the SWI/SNF complex, or impact on other unknown functions. Further *in vitro* and *in vivo* experiments are required to determine whether these types of atypical mutations are indeed functional mutations or non-pathogenic bystander mutations. Initially, mutants could be exogenously expressed in cell lines and their impact on cell cycle, proliferative capacity and ability to differentiate analysed. In addition, knock-in mouse models of these mutants would be useful to further explore the implications of these mutations.

7.2 Conclusions

This thesis describes an investigation of *CEBPA* mutations in AML, clarifying their impact on patient prognosis and the factors that modify this effect. Novel findings were also made with regard to *CEBPA* mutant level in patient samples and the impact of non-classical mutations on protein function. This data has made a significant contribution to the current understanding of both the clinical and biological role of *CEBPA* mutations in AML, as well as provided further avenues for exploration.

PUBLICATION ARISING FROM THE WORK IN THIS THESIS

Green,C.L., Koo,K.K, Hills,R.K., Burnett,A.K., Linch,D.C., & Gale,R.E. (2010) Prognostic significance of CEBPA mutations in a large cohort of younger adult patients with acute myeloid leukemia: impact of double CEBPA mutations and the interaction with FLT3 and NPM1 mutations. *J.Clin.Oncol.*, **28**, 2739-2747.

REFERENCES

Abbas,S., Lugthart,S., Kavelaars,F.G., Schelen,A., Koenders,J.E., Zeilemaker,A., van Putten,W.J., Rijneveld,A.W., Lowenberg,B., & Valk,P.J. (2010) Acquired mutations in the genes encoding IDH1 and IDH2 both are recurrent aberrations in acute myeloid leukemia: prevalence and prognostic value. *Blood*, **116**, 2122-2126.

Abu-Duhier,F.M., Goodeve,A.C., Wilson,G.A., Care,R.S., Peake,I.R., & Reilly,J.T. (2001) Identification of novel FLT-3 Asp835 mutations in adult acute myeloid leukaemia. *Br.J.Haematol.*, **113**, 983-988.

Ahn,J.Y., Seo,K., Weinberg,O., Boyd,S.D., & Arber,D.A. (2009) A Comparison of Two Methods for Screening CEBPA Mutations in Patients with Acute Myeloid Leukemia. *J.Mol.Diagn.*

Allen,C.G., Hills,R.K., Evans,C.M., Lamb,K., Sellar,R., Moorman,A.V., Sale,M., Liu-Yin,J., Burnett,A.K., Linch,D.C., & Gale,R.E. (2011) Mutations in a Large Cohort of Young Adult Patients with Core Binding Factor Acute Myeloid Leukemia: Impact on Outcome and the Selection of Patients for Alternative Treatment Including Transplantation in First Complete Remission. *ASH Annual Meeting Abstracts*, **118**, 419.

Appelbaum,F.R., Gundacker,H., Head,D.R., Slovak,M.L., Willman,C.L., Godwin,J.E., Anderson,J.E., & Petersdorf,S.H. (2006) Age and acute myeloid leukemia. *Blood*, **107**, 3481-3485.

Ashworth,A., Lord,C.J., & Reis-Filho,J.S. (2011) Genetic interactions in cancer progression and treatment. *Cell*, **145**, 30-38.

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*, **111**, 2527-2537.

Bacher,U., Haferlach,T., Schoch,C., Kern,W., & Schnittger,S. (2006) Implications of NRAS mutations in AML: a study of 2502 patients. *Blood*, **107**, 3847-3853.

Barjesteh van Waalwijk van Doorn-Khosrovani, Erpelinck,C., Meijer,J., van,O.S., van Putten,W.L., Valk,P.J., Berna,B.H., Tenen,D.G., Lowenberg,B., & Delwel,R. (2003) Biallelic mutations in the CEBPA gene and low CEBPA expression levels as prognostic markers in intermediate-risk AML. *Hematol.J.*, **4**, 31-40.

Beer,P.A., Ortmann,C.A., Campbell,P.J., & Green,A.R. (2010) Independently acquired biallelic JAK2 mutations are present in a minority of patients with essential thrombocythemia. *Blood*, **116**, 1013-1014.

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.*, **33**, 451-458.

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

Benthaus,T., Schneider,F., Mellert,G., Zellmeier,E., Schneider,S., Kakadia,P.M., Hiddemann,W., Bohlander,S.K., Feuring-Buske,M., Braess,J., Spiekermann,K., & Dufour,A. (2008) Rapid and sensitive screening for CEBPA mutations in acute myeloid leukaemia. *Br.J.Haematol.*, **143**, 230-239.

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*, **16**, 390-400.

Bienz,M., Ludwig,M., Leibundgut,E.O., Mueller,B.U., Ratschiller,D., Solenthaler,M., Fey,M.F., & Pabst,T. (2005) Risk assessment in patients with acute myeloid leukemia and a normal karyotype. *Clin.Cancer Res.*, **11**, 1416-1424.

Biggio,V., Renneville,A., Nibourel,O., Philippe,N., Terriou,L., Roumier,C., Amouyel,P., Cottel,D., Castaigne,S., Dombret,H., Thomas,X., Fenaux,P., & Preudhomme,C. (2008) Recurrent in-frame insertion in C/EBPalpha TAD2 region is a polymorphism without prognostic value in AML. *Leukemia*, **22**, 655-657.

Boissel,N., Leroy,H., Brethon,B., Philippe,N., de,B.S., Auvrignon,A., Raffoux,E., Leblanc,T., Thomas,X., Hermine,O., Quesnel,B., Baruchel,A., Leverger,G., Dombret,H., & Preudhomme,C. (2006) Incidence and prognostic impact of c-Kit, FLT3, and Ras gene mutations in core binding factor acute myeloid leukemia (CBF-AML). *Leukemia*, **20**, 965-970.

Boissel,N., Nibourel,O., Renneville,A., Gardin,C., Reman,O., Contentin,N., Bordessoule,D., Pautas,C., de,R.T., Quesnel,B., Huchette,P., Philippe,N., Geffroy,S., Terre,C., Thomas,X., Castaigne,S., Dombret,H., & Preudhomme,C. (2010) Prognostic impact of isocitrate dehydrogenase enzyme isoforms 1 and 2 mutations in acute myeloid leukemia: a study by the Acute Leukemia French Association group. *J.Clin.Oncol.*, **28**, 3717-3723.

Bonnet,D. & Dick,J.E. (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat.Med.*, **3**, 730-737.

Bornhauser,M., Illmer,T., Schaich,M., Soucek,S., Ehninger,G., & Thiede,C. (2007) Improved outcome after stem-cell transplantation in FLT3/ITD-positive AML. *Blood*, **109**, 2264-2265.

Bowen,D., Groves,M.J., Burnett,A.K., Patel,Y., Allen,C., Green,C., Gale,R.E., Hills,R., & Linch,D.C. (2009) TP53 gene mutation is frequent in patients with acute myeloid leukemia and complex karyotype, and is associated with very poor prognosis. *Leukemia*, **23**, 203-206.

Bowen,D.T., Frew,M.E., Hills,R., Gale,R.E., Wheatley,K., Groves,M.J., Langabeer,S.E., Kottaridis,P.D., Moorman,A.V., Burnett,A.K., & Linch,D.C. (2005) RAS mutation in acute myeloid leukemia is associated with distinct cytogenetic subgroups but does not influence outcome in patients younger than 60 years. *Blood*, **106**, 2113-2119.

Breems,D.A. & Lowenberg,B. (2007) Acute myeloid leukemia and the position of autologous stem cell transplantation. *Semin.Hematol.*, **44**, 259-266.

Breems,D.A., van Putten,W.L., De Greef,G.E., Van Zelderen-Bhola,S.L., Gerssen-Schoorl,K.B., Mellink,C.H., Nieuwint,A., Jotterand,M., Hagemeijer,A., Beverloo,H.B., & Lowenberg,B. (2008) Monosomal karyotype in acute myeloid

leukemia: a better indicator of poor prognosis than a complex karyotype. *J.Clin.Oncol.*, **26**, 4791-4797.

Burnett,A., Wetzler,M., & Lowenberg,B. (2011) Therapeutic Advances in Acute Myeloid Leukemia. *J.Clin.Oncol.*

Burnett,A.K., Hills,R.K., Milligan,D.W., Goldstone,A.H., Prentice,A.G., McMullin,M.F., Duncombe,A., Gibson,B., & Wheatley,K. (2010) Attempts to optimize induction and consolidation treatment in acute myeloid leukemia: results of the MRC AML12 trial. *J.Clin.Oncol.*, **28**, 586-595.

Burnett,A.K., Hills,R.K., Wheatley,K., Goldstone,A.H., Prentice,A.G., & Milligan,D. (2006) A Sensitive Risk Score for Directing Treatment in Younger Patients with AML. *ASH Annual Meeting Abstracts*, **108**, 18.

Cairoli,R., Beghini,A., Grillo,G., Nadali,G., Elice,F., Ripamonti,C.B., Colapietro,P., Nichelatti,M., Pezzetti,L., Lunghi,M., Cuneo,A., Viola,A., Ferrara,F., Lazzarino,M., Rodeghiero,F., Pizzolo,G., Larizza,L., & Morra,E. (2006) Prognostic impact of c-KIT mutations in core binding factor leukemias: an Italian retrospective study. *Blood*, **107**, 3463-3468.

Calkhoven,C.F., Muller,C., & Leutz,A. (2000) Translational control of C/EBPalpha and C/EBPbeta isoform expression. *Genes Dev.*, **14**, 1920-1932.

Care,R.S., Valk,P.J., Goodeve,A.C., Abu-Duhier,F.M., Geertsma-Kleinekoort,W.M., Wilson,G.A., Gari,M.A., Peake,I.R., Lowenberg,B., & Reilly,J.T. (2003) Incidence and prognosis of c-KIT and FLT3 mutations in core binding factor (CBF) acute myeloid leukaemias. *Br.J.Haematol.*, **121**, 775-777.

Carnicer,M.J., Lasa,A., Buschbeck,M., Serrano,E., Carricondo,M., Brunet,S., Aventin,A., Sierra,J., Di,C.L., & Nomdedeu,J.F. (2008) K313dup is a recurrent CEBPA mutation in de novo acute myeloid leukemia (AML). *Ann.Hematol.*, **87**, 819-827.

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.*, **21**, 4642-4649.

Chim,C.S., Wong,A.S., & Kwong,Y.L. (2002) Infrequent hypermethylation of CEBPA promotor in acute myeloid leukaemia. *Br.J.Haematol.*, **119**, 988-990.

Chou,W.C., Chou,S.C., Liu,C.Y., Chen,C.Y., Hou,H.A., Kuo,Y.Y., Lee,M.C., Ko,B.S., Tang,J.L., Yao,M., Tsay,W., Wu,S.J., Huang,S.Y., Hsu,S.C., Chen,Y.C., Chang,Y.C., Kuo,Y.Y., Kuo,K.T., Lee,F.Y., Liu,M.C., Liu,C.W., Tseng,M.H., Huang,C.F., & Tien,H.F. (2011a) TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. *Blood*, **118**, 3803-3810.

Chou,W.C., Hou,H.A., Chen,C.Y., Tang,J.L., Yao,M., Tsay,W., Ko,B.S., Wu,S.J., Huang,S.Y., Hsu,S.C., Chen,Y.C., Huang,Y.N., Chang,Y.C., Lee,F.Y., Liu,M.C., Liu,C.W., Tseng,M.H., Huang,C.F., & Tien,H.F. (2010) Distinct clinical and biological characteristics in adult acute myeloid leukemia bearing isocitrate dehydrogenase 1 (IDH1) mutation. *Blood*.

Chou,W.C., Lei,W.C., Ko,B.S., Hou,H.A., Chen,C.Y., Tang,J.L., Yao,M., Tsay,W., Wu,S.J., Huang,S.Y., Hsu,S.C., Chen,Y.C., Chang,Y.C., Kuo,K.T., Lee,F.Y., Liu,M.C., Liu,C.W., Tseng,M.H., Huang,C.F., & Tien,H.F. (2011b) The prognostic impact and stability of Isocitrate dehydrogenase 2 mutation in adult patients with acute myeloid leukemia. *Leukemia*, **25**, 246-253.

Cleaves,R., Wang,Q.F., & Friedman,A.D. (2004) C/EBPalphap30, a myeloid leukemia oncoprotein, limits G-CSF receptor expression but not terminal granulopoiesis via site-selective inhibition of C/EBP DNA binding. *Oncogene*, **23**, 716-725.

Colombo,E., Alcalay,M., & Pelicci,P.G. (2011) Nucleophosmin and its complex network: a possible therapeutic target in hematological diseases. *Oncogene*, **30**, 2595-2609.

Corbacioglu,A., Frohling,S., Mendla,C., Eiwen,K., Habdank,M., Dohner,H., Schlenk,R.F., & Dohner,K. (2007) CEBPA Germline Mutation Screening in Cytogenetically Normal Acute Myeloid Leukemia with Somatically Acquired CEBPA Mutations. *ASH Annual Meeting Abstracts*, **110**, 363.

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 Lelie,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*, **109**, 3658-3666.

D'Alo',F., Johansen,L.M., Nelson,E.A., Radomska,H.S., Evans,E.K., Zhang,P., Nerlov,C., & Tenen,D.G. (2003) The amino terminal and E2F interaction domains are critical for C/EBP alpha-mediated induction of granulopoietic development of hematopoietic cells. *Blood*, **102**, 3163-3171.

Damm,F., Heuser,M., Morgan,M., Wagner,K., Gorlich,K., Grosshennig,A., Hamwi,I., Thol,F., Surdziel,E., Fiedler,W., Lubbert,M., Kanz,L., Reuter,C., Heil,G., Delwel,R., Lowenberg,B., Valk,P.J., Krauter,J., & Ganser,A. (2011) Integrative prognostic risk score in acute myeloid leukemia with normal karyotype. *Blood*, **117**, 4561-4568.

de The,H. & Chen,Z. (2010) Acute promyelocytic leukaemia: novel insights into the mechanisms of cure. *Nat.Rev.Cancer*, **10**, 775-783.

Dick,J.E. (2008) Stem cell concepts renew cancer research. *Blood*, **112**, 4793-4807.

Ding,L., Ley,T.J., Larson,D.E., Miller,C.A., Koboldt,D.C., Welch,J.S., Ritchey,J.K., Young,M.A., Lamprecht,T., McLellan,M.D., McMichael,J.F., Wallis,J.W., Lu,C., Shen,D., Harris,C.C., Dooling,D.J., Fulton,R.S., Fulton,L.L., Chen,K., Schmidt,H., Kalicki-Veizer,J., Magrini,V.J., Cook,L., McGrath,S.D., Vickery,T.L., Wendl,M.C., Heath,S., Watson,M.A., Link,D.C., Tomasson,M.H., Shannon,W.D., Payton,J.E., Kulkarni,S., Westervelt,P., Walter,M.J., Graubert,T.A., Mardis,E.R., Wilson,R.K., & DiPersio,J.F. (2012) Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*, **481**, 506-510.

Dohner,H., Estey,E.H., Amadori,S., Appelbaum,F.R., Buchner,T., Burnett,A.K., Dombret,H., Fenaux,P., Grimwade,D., Larson,R.A., Lo-Coco,F., Naoe,T., Niederwieser,D., Ossenkoppele,G.J., Sanz,M.A., Sierra,J., Tallman,M.S., Lowenberg,B., & Bloomfield,C.D. (2010) Diagnosis and management of acute

myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood*, **115**, 453-474.

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*, **106**, 3740-3746.

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. *J.Clin.Oncol.*, **20**, 3254-3261.

Dufour,A., Schneider,F., Metzeler,K.H., Hoster,E., Schneider,S., Zellmeier,E., Benthaus,T., Sauerland,M.C., Berdel,W.E., Buchner,T., Wormann,B., Braess,J., Hiddemann,W., Bohlander,S.K., & Spiekermann,K. (2010) Acute myeloid leukemia with biallelic CEBPA gene mutations and normal karyotype represents a distinct genetic entity associated with a favorable clinical outcome. *J.Clin.Oncol.*, **28**, 570-577.

Early Breast Cancer Trialists' Collaborative Group (1990) *Treatment of early breast cancer, volume 1: Worldwide Evidence 1985-1990.*, Oxford University Press.

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.*, **352**, 254-266.

Fasan,A., Eder,C., Haferlach,C., Grossmann,V., Kohlmann,A., Dicker,F., Kern,W., Haferlach,T., & Schnittger,S. (2012) GATA2 mutations are frequent in intermediate-risk karyotype AML with biallelic CEBPA mutations and are associated with favorable prognosis. *Leukemia*.

Fathi,A. & Levis,M. (2011) FLT3 inhibitors: a story of the old and the new. *Curr.Opin.Hematol.*, **18**, 71-76.

Figueroa,M.E., Lugthart,S., Li,Y., Erpelinck-Verschueren,C., Deng,X., Christos,P.J., Schifano,E., Booth,J., van,P.W., Skrabanek,L., Campagne,F., Mazumdar,M., Greally,J.M., Valk,P.J., Lowenberg,B., Delwel,R., & Melnick,A. (2010) DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell*, **17**, 13-27.

Fitzgibbon,J., Smith,L.L., Raghavan,M., Smith,M.L., Debernardi,S., Skoulakis,S., Lillington,D., Lister,T.A., & Young,B.D. (2005) Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. *Cancer Res.*, **65**, 9152-9154.

Friedman,A.D. (2007) Transcriptional control of granulocyte and monocyte development. *Oncogene*, **26**, 6816-6828.

Frohling,S., Corbacioglu,A., Schlenk,R.F., Dohner,H., & Dohner,K. (2007) In Reply to Genetic changes of CEBPA in cancer: mutations or polymorphisms? *J.Clin.Oncol.*

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*, **100**, 4372-4380.

Frohling,S., Schlenk,R.F., Krauter,J., Thiede,C., Ehninger,G., Haase,D., Harder,L., Kreitmeier,S., Scholl,C., Caligiuri,M.A., Bloomfield,C.D., Dohner,H., & Dohner,K. (2005) Acute myeloid leukemia with deletion 9q within a noncomplex karyotype is associated with CEBPA loss-of-function mutations. *Genes Chromosomes.Cancer*, **42**, 427-432.

Frohling,S., Schlenk,R.F., Stolze,I., Bihlmayr,J., Benner,A., Kreitmeier,S., Tobis,K., Dohner,H., & Dohner,K. (2004) CEBPA mutations in younger adults with acute myeloid leukemia and normal cytogenetics: prognostic relevance and analysis of cooperating mutations. *J.Clin.Oncol.*, **22**, 624-633.

Gaidzik,V.I., Bullinger,L., Schlenk,R.F., Zimmermann,A.S., Rock,J., Paschka,P., Corbacioglu,A., Krauter,J., Schlegelberger,B., Ganser,A., Spath,D., Kundgen,A., Schmidt-Wolf,I.G., Gotze,K., Nachbaur,D., Pfreundschuh,M., Horst,H.A., Dohner,H., & Dohner,K. (2011) RUNX1 mutations in acute myeloid leukemia:

results from a comprehensive genetic and clinical analysis from the AML study group. *J.Clin.Oncol.*, **29**, 1364-1372.

Gaidzik,V.I., Paschka,P., Spath,D., Habdank,M., Kohne,C.H., Germing,U., von Lilienfeld-Toal,M., Held,G., Horst,H.A., Haase,D., Bentz,M., Gotze,K., Dohner,H., Schlenk,R.F., Bullinger,L., & Dohner,K. (2012) TET2 mutations in acute myeloid leukemia (AML): results from a comprehensive genetic and clinical analysis of the AML study group. *J.Clin.Oncol.*, **30**, 1350-1357.

Gaidzik,V.I., Schlenk,R.F., Moschny,S., Becker,A., Bullinger,L., Corbacioglu,A., Krauter,J., Schlegelberger,B., Ganser,A., Dohner,H., & Dohner,K. (2009) Prognostic impact of WT1 mutations in cytogenetically normal acute myeloid leukemia: a study of the German-Austrian AML Study Group. *Blood*, **113**, 4505-4511.

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*, **111**, 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*, **106**, 3658-3665.

Geletu,M., Balkhi,M.Y., Peer Zada,A.A., Christopeit,M., Pulikkan,J.A., Trivedi,A.K., Tenen,D.G., & Behre,G. (2007) Target proteins of C/EBPalphap30 in AML: C/EBPalphap30 enhances sumoylation of C/EBPalphap42 via up-regulation of Ubc9. *Blood*, **110**, 3301-3309.

Gerlinger,M., Rowan,A.J., Horswell,S., Larkin,J., Endesfelder,D., Gronroos,E., Martinez,P., Matthews,N., Stewart,A., Tarpey,P., Varela,I., Phillimore,B., Begum,S., McDonald,N.Q., Butler,A., Jones,D., Raine,K., Latimer,C., Santos,C.R., Nohadani,M., Eklund,A.C., Spencer-Dene,B., Clark,G., Pickering,L., Stamp,G., Gore,M., Szallasi,Z., Downward,J., Futreal,P.A., & Swanton,C. (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N.Engl.J.Med.*, **366**, 883-892.

Goardon,N., Marchi,E., Atzberger,A., Quek,L., Schuh,A., Soneji,S., Woll,P., Mead,A., Alford,K.A., Rout,R., Chaudhury,S., Gilkes,A., Knapper,S., Beldjord,K., Begum,S., Rose,S., Geddes,N., Griffiths,M., Standen,G., Sternberg,A., Cavenagh,J., Hunter,H., Bowen,D., Killick,S., Robinson,L., Price,A., Macintyre,E., Virgo,P., Burnett,A., Craddock,C., Enver,T., Jacobsen,S.E., Porcher,C., & Vyas,P. (2011) Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell*, **19**, 138-152.

Gombart,A.F., Hofmann,W.K., Kawano,S., Takeuchi,S., Krug,U., Kwok,S.H., Larsen,R.J., Asou,H., Miller,C.W., Hoelzer,D., & Koeffler,H.P. (2002) Mutations in the gene encoding the transcription factor CCAAT/enhancer binding protein alpha in myelodysplastic syndromes and acute myeloid leukemias. *Blood*, **99**, 1332-1340.

Gratwohl,A., Stern,M., Brand,R., Apperley,J., Baldomero,H., de,W.T., Dini,G., Rocha,V., Passweg,J., Sureda,A., Tichelli,A., & Niederwieser,D. (2009) Risk score for outcome after allogeneic hematopoietic stem cell transplantation: a retrospective analysis. *Cancer*, **115**, 4715-4726.

Greaves,M. & Maley,C.C. (2012) Clonal evolution in cancer. *Nature*, **481**, 306-313.

Green,C., Linch,D.C., & Gale,R.E. (2008) Most acute myeloid leukaemia patients with intermediate mutant FLT3/ITD levels do not have detectable bi-allelic disease, indicating that heterozygous disease alone is associated with an adverse outcome. *Br.J.Haematol.*, **142**, 423-426.

Green,C.L., Evans,C.M., Hills,R.K., Burnett,A.K., Linch,D.C., & Gale,R.E. (2010) The prognostic significance of IDH1 mutations in younger adult patients with acute myeloid leukemia is dependent on FLT3/ITD status. *Blood*, **116**, 2779-2782.

Green,C.L., Evans,C.M., Zhao,L., Hills,R.K., Burnett,A.K., Linch,D.C., & Gale,R.E. (2011) The prognostic significance of IDH2 mutations in AML depends on the location of the mutation. *Blood*, **118**, 409-412.

Greif,P.A., Dufour,A., Konstandin,N.P., Ksienzyk,B., Zellmeier,E., Tizazu,B., Sturm,J., Benthaus,T., Herold,T., Yaghmaie,M., Dorge,P., Hopfner,K.P., Hauser,A., Graf,A., Krebs,S., Blum,H., Kakadia,P.M., Schneider,S., Hoster,E., Schneider,F., Stanulla,M., Braess,J., Sauerland,M.C., Berdel,W.E., Buchner,T., Woermann,B.J.,

Hiddemann,W., Spiekermann,K., & Bohlander,S.K. (2012) GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia. *Blood*.

Griffiths,M., Mason,J., Rindl,M., Akiki,S., McMullan,D., Stinton,V., Powell,H., Curtis,A., Bown,N., & Craddock,C. (2005) Acquired isodisomy for chromosome 13 is common in AML, and associated with FLT3-itd mutations. *Leukemia*, **19**, 2355-2358.

Grimwade,D. & Hills,R.K. (2009) Independent prognostic factors for AML outcome. *Hematology.Am.Soc.Hematol.Educ.Program.*, 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. (2010a) 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*, **116**, 354-365.

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.*, **27**, 3650-3658.

Grimwade,D., Vyas,P., & Freeman,S. (2010b) Assessment of minimal residual disease in acute myeloid leukemia. *Curr.Opin.Oncol.*, **22**, 656-663.

Grimwade,D., Walker,H., Harrison,G., Oliver,F., Chatters,S., Harrison,C.J., Wheatley,K., Burnett,A.K., & Goldstone,A.H. (2001) The predictive value of hierarchical cytogenetic classification in older adults with acute myeloid leukemia (AML): analysis of 1065 patients entered into the United Kingdom Medical Research Council AML11 trial. *Blood*, **98**, 1312-1320.

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*, **92**, 2322-2333.

Grisendi,S., Bernardi,R., Rossi,M., Cheng,K., Khandker,L., Manova,K., & Pandolfi,P.P. (2005) Role of nucleophosmin in embryonic development and tumorigenesis. *Nature*, **437**, 147-153.

Grossmann,V., Schnittger,S., Schindela,S., Klein,H.U., Eder,C., Dugas,M., Kern,W., Haferlach,T., Haferlach,C., & Kohlmann,A. (2011) Strategy for robust detection of insertions, deletions, and point mutations in CEBPA, a GC-rich content gene, using 454 next-generation deep-sequencing technology. *J.Mol.Diagn.*, **13**, 129-136.

Gupta,M., Raghavan,M., Gale,R.E., Chelala,C., Allen,C., Molloy,G., Chaplin,T., Linch,D.C., Cazier,J.B., & Young,B.D. (2008) Novel regions of acquired uniparental disomy discovered in acute myeloid leukemia. *Genes Chromosomes.Cancer*, **47**, 729-739.

Hanahan,D. & Weinberg,R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646-674.

Hann,I.M., Stevens,R.F., Goldstone,A.H., Rees,J.K., Wheatley,K., Gray,R.G., & Burnett,A.K. (1997) Randomized comparison of DAT versus ADE as induction chemotherapy in children and younger adults with acute myeloid leukemia. Results of the Medical Research Council's 10th AML trial (MRC AML10). Adult and Childhood Leukaemia Working Parties of the Medical Research Council. *Blood*, **89**, 2311-2318.

Harrison,C.J., Hills,R.K., Moorman,A.V., Grimwade,D.J., Hann,I., Webb,D.K., Wheatley,K., de Graaf,S.S., van den Berg,E., Burnett,A.K., & Gibson,B.E. (2010) Cytogenetics of childhood acute myeloid leukemia: United Kingdom Medical Research Council Treatment trials AML 10 and 12. *J.Clin.Oncol.*, **28**, 2674-2681.

Hegenbart,U., Niederwieser,D., Sandmaier,B.M., Maris,M.B., Shizuru,J.A., Greinix,H., Cordonnier,C., Rio,B., Gratwohl,A., Lange,T., Al-Ali,H., Storer,B., Maloney,D., McSweeney,P., Chauncey,T., Agura,E., Bruno,B., Maziarz,R.T., Petersen,F., & Storb,R. (2006) Treatment for acute myelogenous leukemia by low-dose, total-body,

irradiation-based conditioning and hematopoietic cell transplantation from related and unrelated donors. *J.Clin.Oncol.*, **24**, 444-453.

Helbling,D., Mueller,B.U., Timchenko,N.A., Schardt,J., Eyer,M., Betts,D.R., Jotterand,M., Meyer-Monard,S., Fey,M.F., & Pabst,T. (2005) CBFB-SMMHC is correlated with increased calreticulin expression and suppresses the granulocytic differentiation factor CEBPA in AML with inv(16). *Blood*, **106**, 1369-1375.

Herr,A.L., Labopin,M., Blaise,D., Milpied,N., Potter,M., Michallet,M., Heit,W., Ferrara,F., Esteve,J., Arcese,W., Ehninger,G., Rowe,J.M., Kobbe,G., Rosselet,A., Bunjes,D., Rio,B., Brune,M., Nagler,A., Gorin,N.C., Frassoni,F., & Rocha,V. (2007) HLA-identical sibling allogeneic peripheral blood stem cell transplantation with reduced intensity conditioning compared to autologous peripheral blood stem cell transplantation for elderly patients with de novo acute myeloid leukemia. *Leukemia*, **21**, 129-135.

Ho,P.A., Alonzo,T.A., Kopecky,K.J., Miller,K.L., Kuhn,J., Zeng,R., Gerbing,R.B., Raimondi,S.C., Hirsch,B.A., Oehler,V., Hurwitz,C.A., Franklin,J.L., Gamis,A.S., Petersdorf,S.H., Anderson,J.E., Reaman,G.H., Baker,L.H., Willman,C.L., Bernstein,I.D., Radich,J.P., Appelbaum,F.R., Stirewalt,D.L., & Meshinchi,S. (2010) Molecular alterations of the IDH1 gene in AML: a Children's Oncology Group and Southwest Oncology Group study. *Leukemia*, **24**, 909-913.

Hou,H.A., Lin,L.I., Chen,C.Y., & Tien,H.F. (2009) Reply to 'Heterogeneity within AML with CEBPA mutations; only CEBPA double mutations, but not single CEBPA mutations are associated with favorable prognosis'. *Br.J.Cancer*.

Hou,Y., Song,L., Zhu,P., Zhang,B., Tao,Y., Xu,X., Li,F., Wu,K., Liang,J., Shao,D., Wu,H., Ye,X., Ye,C., Wu,R., Jian,M., Chen,Y., Xie,W., Zhang,R., Chen,L., Liu,X., Yao,X., Zheng,H., Yu,C., Li,Q., Gong,Z., Mao,M., Yang,X., Yang,L., Li,J., Wang,W., Lu,Z., Gu,N., Laurie,G., Bolund,L., Kristiansen,K., Wang,J., Yang,H., Li,Y., Zhang,X., & Wang,J. (2012) Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell*, **148**, 873-885.

Jaffe,E.S., Harris,N.L., Stein,H., & Vardiman,J.W. (2001) *World Health Organization Classification of Tumors. Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues. Lyon, France*, IARC.

Juhl-Christensen,C., Bomberg,M., Melsvik,D., Hokland,P., & Nyvold,C.G. (2008) Capillary gel electrophoresis: a simple method for identification of mutations and polymorphisms in the CEBPA gene in patients with acute myeloid leukaemia. *Eur.J.Haematol.*, **81**, 273-280.

Juliusson,G., Antunovic,P., Derolf,A., Lehmann,S., Mollgard,L., Stockelberg,D., Tidefelt,U., Wahlin,A., & Hoglund,M. (2009) Age and acute myeloid leukemia: real world data on decision to treat and outcomes from the Swedish Acute Leukemia Registry. *Blood*, **113**, 4179-4187.

Kato,N., Kitaura,J., Doki,N., Komeno,Y., Watanabe-Okochi,N., Togami,K., Nakahara,F., Oki,T., Enomoto,Y., Fukuchi,Y., Nakajima,H., Harada,Y., Harada,H., & Kitamura,T. (2011) Two types of C/EBPalpha mutations play distinct but collaborative roles in leukemogenesis: lessons from clinical data and BMT models. *Blood*, **117**, 221-233.

Kelly,L.M. & Gilliland,D.G. (2002) Genetics of myeloid leukemias. *Annu.Rev.Genomics Hum.Genet.*, **3**, 179-198.

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*, **13**, 299-310.

Kondo,M., Wagers,A.J., Manz,M.G., Prohaska,S.S., Scherer,D.C., Beilhack,G.F., Shizuru,J.A., & Weissman,I.L. (2003) Biology of hematopoietic stem cells and progenitors: implications for clinical application. *Annu.Rev.Immunol.*, **21**, 759-806.

Koschmieder,S., Halmos,B., Levantini,E., & Tenen,D.G. (2009) Dysregulation of the C/EBPalpha differentiation pathway in human cancer. *J.Clin.Oncol.*, **27**, 619-628.

Krauter,J., Wagner,K., Schafer,I., Marschalek,R., Meyer,C., Heil,G., Schaich,M., Ehninger,G., Niederwieser,D., Krahl,R., Buchner,T., Sauerland,C., Schlegelberger,B., Dohner,K., Dohner,H., Schlenk,R.F., & Ganser,A. (2009) Prognostic factors in adult patients up to 60 years old with acute myeloid leukemia and translocations of chromosome band 11q23: individual patient data-based meta-

analysis of the German Acute Myeloid Leukemia Intergroup. *J.Clin.Oncol.*, **27**, 3000-3006.

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*, **442**, 818-822.

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*, **367**, 645-648.

Ley,T.J., Ding,L., Walter,M.J., McLellan,M.D., Lamprecht,T., Larson,D.E., Kandoth,C., Payton,J.E., Baty,J., Welch,J., Harris,C.C., Lichti,C.F., Townsend,R.R., Fulton,R.S., Dooling,D.J., Koboldt,D.C., Schmidt,H., Zhang,Q., Osborne,J.R., Lin,L., O'Laughlin,M., McMichael,J.F., Delehaunty,K.D., McGrath,S.D., Fulton,L.A., Magrini,V.J., Vickery,T.L., Hundal,J., Cook,L.L., Conyers,J.J., Swift,G.W., Reed,J.P., Alldredge,P.A., Wylie,T., Walker,J., Kalicki,J., Watson,M.A., Heath,S., Shannon,W.D., Varghese,N., Nagarajan,R., Westervelt,P., Tomasson,M.H., Link,D.C., Graubert,T.A., DiPersio,J.F., Mardis,E.R., & Wilson,R.K. (2010) DNMT3A mutations in acute myeloid leukemia. *N.Engl.J.Med.*, **363**, 2424-2433.

Ley,T.J., Mardis,E.R., Ding,L., Fulton,B., McLellan,M.D., Chen,K., Dooling,D., Dunford-Shore,B.H., McGrath,S., Hickenbotham,M., Cook,L., Abbott,R., Larson,D.E., Koboldt,D.C., Pohl,C., Smith,S., Hawkins,A., Abbott,S., Locke,D., Hillier,L.W., Miner,T., Fulton,L., Magrini,V., Wylie,T., Glasscock,J., Conyers,J., Sander,N., Shi,X., Osborne,J.R., Minx,P., Gordon,D., Chinwalla,A., Zhao,Y., Ries,R.E., Payton,J.E., Westervelt,P., Tomasson,M.H., Watson,M., Baty,J., Ivanovich,J., Heath,S., Shannon,W.D., Nagarajan,R., Walter,M.J., Link,D.C., Graubert,T.A., DiPersio,J.F., & Wilson,R.K. (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*, **456**, 66-72.

Lin,F.T., MacDougald,O.A., Diehl,A.M., & Lane,M.D. (1993) A 30-kDa alternative translation product of the CCAAT/enhancer binding protein alpha message: transcriptional activator lacking antimitotic activity. *Proc.Natl.Acad.Sci.U.S.A*, **90**, 9606-9610.

Lin,L.I., Chen,C.Y., Lin,D.T., Tsay,W., Tang,J.L., Yeh,Y.C., Shen,H.L., Su,F.H., Yao,M., Huang,S.Y., & Tien,H.F. (2005) Characterization of CEBPA mutations in acute myeloid leukemia: most patients with CEBPA mutations have biallelic mutations and show a distinct immunophenotype of the leukemic cells. *Clin.Cancer Res.*, **11**, 1372-1379.

Lin,L.I., Lin,T.C., Chou,W.C., Tang,J.L., Lin,D.T., & Tien,H.F. (2006) A novel fluorescence-based multiplex PCR assay for rapid simultaneous detection of CEBPA mutations and NPM mutations in patients with acute myeloid leukemias. *Leukemia*, **20**, 1899-1903.

Marcucci,G., Haferlach,T., & Dohner,H. (2011) Molecular Genetics of Adult Acute Myeloid Leukemia: Prognostic and Therapeutic Implications. *J.Clin.Oncol.*

Marcucci,G., Maharry,K., Radmacher,M.D., Mrozek,K., Vukosavljevic,T., Paschka,P., Whitman,S.P., Langer,C., Baldus,C.D., Liu,C.G., Ruppert,A.S., Powell,B.L., Carroll,A.J., Caligiuri,M.A., Kolitz,J.E., Larson,R.A., & Bloomfield,C.D. (2008) Prognostic significance of, and gene and microRNA expression signatures associated with, CEBPA mutations in cytogenetically normal acute myeloid leukemia with high-risk molecular features: a Cancer and Leukemia Group B Study. *J.Clin.Oncol.*, **26**, 5078-5087.

Marcucci,G., Maharry,K., Wu,Y.Z., Radmacher,M.D., Mrozek,K., Margeson,D., Holland,K.B., Whitman,S.P., Becker,H., Schwind,S., Metzeler,K.H., Powell,B.L., Carter,T.H., Kolitz,J.E., Wetzler,M., Carroll,A.J., Baer,M.R., Caligiuri,M.A., Larson,R.A., & Bloomfield,C.D. (2010) IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J.Clin.Oncol.*, **28**, 2348-2355.

Marcucci,G., Metzeler,K.H., Schwind,S., Becker,H., Maharry,K., Mrozek,K., Radmacher,M.D., Kohlschmidt,J., Nicolet,D., Whitman,S.P., Wu,Y.Z., Powell,B.L., Carter,T.H., Kolitz,J.E., Wetzler,M., Carroll,A.J., Baer,M.R., Moore,J.O., Caligiuri,M.A., Larson,R.A., & Bloomfield,C.D. (2012) Age-related prognostic impact of different types of DNMT3A mutations in adults with primary cytogenetically normal acute myeloid leukemia. *J.Clin.Oncol.*, **30**, 742-750.

Mardis,E.R., Ding,L., Dooling,D.J., Larson,D.E., McLellan,M.D., Chen,K., Koboldt,D.C., Fulton,R.S., Delehaunty,K.D., McGrath,S.D., Fulton,L.A., Locke,D.P., Magrini,V.J., Abbott,R.M., Vickery,T.L., Reed,J.S., Robinson,J.S., Wylie,T., Smith,S.M., Carmichael,L., Eldred,J.M., Harris,C.C., Walker,J., Peck,J.B., Du,F., Dukes,A.F., Sanderson,G.E., Brummett,A.M., Clark,E., McMichael,J.F., Meyer,R.J., Schindler,J.K., Pohl,C.S., Wallis,J.W., Shi,X., Lin,L., Schmidt,H., Tang,Y., Haipek,C., Wiechert,M.E., Ivy,J.V., Kalicki,J., Elliott,G., Ries,R.E., Payton,J.E., Westervelt,P., Tomasson,M.H., Watson,M.A., Baty,J., Heath,S., Shannon,W.D., Nagarajan,R., Link,D.C., Walter,M.J., Graubert,T.A., DiPersio,J.F., Wilson,R.K., & Ley,T.J. (2009) Recurring mutations found by sequencing an acute myeloid leukemia genome. *N.Engl.J.Med.*, **361**, 1058-1066.

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*, **110**, 1262-1270.

Meshinchi,S., Arceci,R.J., Sanders,J.E., Smith,F.O., Woods,W.B., Radich,J.P., & Alonzo,T.A. (2006) Role of allogeneic stem cell transplantation in FLT3/ITD-positive AML. *Blood*, **108**, 400-401.

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*, **112**, 4193-4201.

Metzeler,K.H., Maharry,K., Radmacher,M.D., Mrozek,K., Margeson,D., Becker,H., Curfman,J., Holland,K.B., Schwind,S., Whitman,S.P., Wu,Y.Z., Blum,W., Powell,B.L., Carter,T.H., Wetzler,M., Moore,J.O., Kolitz,J.E., Baer,M.R., Carroll,A.J., Larson,R.A., Caligiuri,M.A., Marcucci,G., & Bloomfield,C.D. (2011) TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. *J.Clin.Oncol.*, **29**, 1373-1381.

Miller,M., Shuman,J.D., Sebastian,T., Dauter,Z., & Johnson,P.F. (2003) Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. *J.Biol.Chem.*, **278**, 15178-15184.

Mrozek,K., Radmacher,M.D., Bloomfield,C.D., & Marcucci,G. (2009) Molecular signatures in acute myeloid leukemia. *Curr.Opin.Hematol.*, **16**, 64-69.

Mueller,B.U. & Pabst,T. (2006) C/EBPalpha and the pathophysiology of acute myeloid leukemia. *Curr.Opin.Hematol.*, **13**, 7-14.

Muller,C., Calkhoven,C.F., Sha,X., & Leutz,A. (2004) The CCAAT enhancer-binding protein alpha (C/EBPalpha) requires a SWI/SNF complex for proliferation arrest. *J.Biol.Chem.*, **279**, 7353-7358.

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*, **10**, 1911-1918.

Nanri,T., Uike,N., Kawakita,T., Iwanaga,E., Hoshino,K., Mitsuya,H., & Asou,N. (2006) A Pedigree Harboring a Germ-Line N-Terminal C/EBP{alpha} Mutation and Development of Acute Myeloblastic Leukemia with a Somatic C-Terminal C/EBP{alpha} Mutation. *ASH Annual Meeting Abstracts*, **108**, 1916.

Nerlov,C. (2004) C/EBPalpha mutations in acute myeloid leukaemias. *Nat.Rev.Cancer*, **4**, 394-400.

Orkin,S.H. & Zon,L.I. (2008) Hematopoiesis: an evolving paradigm for stem cell biology. *Cell*, **132**, 631-644.

Pabst,T., Eyholzer,M., Fos,J., & Mueller,B.U. (2009) Heterogeneity within AML with CEBPA mutations; only CEBPA double mutations, but not single CEBPA mutations are associated with favourable prognosis. *Br.J.Cancer*, **100**, 1343-1346.

Pabst,T., Eyholzer,M., Haefliger,S., Schardt,J., & Mueller,B.U. (2008) Somatic CEBPA mutations are a frequent second event in families with germline CEBPA mutations and familial acute myeloid leukemia. *J.Clin.Oncol.*, **26**, 5088-5093.

Pabst,T., Mueller,B.U., Harakawa,N., Schoch,C., Haferlach,T., Behre,G., Hiddemann,W., Zhang,D.E., & Tenen,D.G. (2001a) AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. *Nat.Med.*, **7**, 444-451.

Pabst,T., Mueller,B.U., Zhang,P., Radomska,H.S., Narravula,S., Schnittger,S., Behre,G., Hiddemann,W., & Tenen,D.G. (2001b) Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. *Nat.Genet.*, **27**, 263-270.

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.*, **24**, 3904-3911.

Paschka,P., Marcucci,G., Ruppert,A.S., Whitman,S.P., Mrozek,K., Maharry,K., Langer,C., Baldus,C.D., Zhao,W., Powell,B.L., Baer,M.R., Carroll,A.J., Caligiuri,M.A., Kolitz,J.E., Larson,R.A., & Bloomfield,C.D. (2008) Wilms' tumor 1 gene mutations independently predict poor outcome in adults with cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. *J.Clin.Oncol.*, **26**, 4595-4602.

Paschka,P., Schlenk,R.F., Gaidzik,V.I., Habdank,M., Kronke,J., Bullinger,L., Spath,D., Kayser,S., Zucknick,M., Gotze,K., Horst,H.A., Germing,U., Dohner,H., & Dohner,K. (2010) IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplication. *J.Clin.Oncol.*, **28**, 3636-3643.

Porse,B.T., Pedersen,T.A., Hasemann,M.S., Schuster,M.B., Kirstetter,P., Luedde,T., Damgaard,I., Kurz,E., Schjerling,C.K., & Nerlov,C. (2006) The proline-histidine-rich CDK2/CDK4 interaction region of C/EBPalpha is dispensable for C/EBPalpha-mediated growth regulation in vivo. *Mol.Cell Biol.*, **26**, 1028-1037.

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*, **107**, 247-258.

Preudhomme,C., Sagot,C., Boissel,N., Cayuela,J.M., Tigaud,I., de,B.S., Thomas,X., Raffoux,E., Lamandin,C., Castaigne,S., Fenaux,P., & Dombret,H. (2002) Favorable prognostic significance of CEBPA mutations in patients with de novo acute myeloid leukemia: a study from the Acute Leukemia French Association (ALFA). *Blood*, **100**, 2717-2723.

Quintana-Bustamante,O., Smith,S.L., Griessinger,E., Reyal,Y., Vargaftig,J., Lister,T.A., Fitzgibbon,J., & Bonnet,D. (2012) Overexpression of wild-type or mutants forms of CEBPA alter normal human hematopoiesis. *Leukemia*, **26**, 1537-1546.

Radomska,H.S., Basseres,D.S., Zheng,R., Zhang,P., Dayaram,T., Yamamoto,Y., Sternberg,D.W., Lokker,N., Giese,N.A., Bohlander,S.K., Schnittger,S., Delmotte,M.H., Davis,R.J., Small,D., Hiddemann,W., Gilliland,D.G., & Tenen,D.G. (2006) Block of C/EBP alpha function by phosphorylation in acute myeloid leukemia with FLT3 activating mutations. *J.Exp.Med.*, **203**, 371-381.

Raghavan,M., Lillington,D.M., Skoulakis,S., Debernardi,S., Chaplin,T., Foot,N.J., Lister,T.A., & Young,B.D. (2005) Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. *Cancer Res.*, **65**, 375-378.

Ramji,D.P. & Foka,P. (2002) CCAAT/enhancer-binding proteins: structure, function and regulation. *Biochem.J.*, **365**, 561-575.

Rangarajan,A., Hong,S.J., Gifford,A., & Weinberg,R.A. (2004) Species- and cell type-specific requirements for cellular transformation. *Cancer Cell*, **6**, 171-183.

Reckzeh,K., Bereshchenko,O., Mead,A., Rehn,M., Kharazi,S., Jacobsen,S.E., Nerlov,C., & Cammenga,J. (2012) Molecular and cellular effects of oncogene cooperation in a genetically accurate AML mouse model. *Leukemia*.

Renneville,A., Boissel,N., Gachard,N., Naguib,D., Bastard,C., de,B.S., Nibourel,O., Pautas,C., Reman,O., Thomas,X., Gardin,C., Terre,C., Castaigne,S.,

Preudhomme,C., & Dombret,H. (2009a) The favorable impact of CEBPA mutations in patients with acute myeloid leukemia is only observed in the absence of associated cytogenetic abnormalities and FLT3 internal duplication. *Blood*, **113**, 5090-5093.

Renneville,A., Boissel,N., Nibourel,O., Berthon,C., Helevaut,N., Gardin,C., Cayuela,J.M., Hayette,S., Reman,O., Contentin,N., Bordessoule,D., Pautas,C., Botton,S., Revel,T., Terre,C., Fenaux,P., Thomas,X., Castaigne,S., Dombret,H., & Preudhomme,C. (2012) Prognostic significance of DNA methyltransferase 3A mutations in cytogenetically normal acute myeloid leukemia: a study by the Acute Leukemia French Association. *Leukemia*, **26**, 1247-1254.

Renneville,A., Mialou,V., Philippe,N., Kagialis-Girard,S., Biggio,V., Zabot,M.T., Thomas,X., Bertrand,Y., & Preudhomme,C. (2009b) Another pedigree with familial acute myeloid leukemia and germline CEBPA mutation. *Leukemia*, **23**, 804-806.

Ribeiro,A.F., Pratcorona,M., Erpelinck-Verschueren,C., Rockova,V., Sanders,M., Abbas,S., Figueroa,M.E., Zeilemaker,A., Melnick,A., Lowenberg,B., Valk,P.J., & Delwel,R. (2012) Mutant DNMT3A: a marker of poor prognosis in acute myeloid leukemia. *Blood*, **119**, 5824-5831.

Rollig,C., Bornhauser,M., Thiede,C., Taube,F., Kramer,M., Mohr,B., Aulitzky,W., Bodenstein,H., Tischler,H.J., Stuhlmann,R., Schuler,U., Stolzel,F., von,B.M., Wandt,H., Schafer-Eckart,K., Schaich,M., & Ehninger,G. (2011) Long-term prognosis of acute myeloid leukemia according to the new genetic risk classification of the European LeukemiaNet recommendations: evaluation of the proposed reporting system. *J.Clin.Oncol.*, **29**, 2758-2765.

Rucker,F.G., Schlenk,R.F., Bullinger,L., Kayser,S., Teleanu,V., Kett,H., Habdank,M., Kugler,C.M., Holzmann,K., Gaidzik,V.I., Paschka,P., Held,G., von Lilienfeld-Toal,M., Lubbert,M., Frohling,S., Zenz,T., Krauter,J., Schlegelberger,B., Ganser,A., Lichter,P., Dohner,K., & Dohner,H. (2012) TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype, and dismal outcome. *Blood*, **119**, 2114-2121.

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.*, **358**, 1909-1918.

Schnittger,S., Bacher,U., Eder,C., Lohse,P., Haferlach,C., Kern,W., & Haferlach,T. (2011a) A copy number repeat polymorphism in the transactivation domain of the CEPBA gene is possibly associated with a protective effect against acquired CEBPA mutations: an analysis in 1135 patients with AML and 187 healthy controls. *Exp.Hematol.*, **39**, 87-94.

Schnittger,S., Dicker,F., Kern,W., Wendland,N., Sundermann,J., Alpermann,T., Haferlach,C., & Haferlach,T. (2011b) RUNX1 mutations are frequent in de novo AML with noncomplex karyotype and confer an unfavorable prognosis. *Blood*, **117**, 2348-2357.

Schnittger,S., Kern,W., Tschulik,C., Weiss,T., Dicker,F., Falini,B., Haferlach,C., & Haferlach,T. (2009) Minimal residual disease levels assessed by NPM1 mutation-specific RQ-PCR provide important prognostic information in AML. *Blood*, **114**, 2220-2231.

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*, **14**, 796-804.

Schnittger,S., Schoch,C., Kern,W., Mecucci,C., Tschulik,C., Martelli,M.F., Haferlach,T., Hiddemann,W., & Falini,B. (2005) Nucleophosmin gene mutations are predictors of favorable prognosis in acute myelogenous leukemia with a normal karyotype. *Blood*, **106**, 3733-3739.

Schwieger,M., Lohler,J., Fischer,M., Herwig,U., Tenen,D.G., & Stocking,C. (2004) A dominant-negative mutant of C/EBPalpha, associated with acute myeloid leukemias, inhibits differentiation of myeloid and erythroid progenitors of man but not mouse. *Blood*, **103**, 2744-2752.

Sellick,G.S., Spendlove,H.E., Catovsky,D., Pritchard-Jones,K., & Houlston,R.S. (2005) Further evidence that germline CEBPA mutations cause dominant inheritance of acute myeloid leukaemia. *Leukemia*, **19**, 1276-1278.

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. (2006) 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*, **20**, 604-609.

Slomiany,B.A., D'Arigo,K.L., Kelly,M.M., & Kurtz,D.T. (2000) C/EBPalpha inhibits cell growth via direct repression of E2F-DP-mediated transcription. *Mol.Cell Biol.*, **20**, 5986-5997.

Small,D. (2006) FLT3 mutations: biology and treatment. *Hematology.Am.Soc.Hematol.Educ.Program.*, 178-184.

Smith,L.L., Pearce,D., Smith,M.L., Jenner,M., Lister,T.A., Bonnet,D., Goff,L., & Fitzgibbon,J. (2006) Development of a quantitative real-time polymerase chain reaction method for monitoring CEBPA mutations in normal karyotype acute myeloid leukaemia. *Br.J.Haematol.*, **133**, 103-105.

Smith,M.L., Cavenagh,J.D., Lister,T.A., & Fitzgibbon,J. (2004) Mutation of CEBPA in familial acute myeloid leukemia. *N.Engl.J.Med.*, **351**, 2403-2407.

Smith,M.L., Hills,R.K., & Grimwade,D. (2011) Independent prognostic variables in acute myeloid leukaemia. *Blood Rev.*, **25**, 39-51.

Snaddon,J., Smith,M.L., Neat,M., Cambal-Parrales,M., Dixon-McIver,A., Arch,R., Amess,J.A., Rohatiner,A.Z., Lister,T.A., & Fitzgibbon,J. (2003) Mutations of CEBPA in acute myeloid leukemia FAB types M1 and M2. *Genes Chromosomes.Cancer*, **37**, 72-78.

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*, **37**, 237-251.

Stratton,M.R., Campbell,P.J., & Futreal,P.A. (2009) The cancer genome. *Nature*, **458**, 719-724.

Swerdlow,S.H., Campo,E., Harris,N.L., Jaffe,E.S., Pileri,S.A., Stein,H., Thiele,J., & Vardiman,J.W. (2008) *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, Fourth Edition,* IARC Press.

Tang,J.L., Hou,H.A., Chen,C.Y., Liu,C.Y., Chou,W.C., Tseng,M.H., Huang,C.F., Lee,F.Y., Liu,M.C., Yao,M., Huang,S.Y., Ko,B.S., Hsu,S.C., Wu,S.J., Tsay,W., Chen,Y.C., Lin,L.I., & Tien,H.F. (2009) AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations. *Blood*, **114**, 5352-5361.

Taskesen,E., Bullinger,L., Corbacioglu,A., Sanders,M.A., Erpelinck,C.A., Wouters,B.J., van der Poel-van de Luytgaarde SC, Damm,F., Krauter,J., Ganser,A., Schlenk,R.F., Lowenberg,B., Delwel,R., Dohner,H., Valk,P.J., & Dohner,K. (2011) Prognostic impact, concurrent genetic mutations, and gene expression features of AML with CEBPA mutations in a cohort of 1182 cytogenetically normal AML patients: further evidence for CEBPA double mutant AML as a distinctive disease entity. *Blood*, **117**, 2469-2475.

Taube,F., Illmer,T., Schaich,M., Kroschinsky,F., Oelschlagel,U., Mohr,B., Platzbecker,U., Ehninger,G., & Thiede,C. (2009) CEBPA gene mutations affecting the TAD and BZIP region show differences in their association with clinical parameters and prognosis: an analysis in 1779 adult patients with AML. (Abstract).*Haematologica*, **94 (suppl. 2)**, 218.

Taussig,D.C., Vargaftig,J., Miraki-Moud,F., Griessinger,E., Sharrock,K., Luke,T., Lillington,D., Oakervee,H., Cavenagh,J., Agrawal,S.G., Lister,T.A., Gribben,J.G., & Bonnet,D. (2010) Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction. *Blood*, **115**, 1976-1984.

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*, **107**, 4011-4020.

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*, **99**, 4326-4335.

Thol,F., Damm,F., Ludeking,A., Winschel,C., Wagner,K., Morgan,M., Yun,H., Gohring,G., Schlegelberger,B., Hoelzer,D., Lubbert,M., Kanz,L., Fiedler,W., Kirchner,H., Heil,G., Krauter,J., Ganser,A., & Heuser,M. (2011) Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia. *J.Clin.Oncol.*, **29**, 2889-2896.

Thol,F., Damm,F., Wagner,K., Gohring,G., Schlegelberger,B., Hoelzer,D., Lubbert,M., Heit,W., Kanz,L., Schlimok,G., Raghavachar,A., Fiedler,W., Kirchner,H., Heil,G., Heuser,M., Krauter,J., & Ganser,A. (2010) Prognostic impact of IDH2 mutations in cytogenetically normal acute myeloid leukemia. *Blood*, **116**, 614-616.

Tiesmeier,J., Czwalinna,A., Muller-Tidow,C., Krauter,J., Serve,H., Heil,G., Ganser,A., & Verbeek,W. (2003) Evidence for allelic evolution of C/EBPalpha mutations in acute myeloid leukaemia. *Br.J.Haematol.*, **123**, 413-419.

Timchenko,N.A., Wilde,M., Nakanishi,M., Smith,J.R., & Darlington,G.J. (1996) CCAAT/enhancer-binding protein alpha (C/EBP alpha) inhibits cell proliferation through the p21 (WAF-1/CIP-1/SDI-1) protein. *Genes Dev.*, **10**, 804-815.

Tong,Q., Tsai,J., Tan,G., Dalgin,G., & Hotamisligil,G.S. (2005) Interaction between GATA and the C/EBP family of transcription factors is critical in GATA-mediated suppression of adipocyte differentiation. *Mol.Cell Biol.*, **25**, 706-715.

Umek,R.M., Friedman,A.D., & McKnight,S.L. (1991) CCAAT-enhancer binding protein: a component of a differentiation switch. *Science*, **251**, 288-292.

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.*, **350**, 1617-1628.

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*, **114**, 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. (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*, **106**, 3747-3754.

Virappane,P., Gale,R., Hills,R., Kakkas,I., Summers,K., Stevens,J., Allen,C., Green,C., Quentmeier,H., Drexler,H., Burnett,A., Linch,D., Bonnet,D., Lister,T.A., & Fitzgibbon,J. (2008) Mutation of the Wilms' tumor 1 gene is a poor prognostic factor associated with chemotherapy resistance in normal karyotype acute myeloid leukemia: the United Kingdom Medical Research Council Adult Leukaemia Working Party. *J.Clin.Oncol.*, **26**, 5429-5435.

Wagner,K., Damm,F., Gohring,G., Gorlich,K., Heuser,M., Schafer,I., Ottmann,O., Lubbert,M., Heit,W., Kanz,L., Schlimok,G., Raghavachar,A.A., Fiedler,W., Kirchner,H.H., Brugger,W., Zucknick,M., Schlegelberger,B., Heil,G., Ganser,A., & Krauter,J. (2010) Impact of IDH1 R132 mutations and an IDH1 single nucleotide polymorphism in cytogenetically normal acute myeloid leukemia: SNP rs11554137 is an adverse prognostic factor. *J.Clin.Oncol.*, **28**, 2356-2364.

Wang,C., Chen,X., Wang,Y., Gong,J., & Hu,G. (2007) C/EBPalphap30 plays transcriptional regulatory roles distinct from C/EBPalphap42. *Cell Res.*, **17**, 374-383.

Wang,H., Iakova,P., Wilde,M., Welm,A., Goode,T., Roesler,W.J., & Timchenko,N.A. (2001) C/EBPalpha arrests cell proliferation through direct inhibition of Cdk2 and Cdk4. *Mol.Cell*, **8**, 817-828.

Wang,N.D., Finegold,M.J., Bradley,A., Ou,C.N., Abdelsayed,S.V., Wilde,M.D., Taylor,L.R., Wilson,D.R., & Darlington,G.J. (1995) Impaired energy homeostasis in C/EBP alpha knockout mice. *Science*, **269**, 1108-1112.

Ward,P.S., Patel,J., Wise,D.R., Abdel-Wahab,O., Bennett,B.D., Coller,H.A., Cross,J.R., Fantin,V.R., Hedvat,C.V., Perl,A.E., Rabinowitz,J.D., Carroll,M., Su,S.M., Sharp,K.A., Levine,R.L., & Thompson,C.B. (2010) The common feature of

leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell*, **17**, 225-234.

Welch,J.S., Ley,T.J., Link,D.C., Miller,C.A., Larson,D.E., Koboldt,D.C., Wartman,L.D., Lamprecht,T.L., Liu,F., Xia,J., Kandoth,C., Fulton,R.S., McLellan,M.D., Dooling,D.J., Wallis,J.W., Chen,K., Harris,C.C., Schmidt,H.K., Kalicki-Veizer,J.M., Lu,C., Zhang,Q., Lin,L., O'Laughlin,M.D., McMichael,J.F., Delehaunty,K.D., Fulton,L.A., Magrini,V.J., McGrath,S.D., Demeter,R.T., Vickery,T.L., Hundal,J., Cook,L.L., Swift,G.W., Reed,J.P., Alldredge,P.A., Wylie,T.N., Walker,J.R., Watson,M.A., Heath,S.E., Shannon,W.D., Varghese,N., Nagarajan,R., Payton,J.E., Baty,J.D., Kulkarni,S., Klco,J.M., Tomasson,M.H., Westervelt,P., Walter,M.J., Graubert,T.A., DiPersio,J.F., Ding,L., Mardis,E.R., & Wilson,R.K. (2012) The origin and evolution of mutations in acute myeloid leukemia. *Cell*, **150**, 264-278.

Wheatley,K., Burnett,A.K., Goldstone,A.H., Gray,R.G., Hann,I.M., Harrison,C.J., Rees,J.K., Stevens,R.F., & Walker,H. (1999) A simple, robust, validated and highly predictive index for the determination of risk-directed therapy in acute myeloid leukaemia derived from the MRC AML 10 trial. United Kingdom Medical Research Council's Adult and Childhood Leukaemia Working Parties. *Br.J.Haematol.*, **107**, 69-79.

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.*, **61**, 7233-7239.

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. (2007a) Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood*, **110**, 3706-3714.

Wouters,B.J., Louwers,I., Valk,P.J., Lowenberg,B., & Delwel,R. (2007b) A recurrent in-frame insertion in a CEBPA transactivation domain is a polymorphism rather than a

mutation that does not affect gene expression profiling-based clustering of AML. *Blood*, **109**, 389-390.

Wouters,B.J., Lowenberg,B., Erpelinck-Verschueren,C.A., van Putten,W.L., Valk,P.J., & Delwel,R. (2009) Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood*, **113**, 3088-3091.

Wouters,B.J., Sanders,M.A., Lugthart,S., Geertsma-Kleinekoort,W.M., van,D.E., Beverloo,H.B., Lowenberg,B., Valk,P.J., & Delwel,R. (2007c) Segmental uniparental disomy as a recurrent mechanism for homozygous CEBPA mutations in acute myeloid leukemia. *Leukemia*, **21**, 2382-2384.

Yamamoto,Y., Kiyoi,H., Nakano,Y., Suzuki,R., Kodera,Y., Miyawaki,S., Asou,N., Kuriyama,K., Yagasaki,F., Shimazaki,C., Akiyama,H., Saito,K., Nishimura,M., Motoji,T., Shinagawa,K., Takeshita,A., Saito,H., Ueda,R., Ohno,R., & Naoe,T. (2001) Activating mutation of D835 within the activation loop of FLT3 in human hematologic malignancies. *Blood*, **97**, 2434-2439.

Yan,X.J., Xu,J., Gu,Z.H., Pan,C.M., Lu,G., Shen,Y., Shi,J.Y., Zhu,Y.M., Tang,L., Zhang,X.W., Liang,W.X., Mi,J.Q., Song,H.D., Li,K.Q., Chen,Z., & Chen,S.J. (2011) Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat.Genet.*, **43**, 309-315.

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*, **94**, 569-574.

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.*, **21**, 853-863.

Appendix Table 1 Primer sequences

Primer name	Primer sequence (5' to 3')	Location in gene[1]
CEBPA/1F[2]	TCGCCATGCCGGGAGAACTCTAAC	-30 to -7
CEBPA/1R	TAGCCGGCGGCCGCGCAGCCGTAG	390 to 413
CEBPA/1R3	AGCTGCTTGGCTTCATCCTCCT	497 to 518
CEBPA/2F	GCCCCGGCGGCGCCGTCATGC	341 to 361
CEBPA/2F2	GCTGGTGATCAAGCAGGAGC	471 to 490
CEBPA/2R	CCGCCACTCGCGCGGAGGTCG	783 to 803
CEBPA/3F	GGCAGCGCGCTCAAGGGGCTG	748 to 768
CEBPA/3R[2]	CACGGCTCGGGCAAGCCTCGAGAT	*71 to *94
CEBPA/C693G/R(MM)	CGGGGTGCGGGCTGGGCACCGG	691 to 712
CEBPA/G964T/R(MM)	CTGTTCCACCCGCTTGCGCAGGCGGTAA	963 to 990
CEBPA/G157A/R(MM)	ATGGACGTCTCGTGCTCGCAGATGCCGCTC	156 to 185
CEBPA/Clone/F[3,4]	TGTTCCCTACCG**GAATTC**GGATCC**GCCACC****ATGGAGTCGGCCGACTTCTACG**	1 to 22
CEBPA/Clone-p30/F[3]	TGTTCCCTACCG**GAATTC**GGATCC**GCCACC****ATGCCCGGGGGAGCGCACGGGC**	358 to 379
CEBPA/Clone-UTR/F[3]	TGTTCCCTACCG**GAATTC**GGATCC**ACGCGGCCTGCCGGGTATAAAAGCT**	-165 to -141
CEBPA/Clone/R[3,4]	AAGGCCTTCGAC**CTCGAG****TCACGCGCAGTTGCCCATGGCC**	1056 to 1077
MSCV/F2	CTTGAACCTCCTCGTTCGAC	NA
MSCV/R2	CACACCGGCCTTATTCCAAGCG	NA

[1]Nucleotides in *CEBPA* covered by primer sequence. Numbered with 1 as A of ATG for p42 isoform, 5′ of ATG are -1, -2 etc. and 3′ of stop codon are *1, *2 etc. [2]Primers as per PP1F and PP2R from Pabst *et al* (2001b). [3]Primers were designed so that the 3′ end was complementary to *CEBPA*, in bold, and the 5′ end incorporated a restriction enzyme recognition site, either *Eco*RI (green) or *Xho*I (purple), and a Kozak consensus sequence (red), if required. [4]Primers were designed by Dr. Martin Pule of the Department of Haematology, UCL Cancer Institute.
For mismatch (MM) primers altered nucleotide is underlined. F and R indicate forward and reverse primers, respectively. NA indicates not applicable.

Appendix Table 2 *CEBPA* sequence alterations detected

Patient No.	DNA change 1	Predicted AA change 1	DNA change 2	Predicted AA change 2	*CEBPA*-single, -double or SNP
1	52dupC	H18fs	925_927del	E309del	double
2	63_64delinsA	S21fs	918_935del	N307_Q312del	double
3	68_78del	P23fs			single
4[a]	68delC	P23fs	68dupC	H24fs	single
5	68_78del	P23fs	912_913insTTG	K304_Q305insL	double
6	68dupC	H24fs			single
7	68dupC	H24fs	625delC	Q209fs	double
8	68dupC	H24fs	912_913insTTG	K304_Q305insL	double
9	68dupC	H24fs	912_920dup	K304_R306dup	double
10[#]	67_68dup	H24fs	925_926insTCG	V308dup	double
11	68dupC	H24fs	934_935insTTC	Q311_Q312insL	double
12	86_94delinsT	A29fs	943_945dup	L315dup	double
13	98_116del	F33fs	977_978ins66nt	K326delins23aa	double
14	103delC	R35fs	925_927dup	E309dup	double
15	107_113del	G36fs	543C>A	Y181X	double
16	105_120del	G36fs	890_934dup	R297_Q311dup	double
17	107delG	G36fs	910_936dup	K304_Q312dup	double
18	107_113del	G36fs	920_921insACAGCGCAA	R306_N307insKQR	double
19[#]	106_113del	G36fs	925delinsAAAACGC	E309delinsKTQ	double
20	111_124del	G38fs	929_934del	T310_Q312delinsK	double
21	113delG	G38fs	938_939insTA	K313fs	double
22	117delC	A40fs	925_951dup	E309_L317dup	double
23	126_132del	P43fs			single

Appendix Table 2 Continued

Patient No.	DNA change 1	Predicted AA change 1	DNA change 2	Predicted AA change 2	*CEBPA*-single, -double or SNP
24#	126_132del	P43fs	934_936dup	Q312dup	double
25	129dupC	A44fs	883G>C	A295P	double
26	135dupA	P46fs			single
27	146delC	P49fs	911_946dup	Q305_E316dup	double
28	148_167del	E50fs	905_943dup	V314_L315ins13aa	double
29	158_161del	G53fs	937_939dup	K313dup	double
30	155dupT	G54fs	499dupG	E167fs	double
31	158_159insTCCGCCAC	G54fs	911_916dup	Q305_R306insQQ	double
32	166dupT	C56fs	903_917dup	K302_R306dup	double
33	175G>T	E59X			single
34#	178dupA	T60fs			single
35	179_180dup	S61fs			single
36	183_184insGCACGTC	I62fs	925_926insCAC	E309delinsAQ	double
37	184_186delinsG	I62fs	986_987ins105nt	E329_Q330ins35aa	double
38	196dupG	A66fs			single
39	191_194dup	A66fs	949_950insGTC	E316_L317insR	double
40	198_201dup	I68fs			single
41	198_201dup	I68fs	890G>C	R297P	double
42	201_202del	I68fs	919_920insGGCGCA	R306_N307insRR	double
43	201dupC	I68fs	937_939dup	K313dup	double
44	230_233del	F77fs	917_934del	R306_Q311del	double
45	233delT	L78fs	896_973dup	R300_L324dupinsR	double
46	230_231dup	L78fs	922_981dup	V308_R327dup	double

Appendix Table 2 Continued

Patient No.	DNA change 1	Predicted AA change 1	DNA change 2	Predicted AA change 2	*CEBPA*-single, -double or SNP
47	230_231dup	L78fs	927_986dup	E309_V328dup	double
48	232delC	L78fs	937_939dup	K313dup	double
49	245_246insGTGTT	F82fs	909_923dup	K304_V308dup	double
50	247delC	Q83fs			single
51	247delC	Q83fs	898_936dup	R300_Q312dup	double
52[#]	247delC	Q83fs	937_939dup	K313dup	double
53	248dupA	H84fs	904_912dup	K302_K304dup	double
54[#]	271dupG	A91fs	956_957insACT	S319delinsRL	double
55	278dupC	A94fs			single
56	280_283dup	V95fs	707_708insT	A238Rfs	double
57[b]	282delC	V95fs	922G>T 934_936dup	V308L Q312dup	double
58	291_292insCC	T98fs	928_930dup	T310dup	double
59	296_302del	G99fs	912_913insTTG	K304_Q305insL	double
60	296_299dup	G101fs	934_936dup	Q312dup	double
61	315delC	F106fs	903_926dup	D301_V308dup	double
62	317_318dup	D107fs	897_956dup	S299_T318dup	double
63	318dupT	D107X	901_903del	D301del	double
64[#]	326_327insTA	G110fs	934_936dup	Q312dup	double
65	339_342del	G114fs	934_936dup	Q312dup	double
66[c]	343C>A	P115T	363_364ins19nt	G122fs	single
67	368delG	G123fs			single
68	496G>T	E166X			single

Appendix Table 2 Continued

Patient No.	DNA change 1	Predicted AA change 1	DNA change 2	Predicted AA change 2	*CEBPA*-single, -double or SNP
69	500_501dup	D168fs			single
70	500_501dup	D168fs			single
71†	533_548del	L178fs	827A>G	K276R	double
72	539delC	P180fs			single
73†	548C>A	P183Q			single
74†	558_566del	P187_P189del			single
75†	558_566del	P187_P189del			single
76†	578_589del	H193_P196del			single
77	644_665dup	G223fs			single
78	683delC	P228fs			single
79†	698C>G	P233R			single
80†	724G>A	G242S			single
81†	724G>A	G242S			single
82	745_748dup	G250fs			single
83	756_778del	L253fs			single
84	797_798insAG	S266fs			single
85	822delC	K275fs			single
86†	827A>G	K276R			single
87	854A>C	Y285S			single
88	825_873dup	N292fs			single
89	887T>A	V296E			single
90	899G>C	R300P			single
91†	899_901delinsAAA[HOM]	R300_D301delinsQN			double

Appendix Table 2 Continued

Patient No.	DNA change 1	Predicted AA change 1	DNA change 2	Predicted AA change 2	*CEBPA*-single, -double or SNP
92	899_901delinsAGT	R300_D301delinsQY			single
93	902_946dup	D301_L315dup			single
94†	917_943dup	R306_V314dup	962A>G	N321S	double
95[#]	914_917dup	N307fs			single
96	937_939dup	K313dup			single
97	937_939dup	K313dup			single
98	937_939dup	K313dup			single
99†	937_939dup	K313dup	1035_1045dup	S349fs	double
100	966_968del	R323del			single
101†[#]	941T>G [HOM]	V314G			double
102†	962A>G [HOM]	N321S			double
103	992T>A	L331Q			single
104	1021A>G	I341V			single
105	1027delC	R343fs			single
106	1066_1071del	N356_C357del			single
107	1066_1071del	N356_C357del			single
108	24G>A	E8E			SNP
109	111G>C	A37A			SNP
110	612G>C	P204P			SNP
111	612G>C	P204P			SNP
112	888G>A	V296V			SNP
113	*5C>T				SNP
114	*18C>T				SNP

Appendix Table 2 Continued

Patient No.	DNA change 1	Predicted AA change 1	DNA change 2	Predicted AA change 2	*CEBPA*-single, -double or SNP
115	*18C>T				SNP
116	*18C>T				SNP
117	*37G>A				SNP
118	*44A>T				SNP
119	*69G>A				SNP

Nucleotides numbered from the major translational start codon at nucleotide position 1. All mutations were heterozygous unless specifically stated otherwise. Abbreviations: AA, amino acid; HOM, homozygous mutations; ins, insertion; del, deletion; dup, duplication; nt, nucleotide; SNP, single nucleotide polymorphism. †Cases excluded from analysis of classical N- and/or C-terminal mutations. [a] Mutations on different alleles, however WT sequence still visible at about 50% level so patient scored as single. [b] Case has two C-terminal mutations, not known if on the same allele. [c] Both mutations on the same allele so case scored as single. The 19 bp insertion occurs after second start site so no functional p30 translated from mutated allele. [#] Mutations not present in samples taken during complete remission.