STUDIES TO INVESTIGATE EPIGENETIC FACTORS IN ACUTE MYELOID LEUKAEMIA

Dima Ahmed El-Sharkawi

UCL

A thesis submitted for the Degree of Doctor of Philosophy

2016

DECLARATION

I, Dima Ahmed El-Sharkawi, 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 heterogeneous disease with numerous recurrent cytogenetic and molecular abnormalities. This heterogeneity is reflected in the variation in clinical outcome seen in patients. This disparity in outcome is also seen within groups of patients who have the same mutation or no known molecular abnormalities.

To investigate whether the DNA methylation profile of samples can provide prognostic information, the methylome of forty cytogenetically normal AML samples that were wild-type for *NPM1* and *FLT3* was analysed, 20 were from patients with chemosensitive disease and 20 with chemoresistant disease. Unsupervised cluster analysis revealed the DNA methylation profile to be most associated with underlying *CEBPA* genotype hence a *CEBPA* signature was created using the 25 CpG sites that differed the most between wild-type (n=30) and classic *CEBPA*^{DM} (double mutant) samples (n=10). Two follow-up cohorts were analysed, validating the initial signature in differentiating classic *CEBPA*^{DM} samples from wild-type. *CEBPA*^{SM} (single mutant) samples had profiles more similar to the *CEBPA*^{WT} (wild-type) signature. Non-classic *CEBPA*^{DM} samples with at least one mutation leading to loss of function of the C terminal were associated with a *CEBPA* mutant methylation profile. Methylation of the *CEBPA* promoter was not associated with a classic *CEBPA*^{DM} methylation profile in eight of the nine cases exhibiting hypermethylation.

The *ASXL1* gene, known to have a role in histone regulation, was screened in 371 patients using denaturing HPLC. The overall mutation rate was 9%. Overall survival was significantly lower in patients with an *ASXL1* mutation, however the mutation was associated with secondary disease and older age, and thus in multivariate analysis mutations in *ASX L1* lost significance.

These studies indicate that epigenetic factors are closely linked to other prognostic traits such as age or underlying molecular status of the AML. Given this association, DNA methylation could play an important role in assessing the significance of different types of mutations.

ACKNOWLEDGEMENTS

I would like to thank Professor Rosemary Gale for her help and support throughout the PhD, I have learnt so much from her methodical approach to research and meticulousness in scientific writing. I am incredibly grateful to Professor David Linch who gave me the opportunity to undertake this PhD. He gave me very useful advice and guidance, and helped me understand the importance of asking the right question. I would also like to thank Dr Duncan Sproul for the bioinformatic analysis of the array data. Thanks to everyone in lab 305, who taught me a great deal and helped make the PhD so enjoyable. This work would not have been possible without the patients and clinical investigators involved in the trials.

Thank you to mum, dad, Lamah and Reem for their continuous love, encouragement and enthusiasm in trying to understand my research. Finally, thank you to my team, Trevor, Sammy and Tamara.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

COMMONLY USED ABBREVIATIONS

AA	Amino acid
AML	Acute Myeloid Leukaemia
APL	Acute Promyelocytic Leukaemia
aCGH	Array comparative genomic hybridisation
Asx	Additional sex combs
ASXL1	Additional sex comb like 1 gene
ATRA	All-trans-retinoic acid
BM	Bone Marrow
bp	Base pairs
CBF	Core Binding Factor
cDNA	Complementary DNA
CEBPA	CCAAT-enhancer binding protein α
CGI	CpG Island
ChIP	Chromatin Immunoprecipitation
CI	Confidence intervals
CIR	Cumulative Incidence of Relapse
CLP	Common Lymphoid Progenitor
CMP	Common Myeloid Progenitor
CpG	Cytosine (phosphate) Guanine dinucleotide
CR	Complete Remission
DBD	DNA-binding domain
dHPLC	Denaturing high performance liquid chromatography
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DNMT3A	DNA methyltransferase 3A
FAB	French-American-British
FLT3	Fms-like tyrosine kinase-3
gDNA	Genomic DNA
GMP	Granulocyte/macrophage progenitor
GVL	Graft-versus-leukaemia

HELP	*Hpa*II tiny fragment Enrichment by Ligation-mediated PCR
HLA	Human Leukocyte Antigen
HR	Hazard Ratio
ID	Induction death
IDH	Isocitrate Dehydrogenase
IK	Intermediate Karyotype
ITD	Internal Tandem Duplication
LIC	Leukaemia initiating cell
LZD	Leucine zipper domain
MDS	Myelodysplasia
MeDIP	Methylated DNA immunoprecipitation
MEP	Megakaryocyte/erythroid progenitors
MHC	Major histocompatibility complex
MRC	Medical Research Council
NGS	Next Generation Sequencing
NK	Normal Karyotype
NPM1	Nucleophosmin 1
nt	Nucleotide
OR	Odds Ratio
OS	Overall Survival
PCR	Polymerase chain reaction
PRC2	Polycomb Repressive Complex 2
RD	Resistant Disease
RED	Restriction Enzyme Digest
RNA	Ribonucleic acid
SCT	Stem Cell Transplantation
SNP	Single Nucleotide Polymorphism
TA	Transactivation activity
TAD	Transactivation Domain
TKD	Tyrosine Kinase Domain
TSS	Transcription start site
WHO	World Health Organization
WT	Wild-type

CHAPTER 1:**INTRODUCTION**

Acute myeloid leukaemia (AML) is a heterogeneous malignancy characterised by an accumulation of myeloid precursors. Part of the heterogeneity is explained by the different chromosomal changes and mutations seen in this disease. However, these underlying genetic changes do not account for the full diversity seen, both clinically and biologically, and recent studies have demonstrated the important role of epigenetic factors in contributing to its pathogenesis. This thesis presents studies exploring epigenetic differences identified in AML and mutations in a gene coding for an epigenetic modifier, with a view to determining whether they can aid stratification of patients into prognostic groups.

1.1 Haematopoiesis

Mature blood cells have a life span ranging from a few days to a few months and thus are continually replenished throughout life. This continual source of blood cells is maintained by a few haematopoietic stem cells which have the ability to self-renew and also differentiate further into more committed progenitor cells, culminating in terminally differentiated mature blood cells (Orkin & Zon, 2008). This hierarchical process is summarised in Figure 1.1. Haematopoiesis is a co-ordinated process regulated by specific growth factors and transcription factors. It occurs in waves at different sites that alter during development; in adults the location of haematopoiesis is predominantly in the bone marrow of the central skeleton and proximal parts of the femurs and humeri.

1.2 AML

AML is a relatively uncommon malignancy of haematopoietic cells of the myeloid lineage; the incidence is 2-3 cases per 100,000 individuals per year in the population less than 60 years of age, but it increases to 13-15 in the population over 60 years of age (Lowenberg *et al*, 1999). Patients present with symptoms of pancytopenia such as fatigue, breathlessness, bleeding, bruising and recurrent infections; leukaemic infiltration which can affect anywhere but particularly the spleen, liver, gums, skin and nervous system; or leukostasis

(COPYRIGHT PROTECTED IMAGE REMOVED)

Figure 1.1 Summary of haematopoiesis. The stages at which haematopoietic development is blocked in the absence of a given transcription factor 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.

Taken from Orkin and Zon (2008).

which can lead to neurological or pulmonary complications. Usually symptoms develop rapidly, over days to weeks. Although there are no known causes of AML, there are several factors that increase the risk of developing it, and this is reflected in the classification of the disease.

1.2.1 Diagnosis and Classification of AML

A diagnosis of AML is made if 20% or more of nucleated blood cells in the bone marrow are myeloid blasts, although cases with fewer than 20% blasts that also have an underlying cytogenetic rearrangement such as inv(16), t(8;21) or t(15;17) are also diagnosed as AML. The current classification is based on the World Health Organisation (WHO) proposal (Swerdlow *et al*, 2008) which is summarised in Table 1.1. This incorporates the previous French-American-British classification, which was based only on morphological features and cytochemistry, but has been extended to include subtypes related to predisposing clinical or underlying genetic factors. The main subtypes are AML with recurrent genetic abnormalities; AML with myelodysplasia-related changes; therapy-related myeloid neoplasms; myeloid proliferations related to Down syndrome; AML, not otherwise specified.

1.2.2 Prognostic Factors in AML

Prognostic factors can be grouped into patient-related, disease-related and response to therapy.

1.2.2.1 Patient-related factors

Studies have shown that a higher WHO performance score or the presence of co-morbidities at diagnosis are independent risk factors associated with a poorer outcome (Döhner *et al*, 2010; Giles *et al*, 2007). However, age at diagnosis is the strongest patient-related factor, with the prognosis gradually getting worse with increasing age. The 5-year overall survival rate of patients diagnosed below the age of 14 years is 63%, compared to 35% for those aged 45-59 years and only 14% for those over 60 years (Smith *et al*, 2011). The reason for this is multi-factorial; older patients are more likely to have underlying co-morbidities and worse performance score, as well as differences in ability to tolerate chemotherapy and the complications of chemotherapy. Furthermore,

Table 1.1 WHO classification of AML

Acute myeloid leukaemia with recurrent genetic abnormalities

- AML with t(8;21)(q22;q22); RUNX1-RUNX1T1
- AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); CBFB-MYH11
- APL with t(15;17)(q22;q12); PML-RARA
- AML with t(9;11)(p22;q23); MLLT3-MLL
- AML with t(6;9)(p23;q34); DEK-NUP214
- AML with inv(3)(q21q26.2) or t(3;3)(q21;q26.2); RPN1-EVI1
- AML (megakaryoblastic) with t(1;22)(p13;q13); RBM15-MKL1
- Provisional entity: AML with mutated NPM1
- Provisional entity: AML with mutated CEBPA

Acute myeloid leukaemia with myelodysplasia-related changes

Therapy-related myeloid neoplasms

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
  - Pure erythroid leukaemia
  - Erythroleukemia, erythroid/myeloid
- Acute megakaryoblastic leukaemia
- Acute basophilic leukaemia
- Acute panmyelosis with myelofibrosis

Myeloid sarcoma

Myeloid proliferations related to Down syndrome

- Transient abnormal myelopoiesis
- Myeloid leukaemia associated with Down syndrome

Blastic plasmacytoid dendritic cell neoplasm

the disease biology itself is different in the older age group, with a higher frequency of adverse cytogenetics and higher incidence of secondary AML. However, even when all these factors are taken into account, age is still an independent prognostic variable (Liersch *et al*, 2014).

1.2.2.2 Disease-related factors

Certain factors such as previous chemotherapy or myelodysplastic changes not only increase the risk of developing AML (secondary AML), but are also associated with a worse prognosis when compared to AML arising *de novo* (Arber *et al*, 2003), hence patients in these categories are considered separately in the classification system.

The majority of patients develop AML *de novo*. Approximately 60% of these have recurrent cytogenetic alterations, many of which are of major prognostic importance and have been used as the basis of risk-stratified treatment for over 20 years. A recent cytogenetic analysis of 5876 patients aged between 15 and 59 years treated in UK MRC trials found that approximately one-quarter of patients had either t(15;17), inv(16) or t(16;16), or t(8;21) and were in a favourable prognostic group (Table 1.2) (Grimwade *et al*, 2010). The latter two translocations lead to CBFB-MYH11 and RUNX1-RUNXT1 fusion proteins respectively and are termed core-binding factor leukaemias; patients with these changes had a 10-year overall survival of approximately 60%. The 17% of patients with poor prognostic changes, as specified in Table 1.2, had a 10-year survival of 12%. The remaining 58% of patients either had no karyotypic change or changes with no known prognostic impact, and this group had a 10-year overall survival of approximately 35%.

However, an increasing number of recurrent mutations has now been found in AML, some of which have prognostic impact (Table 1.3). The list given is not exhaustive and many new less common mutations have been identified, particularly with the increase in availability of whole genome sequencing (2013). One of the most commonly found mutations in AML is an internal tandem duplication in the FMS-like tyrosine kinase 3 gene (*FLT3*/ITD). *FLT3* encodes a receptor tyrosine kinase that is expressed in early haematopoietic cells.

Table 1.2 Cytogenetic risk group classification based on analysis of young adult patients with AML treated on UK MRC protocols

Risk Group	10-year OS[2]	Proportion of patients[2]	Original MRC[1]	Refined MRC[2]
Favourable	69%	7%	t(8;21)	t(8;21)
		13%	t(15;17)	t(15;17)
		5%	inv(16) or t(16;16)	inv(16) or t(16;16)
Intermediate	38%	41%	Normal karyotype	Normal karyotype
	33%	17%	Other non-complex	Other non-complex
Adverse	12%	17%	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, excluding those with favourable changes	Complex ≥4 unrelated abn, excluding those with favourable changes

[1](Grimwade *et al*, 1998) [2](Grimwade *et al*, 2010)

Abbreviations: OS, overall survival; MRC, Medical Research Council; abn, abnormality; del, deletion; add, addition; t, translocation.

Table 1.3 Recurrent mutations seen in AML

Gene	Chr	Proposed Function	Reported Incidence	Type of mutations commonly detected	Overall impact on outcome	Associations	References
ASXL1	20q11	Additional sex comb like 1 - Stablilises PRC2 which catalyses repressive histone trimethylation mark H3K27Me3	3-19%	Frameshift Nonsense	Adverse	Older adults Secondary disease Inv. with *NPM1*^{MUT}	Chou *et al* (2010b) Paschka *et al* (2011) Pratcorona *et al* (2012) Schnittger *et al* (2013) El-Sharkawi *et al* (2013)
BCOR	Xp11.4	BCL-6 co-repressor	4% (NK-AML)	Frameshift Nonsense Splice site	Adverse	NK-AML *DNMT3A*^{MUT} Inv. with *NPM1*^{MUT}	Grossmann *et al* (2011b)
BCORL1	Xq25-26.1	Transcriptional co-repressor	6%	Frameshift Nonsense Missense Splice site	N/A		Li *et al* (2011)
CBL	11q23.3	Casitas B cell lymphoma gene- involved in the degradation of tyrosine kinases	1-3%	Missense	N/A	CBF-leukaemia	Bacher *et al* (2010) Allen *et al* (2011) Ibanez *et al* (2012)

Table 1.3 continued

Gene	Chr	Proposed Function	Reported Incidence	Type of mutations commonly detected	Overall impact on outcome	Associations	References
CEBPA	19q13.1	Transcription factor promoting cell differentiation	6%	Frameshift or nonsense at N terminus In-frame insertion or deletion at C terminus Often biallelic	Favourable (*CEBPA*^{DM} only)	NK-AML	Hou *et al* (2009) Wouters *et al* (2009) Dufour *et al* (2010) Green *et al* (2010b) Taskesen *et al* (2011)
c-KIT	4q12	Tyrosine kinase receptor	2-3%	MIssense (D816 or N822) Inframe indel in extracellular domain	Adverse	CBF-leukaemia	Care *et al* (2003) Boissel *et al* (2006) Cairoli *et al* (2006) Paschka *et al* (2006) Schnittger *et al* (2006) Allen *et al* (2011) Chen *et al* (2016)
DNMT3A	2p23	DNA methyltransferase 3A: catalyses *de novo* methylation of CpG dinucleotides	15-25%	Missense (predominantly R882) Nonsense Frameshift Splice site	Adverse	FAB M4/ M5 IR-AML/ NK-AML *NPM1*^{MUT}	Ley *et al* (2010) Thol *et al* (2011a) Yan *et al* (2011) Marcucci *et al* (2012) Renneville *et al* (2012) Ribeiro *et al* (2012) Gale *et al* (2015)

Table 1.3 continued

Gene	Chr	Proposed Function	Reported Incidence	Type of mutations commonly detected	Overall impact on outcome	Associations	References
EZH2	7q35-q36	Histone lysine N-methyltransferase- member of the polycomb protein family involved in histone methylation and chromatin silencing	1-3%	Nonsense Frameshift Missense	N/A	Del7q	Wang *et al* (2013b)
FLT3	13q12	Tyrosine kinase receptor	25% 7%	FLT3/ITD FLT3/TKD (Missense)	Adverse No impact	*NPM1* ^{MUT} NK-AML	Kottaridis *et al* (2001) Thiede *et al* (2002) Mead *et al* (2007) Gale *et al* (2008)
GATA2	3q21.3	Transcription factor involved in haematopoiesis	3% in *CEBPA*^{WT} 18-27% in *CEBPA*^{DM}	Missense	Unclear: Favourable/ No impact	*CEBPA* double mutant AML	Fasan *et al* (2013b) Green *et al* (2013) Grossmann *et al* (2013) Pasquet *et al* (2013)

Table 1.3 continued

Gene	Chr	Proposed Function	Reported Incidence	Type of mutations commonly detected	Overall impact on outcome	Associations	References
IDH1	2q33.3	Isocitrate Dehydrogenase 1: Cytosolic metabolic enzyme catalyses the conversion of isocitrate to α ketoglutarate	7-16%	Missense at R132	Unclear: Adverse/No impact/ Favourable in *FLT3*/ITD positive subgroup	NK-AML *NPM1* ^{MUT} inv. with *CEBPA* ^{MUT}, *TET2* ^{MUT} and *WT1* ^{MUT}	Abbas *et al* (2010) Boissel *et al* (2010) Chou *et al* (2010a) Green *et al* (2010a) Marcucci *et al* (2010) Paschka *et al* (2010) Wagner *et al* (2010) Nomdedeu *et al* (2012)
IDH2	15q26.1	Isocitrate Dehydrogenase 2: Mitochondrial metabolic enzyme catalyses the conversion of isocitrate to α ketoglutarate	8-15%	Missense at R140 or R172	Unclear: Adverse/No impact/ Favourable	*NPM1* ^{MUT} Inv. with *TET2* ^{MUT} and *WT1* ^{MUT}	Abbas *et al* (2010) Boissel *et al* (2010) Chou *et al* (2011b) Green *et al* (2011) Paschka *et al* (2010)
JAK2	9p24	Janus Kinase 2: implicated in cell signalling	1%	V617F	Adverse	CBF-leukaemia	Fröhling *et al* (2006) Illmer *et al* (2007)
KRAS	12p12.1	GTPase	5-20%	Missense	No impact	Inv(16)	Bowen *et al* (2005) Rocquain *et al* (2010)
MLL	11q23	Histone methyltransferase: mediates H3K4me, active transctiptional mark	5-10%	PTD	Adverse	NK-AML	Schnittger *et al* (2000) Dohner *et al* (2002) Steudel *et al* (2003)

Table 1.3 continued

Gene	Chr	Proposed Function	Reported Incidence	Type of mutations commonly detected	Overall impact on outcome	Associations	References
NPM1	5q35.1	Nucleolar protein	20-35%	Frameshift	Favourable	*FLT3*	Dohner *et al* (2005) Schnittger *et al* (2005) Verhaak *et al* (2005) Thiede *et al* (2006)
NRAS	1p13.2	GTPase	5-20%	Missense	No impact		Bacher *et al* (2006) Bowen *et al* (2005)
PHF6	Xq26.3	Plant Homeodomain finger 6:X linked transcriptional regulator	2-4%	Nonsense Frameshift	Adverse	Male sex	Van Vlierberghe *et al* (2011)
PTPN11	12q24	Encodes SHP2, a non-tyrosine phosphatase signalling molecule	3%	Missense	N/A	Paediatric AML	Loh *et al* (2004) Hugues *et al* (2005)
RUNX1	21q22.3	Runt related transcription factor 1: transcription factor required for haematopoiesis	10-15%	Frameshift Nonsense Missense	Adverse	CN-AML *MLL*^{MUT} Inv. with *CEBPA*^{MUT} and *NPM1*^{MUT}	Tang *et al* (2009) Gaidzik *et al* (2011) Mendler *et al* (2012)
TET2	4q24	Catalyses conversion of methylcytosine to hydroxymethylcytosine	8-12%	Frameshift Nonsense Missense	Adverse	NK AML Inv. with *IDH1*^{MUT}, *IDH2*^{MUT} and *WT1*^{MUT}	Chou *et al* (2011a) Metzeler *et al* (2011b) Gaidzik *et al* (2012)

Table 1.3 continued

Gene	Chr	Proposed Function	Reported Incidence	Type of mutations commonly detected	Overall impact on outcome	Associations	References
TP53	17p13.1	Tumour suppressor	2-5%	Frameshift Nonsense Missense	Adverse	Complex karyotype Secondary AML	Bowen *et al* (2009) Rucker *et al* (2012) Wong *et al* (2015)
WT1	11p13	Transcription factor that acts as a tumour suppressor Recruits and aids *TET2*	10-15%	Frameshift Missense Nonsense	Unclear: adverse/ no impact	*FLT3/ITD* *CEBPA*^{MUT} Inv. with *IDH1*^{MUT}, *IDH2*^{MUT} and *TET2*^{MUT}	Paschka *et al* (2008) Virappane *et al* (2008) Gaidzik *et al* (2009) Rampal *et al* (2014)

Abbreviations: CBF, core binding factor; *CEBPA*^{DM}, Double mutation in *CEBPA;* Chr, chromosome; Inv., inversion; Inv. with, inverse association with; IR-AML, intermediate risk AML; ITD, internal tandem duplication; MUT, mutation; N/A, data not available; NK-AML, normal karyotype AML; PTD, partial tandem duplication; TKD, tyrosine kinase domain; WT, wild type.

A *FLT3*/ITD, seen in approximately 25% of AML patients, leads to a constitutively active receptor and is associated with a poor prognosis (Kottaridis *et al*, 2001; Thiede *et al*, 2002). A second type of mutation within this gene, either a missense mutation or small in-frame size change within the tyrosine kinase domain (*FLT3*/TKD), is less frequent and its impact on prognosis is not as clear (Mead *et al*, 2007; Thiede *et al*, 2002; Whitman *et al*, 2008). Another gene frequently mutated in patients with AML is nucleophosmin (*NPM1*), which encodes a nucleolar transporter protein. Mutations in *NPM1* are found in approximately 20-35% of cases, and in the absence of a *FLT3*/ITD are associated with a good prognosis (Gale *et al*, 2008; Schnittger *et al*, 2005; Shen *et al*, 2011; Thiede *et al*, 2006). Mutations in the CCAAT/ enhancer-binding protein alpha gene, *CEBPA*, are found in approximately 10% of patients, the majority of whom have an intermediate karyotype (IK) and are wild-type (WT) for *NPM1*, and are associated with a good prognosis. They are discussed further in Chapter 3.

More recent studies have identified recurrent mutations in genes that encode proteins involved with maintaining the “epigenetic equilibrium”. These mutations in epigenetic modifiers are thought to exert their effects via changes in histone structure, e.g. methylation, ubiquitination, phosphorylation and DNA methylation, which in turn are thought to lead to global alterations in gene expression. Examples include the DNA methyltransferase 3A (*DNMT3A)*, Ten Eleven Translocation methylcytosine dioxygenase 2 (*TET2*), Isocitrate Dehydrogenase 1 and 2 (*IDH1*, *IDH2*), Additional sex comb like 1 (*ASXL1*) and *EZH2* genes. The significance of changes in methylation in cancer is further discussed in section 1.3.6, and of these mutations in chapter 5.

It should be noted that as more is discovered about the underlying cytogenetic and molecular mutations that occur recurrently in AML, the factors that are considered to be associated with good or poor prognosis are constantly coming under review. As more mutations are identified, assessing their impact on prognosis is becoming more complex due to their co-incidence, particularly where factors associated with both good and poor prognosis are both present, for example *FLT3*/ITDs and *NPM1* mutations. Furthermore, as new therapies

are introduced, particularly targeted therapies, this may change how prognostic factors are interpreted.

1.2.2.3 Response to therapy

Patients who are in complete remission following the first cycle of induction chemotherapy have a better prognosis than those who still have detectable disease (Wheatley *et al*, 1999). The likelihood of achieving a complete remission is dictated in part by the risk factors discussed above, for example age. The length of remission has also been associated with outcome; those patients who relapse within 6 months of induction therapy have a worse outcome compared to those who relapse after 6 months (Döhner *et al*, 2010). These studies define remission morphologically. However, with the identification of recurrent mutations, use of more sensitive molecular methods to assess minimal residual disease are becoming increasingly more common, and a recent study has shown that patients with persistent subclinical disease do worse than those who are in molecular remission as defined by these methods (Ivey *et al*, 2016).

1.2.3 Treatment of AML

The mainstay of treatment of AML is chemotherapy using protocols that have not changed dramatically over the past few decades. There are two phases of chemotherapy; the first is remission induction and the second is consolidation. Standard remission induction for non-acute promyelocytic leukaemia (APL), in patients who can tolerate intensive chemotherapy, consists of a nucleoside analogue, usually cytarabine, in combination with an anthracycline such as daunorubicin. Several studies have looked at different chemotherapeutic agents, dosing schedules or additional drugs, but to date these changes have not led to a significant survival advantage. The aim of this phase of chemotherapy is to induce a remission, CR, defined as less than 5% blasts in the bone marrow, detected morphologically, with recovery of blood counts. For patients less than 60 years of age, approximately 80% will achieve CR, 10-20% will have persistent disease and approximately 5% will die as a result of complications of the disease and/or treatment (Burnett, 2013).

Once in remission, the aim of the second phase of chemotherapy is to prevent relapse by eliminating residual subclinical disease. The most common agent given for this, outside of a trial setting, is high-dose cytarabine. However, again, the number of cycles to give and alternative combination chemotherapies are under investigation. Stem cell transplants (SCT) are an important tool in preventing relapse, however this comes at a cost as they are associated with an increased risk of treatment-related morbidity and mortality. There is no current evidence that autologous SCT are superior to chemotherapy alone in outcome, and so these are not standard practice (Burnett *et al*, 2011). Several factors are involved in deciding whether an allogeneic SCT should be performed in first remission, the first of which is donor availability, either a well-matched sibling or unrelated donor. A second consideration is whether the patient is fit enough to tolerate the procedure. The introduction of reduced intensity allografts, which use less myeloablative chemotherapeutic regimes and rely more on the graft-versus-leukaemia effect, has meant that this option can be used in older less fit patients than previously. Thirdly, the choice is based on risk stratifying the disease, which in practice is predominantly done using cytogenetics. Generally, if the risk of relapse is greater than 35% at 4 years post-remission, then it is considered high enough to outweigh the risks of the transplant itself (Cornelissen *et al*, 2007). Standard current practice in the UK is that patients with adverse risk cytogenetics are recommended for a SCT in first remission, whereas those with favourable risk cytogenetics or who have a mutation in *NPM1* in the absence of *FLT3*/ITD are not. Other patients within the intermediate risk category are usually considered for a SCT, but there is no clear consensus whether this is the best option, perhaps reflecting the heterogeneity of this group (Burnett & Hills, 2011).

The overall survival (OS) rate in younger patients treated with standard therapy has steadily improved over the past few decades from a 5-year survival rate of 25% to 40% (Burnett, 2013). This improvement is primarily due to decreased treatment-related mortality because of better supportive care and treatment of complications rather than due to changes in the chemotherapy itself. Unfortunately, this improvement has not been replicated in the older age group, as survival rate in patients over 60 years of age still remains at approximately

20% at 5 years (Burnett, 2013). Even those patients who are considered fit enough to receive the same intensive chemotherapy as younger patients have worse outcomes when compared to the younger patients undergoing the same therapy.

Research into new chemotherapy drugs is ongoing, however currently most clinical trials assessing new agents have adopted a more targeted approach to treatment. Examples include antibody-drug conjugates such as gemtuzumab ozogamicin, which targets CD33 found on leukaemic blasts, and drugs that target particular mutations, for example *FLT3* inhibitors such as lestaurtinib (Stein & Tallman, 2016). Drugs aimed at epigenetic targets are also of particular interest given that epigenetic dysregulation is a feature of AML (see Chapter 3), and the potential reversibility of this dysregulation. The most widely used drugs for AML in this category are the pyrimidine analogues 5-azacytidine (azacitidine) and 5-aza-2'deoxycytidine (decitabine). These function as inhibitors of DNA methyltransferases by being incorporated into the DNA as cytosine analogues, leading to hypomethylation at lower doses and preventing cell proliferation at higher doses (Leone *et al*, 2002). There is some evidence that these drugs are effective, especially in the older age group when compared to current conventional care. One study showed an improvement in overall survival in patients taking azacitidine compared to those who did not,10.4 months and 6.5 months respectively, $p = 0.10$ (Dombret *et al*, 2015). However, there is little evidence that these agents lead to a reversal of aberrant methylation and restoration of the expression of critical tumour suppressor genes (Voso *et al*, 2014). There are also several drugs in clinical phase trials that act on enzymes affecting histone modifications, such as histone deacetylase inhibitors, and on proteins that "read" acetyl marks, such as BET inhibitors (Wouters & Delwel, 2016).

1.3 Epigenetics

Epigenetics is defined as heritable changes in gene expression that are not due to alterations in the DNA sequence itself (Holliday, 1987). More recently it has been suggested that this should encompass "the structural adaptation of

chromosomal regions so as to register, signal or perpetuate altered activity states" (Bird, 2007). This subtle change allows for the inclusion of chromatin marks that may alter gene expression but are transient. As all cells have the same DNA, the epigenome dictates which genes are active and therefore why gene expression is so varied in different cell types. There are currently three systems that are known to have an effect on gene expression and thus are considered "epigenetic", namely DNA methylation, histone modifications, and RNA-associated silencing (which will not be discussed).

1.3.1 DNA methylation

The most studied epigenetic marker is DNA methylation. In 1975, two papers were independently published detailing methylation of cytosines, which involves the addition of a methyl group to the 5' carbon of the cytosine bases in the context of CpG dinucleotides. They proposed it could lead to gene silencing and thus play a role in the development of organisms by regulating gene expression (Holliday & Pugh, 1975; Riggs, 1975). There is evidence of methylation of cytosine bases in non-CpG contexts but this is far less frequent and its role is currently not clear, although studies have demonstrated that it may have an important function in embryonic stem cells and brain tissue (Guo *et al*, 2014; Ramsahoye *et al*, 2000; Ziller *et al*, 2011).

1.3.1.1 CpG dinucleotides

As methylation was found predominantly in the cytosines of CpG dinucleotides, this led to an interest in the distribution of CpGs within the genome. Although there are approximately 28 million CpG sites in the human genome, based on the total GC content, there are fewer CpG sites than would be expected by chance alone (Russell *et al*, 1976). Moreover, the distribution of CpG dinucleotides is not random. Most of the genome is deplete of CpG dinucleotides, however there are CpG-rich regions (termed CpG islands, CGIs) that are mainly concentrated around the promoter region and first exon of 60-70% of genes in the human genome (Saxonov *et al*, 2006). A formal definition of a CGI is that it is a sequence of at least 200 bases with a GC content greater than 50%, and the observed to expected ratio of CpG sites is greater than 60% (Gardiner-Garden & Frommer, 1987). Given the strong link between CGIs and

promoter regions, CGIs are sometimes used to elucidate promoters and genes that are hitherto unknown (Illingworth *et al*, 2010; Larsen *et al*, 1992). For example, Macleod *et al* (1998) discovered a novel intronic promoter within the MHC class II-I Aβ gene in mice which, when deleted, led to decreased transcription of the gene.

Differences have been observed between genes that have CGIs within their promoters and those that do not. All constitutively expressed genes have a CGI located around the transcription start site, and approximately 40% of genes that show differential expression across tissue types have an associated CGI, although the location of the CGI in these cases is more variable (Larsen *et al*, 1992; Zhu *et al*, 2008). Furthermore, CGI promoters are able to initiate transcription from multiple positions, whereas non-CGI promoters are generally associated with a single initiation site (Sandelin *et al*, 2007).

1.3.1.2 Distribution of methylation of CpG sites

In normal human cells, the CpGs within islands are often unmethylated whereas CpG sites outside the CGIs are generally methylated (Weber *et al*, 2007). Although the density of CpGs is less outside CGIs, given the size of the genome this still amounts to approximately 70% of the total CpGs being methylated (Ehrlich *et al*, 1982). The relative paucity of CpG sites throughout the genome is thought to be due to spontaneous deamination of methylated cytosines to form thymine (Holliday & Grigg, 1993); conversely CGIs are thought to be “protected” as they usually remain unmethylated (Smallwood *et al*, 2011).

Although most CGIs associated with genes remain unmethylated, there are a subset that are methylated, and this can be tissue specific (Shen *et al*, 2007). For example, analysis of chromosomes 6, 20 and 22 identified 11 CGIs out of 2279 which were differentially methylated in eight different tissue types (Eckhardt *et al*, 2006). Investigation of the methylation status of CGIs on chromosome 21q revealed 31 out of 149 CGIs (21%) detected were fully methylated in normal peripheral blood cells (Yamada *et al*, 2004). Interestingly, the most variation in methylation levels seen across tissue types is in areas of intermediate rather than high CpG density. Irizarry *et al* (2009) found that most

variation occurred in regions up to 2 kilobases away from the CGIs, which they named CpG shores. These tissue-specific differentially methylated regions were not only conserved between different individuals analysed, but also could be extrapolated to mouse tissues, indicating remarkable conservation of the methylation pattern across species.

1.3.1.3 Regulation of DNA methylation

Given the stability of the methylation distribution across the genome, many studies have investigated how DNA methylation is initiated and maintained. The addition of methyl groups to cytosine is catalysed by the DNA methyltransferase (DNMT) family of enzymes, of which DNMT1, DNMT3A and DNMT3B are considered the most important (Bestor, 1988). Traditionally, DNMT3A and DNMT3B were thought of as *de novo* methyltransferases and the function of DNMT1 was to maintain methylation patterns due to its affinity for hemimethylated DNA and its ability to target foci of DNA replication (Bestor & Ingram, 1983; Leonhardt *et al*, 1992). However, all three are necessary for embryonic development (Chen *et al*, 2003; Li *et al*, 1992; Okano *et al*, 1999) and maintenance of methylation pattern (Jones & Liang, 2009).

During early embryonic development, there is an initial loss of methylation followed by a period of global *de novo* methylation, to which CGIs remain immune (Monk *et al*, 1987). The mechanism by which CGIs remain unmethylated is unknown. Evidence exists that this may be due to bound transcription factors at the CGIs that preclude the methylation of the underlying island. For example, ablation of the binding site for the transcription factor Sp1 facilitates *de novo* methylation of the *APRT* promoter CGI (Brandeis *et al*, 1994). Another hypothesis is that the methyl marks are removed from CGIs by active demethylation (Frank *et al*, 1991).

The process of demethylation is not as well understood. Although none have been identified, demethylases that can actively remove the methyl group from DNA are thought to exist (Ooi & Bestor, 2008). Many papers have looked at other mechanisms by which methyl groups are removed, for example, the TET proteins catalyse the oxidation of 5-methylcytosine to 5-hydroxymethylcytosine, which is thought to be an intermediate in the subsequent demethylation of the

cytosine base (Tahiliani *et al*, 2009). TET proteins can also cause further oxidation to produce 5-formylcytosine and 5-carboxylcytosine (Ito *et al*, 2011). Absence of TET3 in mice led to a failure of demethylation of CpG sites in key genes such as *Nanog* involved with embryogenesis, leading to a delay in development (Gu *et al*, 2011). Furthermore, during embryonic stem cell differentiation, the levels of TET1 and TET2 proteins decrease with a concomitant decrease in 5-hydoxycytosine and increase in 5-methylcytosine. Knockdown of *TET1* and *TET2* leads to downregulation of genes, including those associated with pluripotency, and an increase in methylation at the promoters of these genes (Ficz *et al*, 2011). However, TET proteins have also been implicated in repression of genes through interactions with other proteins such as the SIN3A co-repressor complex (Williams *et al*, 2011).

1.3.2 The significance of DNA methylation

Methylation is essential for viability of somatic cells, as cultured fibroblasts with Cre-mediated deletion of *DNMT1* underwent apoptosis (Jackson-Grusby *et al*, 2001), however there are many questions as to its exact function. Most work to investigate the role of DNA methylation has concentrated on investigation of CGIs within promoter regions of genes. The initial reports that described DNA methylation postulated its role was to silence gene expression (Holliday & Pugh, 1975; Riggs, 1975). Indeed, in vitro experiments showed that methylation of the adenine phosphoribosyltransferase gene inhibited its expression when transduced into mouse L cells (Stein *et al*, 1982). Since then methylation of CGIs has been linked with X chromosome inactivation in females, gene imprinting, and tissue-specific gene expression/repression (Illingworth *et al*, 2008; Li *et al*, 1993; Venolia & Gartler, 1983), all of which are associated with allele or gene silencing. Moreover, treatment of mammalian cells with the demethylating agent 5'-aza 2'-deoxycytidine has been shown to re-activate silenced genes on the inactive X chromosome (Mohandas *et al*, 1981).

However, there is still some debate about whether DNA methylation is the cause of repression (De Smet *et al*, 1999) or is rather a result of repression. For example, methylation of the *Hrpt* gene on the inactive X chromosome occurs after chromosome inactivation (Lock *et al*, 1987), and silencing of the X

chromosome is not dependent on DNMTs (Sado *et al*, 2000; Sado *et al*, 2004). In some circumstances, methylation is thought to stabilise DNA silencing, as when X inactivation takes place in extra-embryonic tissue; if there is no DNA methylation, the genes on the inactive chromosome slowly become reactivated (Samollow *et al*, 1995). It has also been suggested it may play a role in controlling transcription from multiple promoters, for example, despite methylation of the CGI promoter for *PARP12*, expression was still evident from an alternative start site, downstream of the CGI (Rauch *et al*, 2009).

The function of DNA methylation may be dependent on the genomic region. Although methylation within CGIs is associated with gene repression, within gene bodies it is linked with gene expression (Lister *et al*, 2009). Within CpG poor regions, the majority of CpGs are methylated, but little is known about the significance of this. It has been suggested that methylation in this context provides genomic stability, for example in telomeres (Gonzalo *et al*, 2006), and *DNMT3B* mutation in the severe developmental disorder ICF syndrome, which is associated with hypomethylation of the centromeric regions and frequent cytogenetic alterations (Okano *et al*, 1999).

1.3.3 Techniques for analysis of DNA methylation

There are several techniques available to analyse DNA methylation, summarised in Table 1.4. At present, these are all based on one of three methods, namely, paired restriction enzyme isoschizomers, both of which recognise the same CpG-containing restriction site but with differing methylation sensitivities; antibodies to either 5'-methylcytidine or methyl-binding proteins to specifically pull down methylated DNA; or bisulfite conversion, which is considered the gold standard as it can be used to measure methylation at individual base resolution (Patterson *et al*, 2011). Bisulfite treatment converts unmethylated cytosine to uracil, which is subsequently amplified as thymine, whereas 5' methylcytosine remains unchanged. Hence bisulfite conversion creates single nucleotide polymorphisms (SNPs) that reflect the underlying level of methylation originally present at each CpG site (Frommer *et al*, 1992). All these methods can be used to analyse specific regions of DNA to produce binary, semi-quantitative or fully quantitative results. All three techniques have

also been coupled with array platforms to assess CpG methylation on a "genome-wide" scale, although at present these technologies are not truly genome-wide and at most look at only 1-2% of the CpG sites.

An example of a restriction digestion enzyme technique combined with an array that has been used to analyse AML is HELP, *Hpa*II tiny fragment Enrichment by Ligation-mediated PCR. The restriction enzymes *Hpa*II and *Msp*I both recognise the restriction site 5'-CCGG-3'. *Hpa*II only cleaves if the internal CpG within the restriction site is unmethylated, and *Msp*I will cleave regardless of methylation status. Fragments of differing lengths are thus created by the restriction enzymes depending on the methylation status of the CpG sites, and this can then be analysed on an array (Khulan *et al*, 2006). One disadvantage of this technique is that it is limited to analysing CpG sites that are within the 5'-CCGG-3' recognition sequence.

Antibodies to 5'-methylcytidine or methyl-binding proteins to specifically pull down methylated DNA followed by hybridisation with a tiling array have also been used, termed MeDIP, ChIP-chip or MBDCap (Cross *et al*, 1994; Rauch *et al*, 2006; Weber *et al*, 2005; Yalcin *et al*, 2013). Both MeDIP and MDBCap are subject to bias according to CpG density. MeDIP is based on immunoprecipitation of single-stranded DNA fragments, and as CpG-rich regions often remain double-stranded this favours the pull down of CpG-poor regions. In contrast, MBDCap has increased affinity for CpG-rich regions (Robinson *et al*, 2010). Enrichment-based techniques do not provide information at the single CpG level, and require statistical modification to account for differing CpG densities in different regions. Furthermore, they can be susceptible to measurement errors in cases with copy number variation.

The Illumina Infinium Human Methylation array uses bisulfite-converted DNA that is whole genome amplified and hybridised to an array with 50mer probes targetting specific CpGs to quantify the methylation at those sites. The choice of CpG sites has been selected by a panel of experts, with most CpGs interrogated being in CGIs and shores (Bibikova *et al*, 2011; Bibikova *et al*, 2009; Sandoval *et al*, 2011). Given that this methodology analyses single CpG sites, there is an assumption that differentially-methylated CpGs are

Table 1.4 Main methods of analysing DNA methylation

	Sample treatment		
	Enzyme digestion	Affinity enrichment	Bisulfite conversion
Regional analysis-qualitative results	*Hpa*II-PCR	MeDIP-PCR	COBRA MSP
Regional analysis-quantitative results	MSRE-qPCR		Pyrosequencing Sanger sequencing MALDI-TOF
Array-based analysis	HELP	MeDIP ChIP-chip	Illumina Infinium
NGS-based analysis	HELP-seq	MeDIP-seq	RRBS WGBS

COBRA, combined bisulfite conversion and restriction assay; HELP, *Hpa*II tiny fragment enrichment by ligation-mediated PCR; MALDI-TOF, matrix-assisted laser deionisation time-of flight analysis mass spectrometry; MeDIP, Methylated DNA immunoprecipitation; MSP, methylation specific PCR; MSRE-qPCR, methylation sensitive restriction enzyme and quantitative PCR; NGS, next generation sequencing; qPCR, quantitative PCR; RRBS, reduced representation bisulfite sequencing; seq, sequencing; WGBS, whole genome bisulfite sequencing.

representative of differentially-methylated regions, which are considered biologically the more important (Eckhardt *et al*, 2006). This technology is reliant on efficient bisulfite conversion for accurate quantification, however as a control for the conversion it does contain probes that interrogate non-CpG cytosines, which would be expected to be fully unmethylated and thus fully converted to thymine. Another issue is if the cytosine analysed is also the location of a C/T SNP, the array will not be able to differentiate between a T-containing allele and an unmethylated cytosine that had been converted to thymine by bisulfite conversion. This may be seen in up to 4% of the cytosines analysed (Price *et al*, 2013).

Studies published using these different techniques are difficult to compare directly with one another, as they generally analyse different CpG sites or, in the case of affinity-based methods, methylation levels in regions rather than single CpGs. It should also be noted that none of these approaches can differentiate hydroxymethylcytosine from 5-methylcytosine without additional steps being included in the protocols (Bhattacharyya *et al*, 2013), and hydroxymethylation is increasingly being recognised as biologically important due to the significant levels seen in the bodies of active genes in somatic tissues (Nestor *et al*, 2012). Furthermore, these array-based techniques do not interrogate cytosine methylation in a non-CpG context, which is rare in mammalian cells but does exist, particularly in embryonic stem cells (Ramsahoye *et al*, 2000; Ziller *et al*, 2011). There are also important statistical considerations when analysing the array results from any of the platforms, for example, the methylation scale is finite and not normally distributed; and inter-array normalisation, such as that used for gene expression arrays, cannot be used as the overall amount of DNA methylation is different between samples (Aryee *et al*, 2011).

1.3.4 Histone modification

The DNA of eukaryotic cells is organised into higher order structures that can also have an impact on gene expression. Every 147 bases of DNA are wrapped around an octamer complex consisting of four different histones, H2A, H2B, H3 and H4. These repeating units, known as nucleosomes, are attached to one another by the continuation of the DNA strands and also by H1 histone linkers.

The modification of histones is more complex than DNA, not only can the position of modification vary, but also there are a number of modifications that can occur, including methylation, acetylation, ubiquitination and phosphorylation. All these alterations can lead to changes in chromatin structure, and thereby affect gene expression.

1.3.4.1 Lysine methylation by Polycomb proteins as an example of histone modifications

The Polycomb and trithorax families of proteins were initially discovered in *Drosophila* as, respectively, repressors and activators of *Hox* genes that encode transcription factors specifying cell identity along the anteroposterior axis of segmented animals. Subsequently, they were shown to maintain the expression state of the *Hox* genes through modification of histones, not just in *Drosophila* but also in vertebrates (Schuettengruber *et al*, 2007). Two different complexes of Polycomb proteins are seen in humans. Polycomb repressive complex 2 (PRC2) is comprised of EZH2, SUZ12, EED and RbAp46/48. EZH2 has been shown to catalyse the trimethylation of lysine 27 on histone 3 (H3K27me3), and deletion of the *Drosophila* homologue E(z) leads to absence of H3K27me3 (Cao *et al*, 2002). This trimethylation mark is recognised by the polycomb complex PRC1, which ubiquitinates histone H2A at lysine 119 (H2AK119ub) and leads to transcriptional repression by blocking the action of RNA polymerase (Wang *et al*, 2004). Using murine embryonic stem (ES) cells, Boyer *et al* (2006) showed that PRC1 and PRC2 complexes co-localise to nucleosomes with the H3K27me3 marks, and that this was associated with many genes important in development. Furthermore, three Polycomb target genes with low transcript levels in wild type ES cells had significantly increased transcript levels in ES cells deficient for Eed, a component of the PRC2.

However, this is likely to be an oversimplification of transcriptional control by the polycomb proteins. For example, PRC2 and PRC1 do not always co-localise (Ku *et al*, 2008). Moreover, H3K27me3 is not always associated with repressed genes, and is sometimes found together with active histone marks such as H3K4me3 at so called bivalent domains (Young *et al*, 2011).

1.3.5 Interaction of histone modification and DNA methylation

There is a tight link between histone regulation and DNA methylation, with one being able to impact on the other and vice versa. Confirmation of this connection has been seen in studies of how methylation marks are established in embryonic development. Recent evidence suggests that it may be histone marks that allow CGIs to remain unmethylated during the embryonic stage. In the proposed model, RNA polymerase II, which is located at actively transcribed regions of DNA, recruits histone methyltransferases, which in turn mediate the methylation of lysine 4 on histone 3 (H3K4) at these points (Guenther *et al*, 2007). Usually DNMT3L binds to H3 to recruit DNMT3A and DNMT3B to allow *de novo* methylation of the associated DNA, however in regions where there is H3K4 methylation, DNMT3L is inhibited from binding and thus the underlying DNA remains unmethylated (Ooi *et al*, 2007). Indeed, it has been shown in many tissue types that the presence or absence of H3K4me is associated with the absence or presence respectively of DNA methylation (Meissner *et al*, 2008).

As well as histone marks directing the location of DNA methylation, it has been suggested that DNA methylation, which is maintained by DNMT1, serves as a "memory" for the chromatin structure, allowing it to reform when it has been disrupted, for example after cell replication. Chromatin immunoprecipitation studies have shown that unmethylated DNA tends to assemble in nucleosomes containing acetylated histones, whereas methylated DNA assembles in nucleosomes containing non-acetylated H3 and H4 and adopts a more compact heterochromatin conformation (Eden *et al*, 1998; Hashimshony *et al*, 2003). There is also evidence that either DNA methylation itself, or methylcytosine binding proteins such as MECP2 or MBD2, may direct or inhibit enzymes capable of histone modification. For example, DNA methylation directs H3K9 methylation, a mark of repressive chromatin, through interaction of G9a and DNMT1 (Esteve *et al*, 2006).

Further evidence of this link comes from comparison of the genome-wide distribution of histone methylation patterns and DNA methylation patterns (Meissner *et al*, 2008). For example, in ES cells regions with the repressive H3K27me3 are highly correlated with CGIs that develop DNA methylation as

the cells differentiate (Mohn *et al*, 2008). The relationship between the histone methylation mark and subsequent DNA methylation is considered to be through interacting enzymes, for example, G9a and EZH2, both of which catalyse histone methylation through SET domains, are responsible for recruiting the DNMTs from a separate domain. Thus mutations can occur in the SET domain of these enzymes that will affect the histone modification but not DNA methylation (Dong *et al*, 2008).

1.3.6 Epigenetic changes in cancer

When compared to the normal counterpart tissue, the cancer epigenome shows both global changes in chromatin structure and DNA methylation as well as local changes in specific regions. For example, the overall 5-methylcytosine content decreases in malignancy whereas some CGIs become aberrantly hypermethylated (Feinberg & Vogelstein, 1983; Gonzalez-Zulueta *et al*, 1995; Greger *et al*, 1989; Herman *et al*, 1994). Similarly, loss of H4 acetylation at lysine 16 and trimethylation at lysine 20 are also common hallmarks of malignancy (Fraga *et al*, 2005). Less is known about histone changes in malignancy due to the current limitations in analysing these changes on a genome-wide scale.

1.3.6.1 The cancer methylome

The lower level of DNA methylation seen in malignant cells compared to their normal counterpart cell is due to loss of methyl groups in repetitive DNA sequences and demethylation of coding regions and introns. In mouse models with progressive tumours, as the neoplasm progresses from a benign proliferation to an invasive cancer, the degree of hypomethylation increases (Fraga *et al*, 2004). It has been proposed that this generalised hypomethylation contributes to tumour development either by increasing chromosomal instability, which increases the risk of deletion or translocation of chromosomes (Eden *et al*, 2003), or by reactivation of transposable elements that had been silenced by methylation, or through loss of imprinting.

In contrast, since the initial discovery of hypermethylation of the Rb promoter in retinoblastoma (Greger *et al*, 1989; Sakai *et al*, 1991), tumour-specific hypermethylated CGIs associated with inactive tumour suppressor genes have

been identified in several tumour types (Costello *et al*, 2000). However, when the methylome of a colorectal tumour was analysed, the regions showing the most differential methylation compared to normal colorectal tissue were located in the CpG shores, i.e. overlapping with the tissue-specific differentially methylated regions (Irizarry *et al*, 2009).

Aberrant CGI hypermethylation in tumours was therefore proposed as an alternative to inactivating mutations of tumour suppressor genes. The sites of the aberrant hypermethylation often correspond to genes that are known to be mutated in that type of tumour. For example, the colorectal carcinoma cell line HCT116 has a heterozygous mutation in $p16^{INK4A}$, the cyclin-dependent kinase inhibitor, and hypermethylation of the wild type allele, so that only the mutant allele is expressed. Transcription of the wild type allele can be restored by treatment with the demethylating agent 5-aza-2'-deoxycytidine (Myohanen *et al*, 1998). Similarly, *CEBPA*, a gene that is commonly mutated in AML, has been shown to be hypermethylated in a subset of AML patients, although they had wild type *CEBPA* (Jost *et al*, 2009).

However, studies of a prostate cancer cell line, PC3, found that many of the aberrantly hypermethylated genes were already silenced in the corresponding normal prostate epithelial cells (Gal-Yam *et al*, 2008). Comparison of gene expression and methylation data has now demonstrated this association in several other different cancer types, including AML (Sproul *et al*, 2012). This may explain the lack of correlation reported in many studies between gene expression and the corresponding methylome (Pike *et al*, 2008), suggesting that rather than aberrant DNA methylation being a result of selection pressure, it may be determined by a pre-programmed targeting mechanism. For example, some de novo methylated CGIs identified in embryonal carcinoma, embryonal stem cells and the colon tumour cell line, Caco-2, correspond to target sites for Polycomb protein binding (Ohm *et al*, 2007; Schlesinger *et al*, 2007).

1.3.6.2 Interaction of epigenetic and genetic factors in cancer

There are marked changes both in the genetic and epigenetic milieu in malignancy, but how they interact and influence one another is still being elucidated. Methylated cytosines are inherently prone to deamination to become

thymine, and this process has also been associated with many recurrent mutations seen in cancer, for example p53 mutations (Rideout *et al*, 1990). However, although epigenetic modifiers are enzymes that are either involved in creating the epigenetic code, e.g. methyltransferases, or proteins that interpret the epigenetic code to affect further change, e.g. methyl binding proteins, and are encoded by genes that are recurrently mutated in several malignancies, the mutations do not necessarily lead to the expected epigenetic changes. For example, mutations in *DNMT3A* found in AML may be expected to lead to very marked changes in the methylation status of malignant cells, but one study showed no difference in global methylation levels in AML samples with and without *DNMT3A* mutations, and another only showed a small, albeit significant, reduction in methylation in samples with the most common *DNMT3A* mutation, R882 (Ley *et al*, 2010; Russler-Germain *et al*, 2014). However, both studies demonstrated specific regions with differential methylation between AML samples with and without *DNMT3A* mutations. Hence, inactivating mutations in *DNMT3A* do not disturb the methylome as much as would be expected given its role in methylation maintenance.

Thus, despite malignancies displaying aberrant DNA methylation and histone modifications characteristic of the type and subtype of cancer (Costello *et al*, 2000; Esteller *et al*, 2001), the role of these epigenetic changes remains unclear. Nevertheless, an improved understanding of these changes is important and relevant to clinical practice as their potential reversibility makes them an attractive therapeutic target.

1.4 Aims of this thesis

The preliminary aim of the studies reported in this thesis was to investigate whether genome-wide alterations in the epigenetic make-up of patients with AML could provide additional information in predicting response to treatment, and whether this could be used to refine the current prognostic classification. Samples from two highly selected cohorts of patients with chemosensitive and chemoresistant disease were first interrogated to assess whether there were differences in methylation between the two groups (chapter 3). The results of

this array are explored further with particular reference to the underlying *CEBPA* genotype. The different types of *CEBPA* mutations and their methylation patterns are examined in chapter 4, and data is presented looking specifically at methylation of the *CEBPA* promoter region. In addition, patient samples were screened for mutations in the epigenetic modifier *ASXL1* and the prognostic impact determined in cohorts stratified according to age and disease status (Chapter 5).

CHAPTER 2:**MATERIALS AND METHODS**

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)

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)

GoTaq® DNA polymerase (Promega, Madison, USA)

GoTaq® Colourless Flexi reaction buffer and magnesium chloride (Promega, Madison, USA)

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

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

Optimase® DNA Polymerase (Transgenomic Ltd, Glasgow, UK)

Optimase® Buffer and Magnesium Chloride (Transgenomic Ltd, Glasgow, UK)

peqGOLD MicroSpin Cycle-Pure PCR Purification Kit (Peqlab, Salisbury, 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)

QIAquick PCR Purification kit (QIAGEN, Crawley, UK)

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

Super optimal broth with catabolite repression (S.O.C.) (Invitrogen Life Technologies, Paisley, 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)

PCR was used to amplify specific regions of genomic DNA. It required addition of short oligonucleotide primers complementary to either end of the region of interest, to polymerase enzyme, nucleotides and appropriate buffers. A mastermix of this was created and an aliquot added to each DNA sample to be analysed. These mixtures were then subjected to successive cycles of 3 different temperatures to allow denaturation of the template DNA, annealing of the primers, and extension of newly formed DNA strands respectively. A negative control (water added instead of

DNA) was included with each PCR, to ensure there was no contamination of the master mix. Four different DNA polymerase enzymes were used in this work.

BIOTAQ™ DNA polymerase (Bioline, London, UK) was used when experiments did not require a hot start or proof reading polymerase, or if the PCR product was to be cloned. A standard master mix comprised 1x manufacturer's buffer, 1.5mM $MgCl_2$, 200µM each dNTP, 0.5 µM each primer, and 0.5 units of polymerase, to which 30ng of DNA was added, total volume 20µl. The mixture was then subjected to 35 cycles of denaturation at 95°C for 30 seconds, an annealing step at a temperature appropriate to the primers for 30 seconds, and an extension at 72°C for 30 seconds, followed by a final extension of 72°C for 5 minutes. Conditions for the other three DNA polymerases are given in the appropriate sections. GoTaq® Hot Start polymerase (Promega, Wisconsin, USA) was used for amplification of bisulfite-converted DNA, as it has improved specificity compared to non-hotstart enzymes. Optimase® polymerase (Transgenomic Ltd., Glasgow, UK) was used for PCR of samples to be analysed on the WAVE platform, as this enzyme has 3'-5' exonuclease proofreading capabilities necessary for high-fidelity amplification. If samples were not successfully amplified with BIOTAQ™ or Optimase®, then Phusion® Hot Start High Fidelity DNA polymerase (New England Biolabs, Hitchin, UK) was used as an alternative.

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 detect and assess the quality of the PCR products and to ensure the negative control was as expected. It was also used to detect fragments of different sizes if a digest had been performed. Agarose (1-2g according to the percentage of gel required)

was dissolved in 50ml of 1xTBE by heating. Once the solution had cooled slightly, 5µl of 1mg/ml ethidium bromide was added and the gel poured into the mould, with combs inserted to create wells. Once set, the gel was covered in 1xTBE running buffer containing ethidium bromide (0.01µg/ml) and the combs removed. An aliquot (5-10µl) of product was mixed with loading buffer and pipetted into each well. Samples were electrophoresed at a current of approximately 70mA for 5-30 minutes. Fragments were detected by UV illumination and recorded in a digital photograph.

2.1.4 Screening for mutations using dHPLC

Denaturing high performance liquid chromatography (dHPLC) using the WAVE® platform is a fast, sensitive method of screening a large number of samples for nucleotide substitutions, insertions or deletions. It requires the presence of heteroduplexes and relies on the different chemical and physical properties of heteroduplexes and homoduplexes. DNA template (30ng) was amplified using Optimase® DNA polymerase in an Optimase® master mix containing 1x manufacturer's buffer, 1.5mM $MgSO_4$, 200µM each dNTP, 0.5 µM each oligonucleotide primer, 0.5U of Optimase® polymerase and sterile ddH_2O to make up to 20µl. The mixture was subjected to an initial denaturation step of 95°C for 5 minutes followed by 35 cycles of denaturation at 95°C for 45 seconds, a 45 second annealing step at a temperature that was dependent on the primers and a 45 second extension step at 72°C, and then a final extension step of 72°C for 15 minutes. If no products were produced with Optimase®, a PCR using Phusion Hot Start High Fidelity DNA polymerase (0.2U per reaction) was performed using a master mix of 1x manufacturer's HF buffer, 0.2U polymerase and the same concentration of primers, dNTPs and magnesium as for the Optimase PCR. Cycling conditions in this case were an initial denaturation step of 98°C for 2 minutes followed by 35 cycles of denaturation at 98°C for 42 seconds, annealing at the appropriate temperature for 42 seconds and extension at 72°C for 42 seconds and then a final extension step at 72°C for 15 minutes.

PCR products to be analysed were mixed with known wild-type PCR product in a ratio of approximately 4:1, as estimated visually from the agarose gel. The PCR mixes were denatured by incubating at 95°C for 5 minutes and then slowly cooled using 40 cycles, each of one minute, starting at 92°C and gradually decreasing the temperature by 1.5°C for each subsequent cycle, to allow random reannealing of PCR strands. Thus if a mutation was present, even if it was homozygous, the mix would now contain both heteroduplexes and homoduplexes. The samples were individually injected through a polystyrene-divinyl benzene copolymer DNA separation column located in an oven at the temperature required for analysis. This was calculated by the Navigator™ software and was dependent on the sequence and length of the PCR product. Temperatures were chosen so that the sequence of interest was 50-95% helical. Where necessary, the same sample was analysed at two or three different temperatures to account for sequence domains of differing helicity across the amplicon. TEAA acts as an ion-pairing agent and allowed the PCR products to bind to the column. Over the course of a run, the concentration of ACN in the buffer flowing through the column was increased and the resulting change in pH decreased the effectiveness of the TEAA as an ion-pairing agent. Heteroduplexes, due to the mismatch in paired DNA strands, bound less strongly to the column and thus were eluted from the column before homoduplexes. The products released were detected by a UV detector measuring the absorbance at 260nm and this was recorded by the software in real time by a peak on the chromatogram. For a mixture of heteroduplexes and homoduplexes, this would be seen by a minimum of two peaks, whereas a wildtype sample would only have a single homoduplex peak.

2.1.5 Restriction Enzyme Digestion

Digestion of PCR products using restriction enzymes was performed according to manufacturer's instructions. PCR product (4-8µl) was added to 1x appropriate buffer, 10U restriction enzyme, 100pg bovine serum albumin if required and dd$H_2$0 in a total volume of 10µl. This was then incubated at 37°C or 60°C for 2-16 hours depending on the enzyme used.

The products were either analysed by agarose gel electrophoresis with the appropriate DNA ladder (Bioline, London, UK) to estimate fragment size, or on the Beckman Coulter CEQ™ 8000 DNA Genetic Analysis System (Beckman Coulter UK Ltd., Buckinghamshire, UK) if quantification of the fragments was required.

2.1.6 DNA sequencing

PCR products were purified using either QIAquick PCR Purification kit (QIAGEN, Crawley, UK) or peqGold microspin cycle pure kit (Peqlab, Southampton, UK) according to the manufacturer's instructions and then sent to the UCL Cancer Institute Scientific Support Service for direct nucleotide sequencing.

2.1.7 Fragment analysis of PCR and restriction enzyme digestion products

Fragment analysis was used to either detect products of varying lengths that were not adequately separated on an agarose gel, or to quantify the relative amount of multiple products generated by PCR or following restriction enzyme digestion. A BIOTAQ™ PCR was performed using one fluorescently labelled primer and one unlabelled primer. Standard PCR conditions were used (see section 1.1.2) but with only 5pmols of primers and fewer cycles in order to prevent saturation of the fluorescence detector and reduce heteroduplex formation. If required, PCR products were then subject to restriction enzyme digestion.

PCR product or digest (2µl) was added to 37.75µl formamide sample loading solution and 0.25µl DNA Size Standard 600 (Beckman Coulter, High Wycombe, UK) and fragments size-separated by capillary electrophoresis on the Beckman Coulter CEQ™ 8000 DNA Genetic Analysis System (Beckman Coulter UK Ltd., High Wycombe, UK). Fragment size was estimated by the instruments' software based on elution time with reference to the size standards, and the relative amount of each product calculated using the area under the peak as a proportion of the total area under all peaks.

2.1.8 LB Broth and agar plates

LB broth: 6 LB capsules in 250ml ddH_2O

LB agar plates: 12 LB capsules and 7.5g agar in 500ml ddH_2O

The broth and agar mixtures were autoclaved and allowed to cool to 50°C before adding carbenicillin to a final concentration of 100µg/ml. Plates were then poured and left to set. Prior to use, plates were spread with 40µl of 40mg/ml X-gal in dimethylformamide and incubated at 37°C for a minimum of 30 minutes.

2.1.9 TOPO® TA cloning of PCR products

The TOPO® TA Cloning Kit with One Shot® MAX Efficiency® DH5α-T1R *E.coli* (Invitrogen Life Technologies, Paisley, UK) was used for cloning. Fresh PCR products were prepared using the non-proof reading polymerase BIOTAQ as this adds a single deoxyadenosine to the 3' end of the PCR product, and 1-2µl incubated at room temperature for 5 minutes with 1µl linearised pCR 2.1-TOPO vector and 1µl salt solution in a total volume of 6µl. The vector can ligate the PCR product as it has a single 3' thymidine overhang, and has topoisomerase I covalently bound to catalyse the reaction. Ligation reaction (1µl) was added to one vial of thawed One Shot® Max Efficiency® DH5α- T1*E.coli* cells and incubated on ice for 30 minutes. The cells were then heat-shocked at 42°C for 30 seconds; 250µl of SOC medium was added and the mix incubated in a shaking incubator at 37°C for 1 hour. An aliquot of 80µl of the transformed bacteria was mixed with 20µl SOC medium, spread on an LB agar plate (see section 2.1.8) and then incubated at 37°C overnight to allow colony formation.

The pCR 2.1-TOPO vector contains a *LacZα* gene which can hydrolyse X-gal leading to the formation of blue bacterial colonies. However, if a PCR product has been inserted into the vector, this disrupts the gene leading to white bacterial colonies. White colonies were therefore plucked, seeded into 200µl LB broth containing carbenicillin in a 96 well plate, and incubated at 37°C overnight. The individual bacterial cultures were then used as template for PCR amplification using standard conditions but with

an initial step of 95°C for 2 minutes to lyse the cells and inactivate nucleases.

2.2 Analysis of DNA methylation

Bisulfite conversion was the chosen method to detect DNA methylation as this can be used for quantification at the single CpG level (see chapter 1, section 1.3.3).

2.2.1 Reagents

EZ DNA Methylation-Gold Kit™ (Zymo Research, California, USA)

PyroMark Binding Buffer

PyroMark Wash Buffer

Pyromark Annealing Buffer

Sodium Hydroxide Solution

Ethanol

Streptavidin Beads

PyroMark Gold Q96 Reagents

2.2.2 Bisulfite conversion of DNA

Sodium bisulfite treatment of DNA leads to deamination of non-methylated cytosines to form uracil, which is amplified as thymine by downstream techniques such as PCR (Frommer *et al*, 1992; Hayatsu *et al*, 1970), whereas methylated cytosines are “protected” from deamination and thus remain as cytosines (Figure 2.1). The proportion of cytosines compared to the total number of cytosines and thymines at a particular site can therefore provide a surrogate marker for the level of methylation.

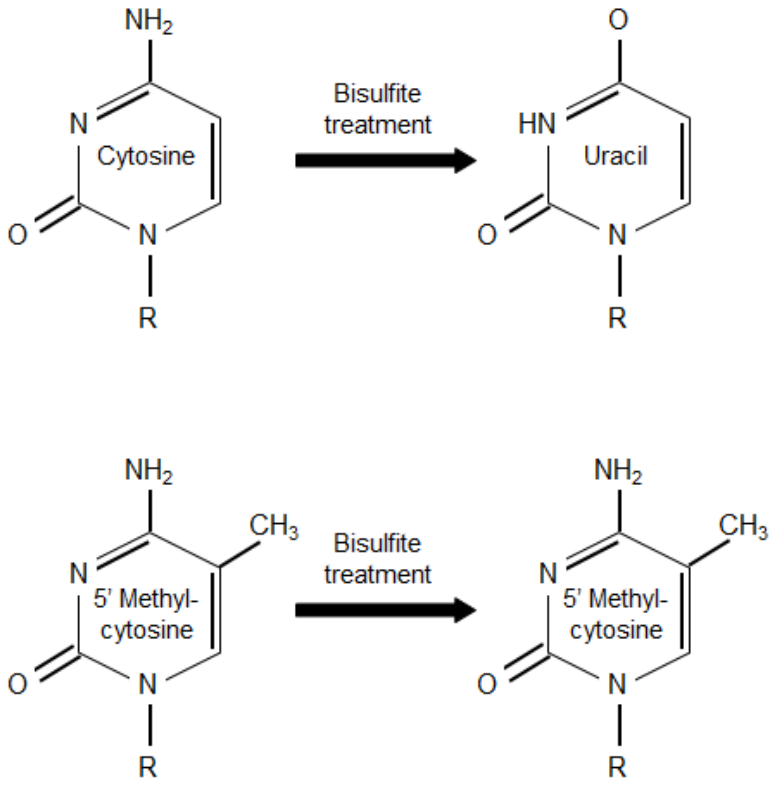


Figure 2.1 Chemical effect of bisulfite treatment on cytosine and 5' methylcytosine.

Using the EZ DNA methylation-gold kit, 350-500ng DNA in 20µl water was added to 130µl CT conversion agent, mixed and incubated at 50°C for 16 hours, with a denaturation step of 95°C for 30 seconds at the beginning of every hour. The periodic cycling has been suggested as an alternative incubation to the manufacturer's recommendations as it improves conversion efficiency (Thirlwell *et al*, 2010). After conversion, the samples were purified according to manufacturer's instructions using the buffers and Zymo-Spin™ IC columns provided in the kit. The bisulfite-converted DNA was eluted into 10-20µl elution buffer depending on the type of downstream analysis planned.

2.2.3 Methylation-Specific PCR to check conversion efficiency

To check for efficiency of conversion, several samples were randomly selected from each converted batch and subjected to methylation-specific

PCR by performing two PCRs of the *HLA-B* gene for each sample, one using primers that would only amplify bisulfite-converted DNA and the other using primers that would only amplify unconverted DNA (Appendix 1). For each PCR, 1µl converted DNA was added to 1x manufacturer's buffer, 25pmols each primer, 200µM dNTPs, 1U GoTaq DNA polymerase (Promega, USA) and 3mM $MgCl_2$ in 25µl total volume. An initial denaturation step of 95°C for 6 minutes was followed by 36 cycles of PCR, each of 94°C for 30 seconds, annealing temperature for 30 seconds and 72°C for 90 seconds, with 2 cycles at annealing temperatures of 60^{o}C, 59^{o}C and 58^{o}C, then 30 cycles at 57°C, followed by a final extension step of 72°C for 15 minutes. PCR products were then run on a 2% agarose gel. Samples were considered to be successfully bisulfite-converted if they had a PCR product with the primers for bisulfite-converted DNA but no product with the primers for unconverted DNA.

2.2.4 Pyrosequencing

Pyrosequencing is a method of sequencing short regions of DNA. Single-stranded PCR product acts as a template to which nucleic acids are added sequentially in a pre-defined order to create a complementary DNA strand. If the nucleotide added is incorporated into the newly forming DNA strand, it releases pyrophosphate, which sets off a cascade of enzymatic reactions culminating in the release of light proportional to the number of nucleotides added.

Pyrosequencing assays to interrogate specific CpG sites were designed using the PyroMark Assay Design Software 2.0 (Qiagen, Germany). PCRs were performed using a master mix of 1.25U GoTaq DNA polymerase, one biotin-labelled and one unlabelled primer each at 0.2 µM, 1x manufacturer's buffer, 3.5mM $MgCl_2$ and 200µM of each dNTP, to which 25ng DNA was added and the final volume made to 20µl by the addition of ddH_2O. The mixtures were denatured at 95°C for 5 minutes then subjected to 50 cycles of denaturation at 95°C for 30 seconds, annealing at the appropriate temperature for 30 seconds and an extension step of 72°C for 45 seconds, then a final extension step at 72°C for 15 minutes. Fifty cycles

of amplification were necessary to ensure that all biotinylated primers were incorporated into PCR products. The presence of PCR products and absence of primer bands was checked by agarose gel electrophoresis, then 5-15µl aliquots were mixed with streptavidin beads to bind the biotin-labelled products. Using the PyroMark Q96 Vacuum Prep Tool, the beads were captured and treated with ethanol, then sodium hydroxide to denature the DNA and leave only the bound single-stranded biotin-labelled PCR product, and finally washed. The beads were dispensed into a pyrosequencing plate containing the appropriate sequencing primer, heated to 85°C, cooled to allow binding of the primer to the template and analysed on the PyroMark Q96 MD Pyrosequencer (Qiagen, Germany) using the Pyromark Gold Q96 enzyme, substrate mixes and nucleotides.

Results of the analysis were displayed as a pyrogram, which shows the order in which the nucleotides were added into the mixture and the amount of light released for each nucleotide added, with the peak height for each nucleotide thus proportional to the number of nucleotides incorporated.

Each pyrogram was assessed by the software for three parameters. At least five non-CpG nucleotides were quantified to ensure that the peak heights were as expected for the number of nucleotides incorporated; a background (negative) control was checked for the absence of nucleotide incorporation; and quantification of a non-CpG C (or G if the complementary strand was being analysed) was checked for complete conversion to T (or A). If these parameters were satisfactory, the proportion of methylated alleles at the CpG site of interest was calculated. The C and T (or G and A) peak heights at the specified CpG site were summed to ensure that they were equivalent to the expected number of nucleotides incorporated, and the proportion of methylated alleles was calculated by the Pyro-Q-CpG software based on the relative peak heights of the pyrogram for the C and T (or G and A) nucleotides at that site.

For each assay, four negative controls were analysed to ensure that the template or primers did not give any background signal. These were respectively adding (i) sequencing primer alone without PCR product, (ii)

biotin-labelled primer alone, (iii) sequencing and biotin-labelled primers, and (iv) PCR product alone. In addition, for each PCR run, the negative water control from the PCR was included to ensure there was no contamination. All samples were analysed in duplicate using fresh PCRs and run on separate occasions. The mean methylation level was calculated and expressed as a percentage of total alleles.

Prior to assessing patient samples, the accuracy, precision and sensitivity of the methylation quantification for each assay was analysed using DNA mixtures with varying proportions of fully unmethylated and fully methylated DNA. These mixes were then bisulfite-converted and assayed. Initially, whole genome amplified DNA was used as the fully unmethylated control (provided by Dr Andy Feber, UCL Cancer Institute, UK), and DNA treated with the methyltransferase enzyme SssI (New England Biolabs, USA) as the fully methylated control DNA. However this did not always provide accurate mixes for the regions of interest, and therefore the Epitect Control DNA set (Qiagen, Crawley, UK) containing bisulfite-converted fully methylated and unmethylated DNA was used to test some assays.

CHAPTER 3: ASSOCIATION OF *CEBPA* GENOTYPE WITH METHYLATION PATTERNS IN AML

3.1 Introduction

There has been an increasing number of recurrent mutations identified in AML, some of which have been shown to have prognostic impact (as discussed in Chapter 1). A two-hit model proposed by Kelly and Gilliland (2002) suggested that for AML to occur, two genetic hits were required, one that led to uncontrolled proliferation and the other a block in differentiation. One of the first genes identified to lead to a block in differentiation when mutated was *CEBPA*.

3.1.1 *CEBPA* structure and function

CCAAT/enhancer-binding protein alpha, C/EBPα, is coded for by the intronless *CEBPA* gene located on chromosome 19q. It is a member of the C/EBP family of basic leucine zipper transcription factors. The full length 42kDa protein has 358 amino acids with several functional regions including two transactivation domains, a DNA binding domain (DBD) and leucine zipper domain (LZD) (Figure 3.1). C/EBPα dimerises with either another C/EBPα molecule or a different member of the C/EBP family, mediated through an α-helix within the LZD, to form a coiled-coil structure which is necessary for its function. The DBD recognises CCAAT motifs in the promoters of downstream target genes. An internal ATG start site exists in the *CEBPA* mRNA and translation from this leads to a truncated p30 isoform that lacks the initial transactivation domain. It has been shown that the p30 protein has reduced transactivation potential when compared to the p42 protein (Pabst *et al*, 2001b). C/EBPα is widely expressed in normal tissue, and at high levels in terminally differentiated liver and adipose cells. Within myeloid cells, the expression of C/EBPα is relatively high in early myeloid progenitors and decreases during granulocytic differentiation (Scott *et al*, 1992). The p30 protein is also expressed in liver and adipose cells and the ratio of p42 to p30 isoform expression is tightly regulated by upstream signalling pathways (Calkhoven *et al*, 2000).

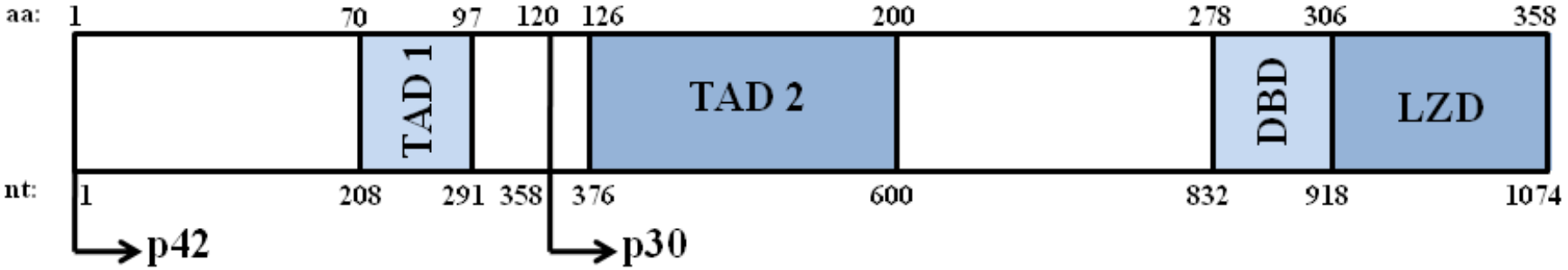


Figure 3.1 Structure and functional domains of C/EBPα There are two transactivation domains (TAD1 and TAD2), a DNA binding domain (DBD) and a leucine zipper domain (LZD). The translational start sites for the p42 and p30 proteins are shown. Amino acid (aa) and nucleotide (nt) numbering are indicated above and below the diagram. Adapted from Mueller and Pabst (2006).

C/EBPα plays an important role in both hepatocyte and adipocyte development (Ramji & Foka, 2002). Mice that are homozygous for deletion of the entire C/ebpα-coding sequence die in the immediate post-natal period due to severe hypoglycaemia, caused by an inability of the mice to store glycogen in the liver (Wang *et al*, 1995). C/ebpα-null mice also lack white adipose tissue. Within the myeloid lineage, absence of mature neutrophils and eosinophils has been observed in null mice, with a corresponding increase in myeloid blasts (Zhang *et al*, 1997). Further work using conditional knock-out mice demonstrated that disruption of C/ebpα blocks the transition of common myeloid progenitors to granulocyte/ monocyte progenitors, but not further differentiation (Zhang *et al*, 2004).

As a transcription factor, C/EBPα binds to DNA at the CCAAT sequence found in the promoters of several genes including those encoding growth factor receptors, for example, the receptor for granulocyte colony-stimulating factor, and secondary granule proteins such as lactoferrin (Koschmieder *et al*, 2009). However, C/EBPα has also been shown to have inhibitory functions which prevent cell proliferation, for example, by binding directly to other proteins such as the cdk2 and cdk4 enzymes, which are cyclin-dependent kinases important in cell cycle regulation (Wang *et al*, 2001), or by blocking the action of the E2F complex and thereby inhibiting cell cycling (Porse *et al*, 2001; Slomiany *et al*, 2000). C/EBPα can also negatively regulate *c-Myc* through this mechanism, allowing early myeloid precursors to enter the differentiation pathway (Johansen *et al*, 2001).

3.1.2 *CEBPA* mutations in AML

Due to the importance of C/EBPα in granulocyte development and the absence of mature granulocytes in *Cebpa*-null mice, Pabst *et al* (2001b) screened the *CEBPA* gene in samples from 137 patients with AML and found that it was mutated in 7% of cases. Since this initial study, several groups have also shown that *CEBPA* is recurrently mutated in AML, with an overall mutation rate of 10% (Table 3.1). The location and type of mutations are non-random, with the majority of mutations falling into one of two categories. Most of the mutations at

Table 3.1 Published studies investigating incidence and impact of *CEBPA* mutations in patients with AML.

Reference	Total no. of patients	*CEBPA*^{MUT} (%)	*CEBPA*^{SM} (% of *CEBPA*^{MUT})	*CEBPA*^{DM} (% of *CEBPA*^{MUT})	Prognostic impact of mutant compared to wild-type *CEBPA*	Notes
Pabst *et al* (2001b)	137	9 (7)	8 (89)	1 (10)	n/a	
Gombart *et al* (2002)	78	6 (8)	5 (83)	1 (17)	n/a	
Preudhomme *et al* (2002)*	135	15 (11)	7 (47)	8 (53)	OS improved in univariate and multivariate analysis	
Barjesteh van Waalwijk van Doorn-Khosrovani *et al* (2003)	277	12 (4)	0 (0)	12 (100)	OS improved in univariate and multivariate analysis	N terminus only examined if C terminus had mutation present
Snaddon *et al* (2003)	99	8 (8)	6 (75)	2 (25)	No difference	
Frohling *et al* (2004)*	236	33 (14)	18 (55)	15 (45)	OS increased in univariate and multivariate analysis	
Bienz *et al* (2005)*	67	12 (18)	5 (42)	7 (58)	OS increased in univariate and multivariate analysis	50% of *CEBPA*^{MUT} had aberrant CD7 expression
Lin *et al* (2005)	104	16 (15)	2 (12)	14 (88)	Trend for longer CR (19 months vs 9 P=0.2)	Higher levels of CD7, CD15, CD34, and HLA-DR expression in *CEBPA*^{MUT}
Frohling *et al* (2005)*	166	17 (10)	2 (12)	15 (88)		Included 41 del(9q) AML samples
Shih *et al* (2006)	149	22 (15)	2 (9)	20 (91)		
Fuchs *et al* (2008)	152	14 (9)	11 (79)	3 (21)		
Juhl-Christensen *et al* (2008)	485	20 (4)	15 (75)	5 (25)		
Benthaus *et al* (2008)	469	38 (8)	18 (47)	20 (53)		CN-AML
Marcucci *et al* (2008)	175	32 (18)	18 (56)	14 (44)	OS increased in univariate and multivariate analysis	CN-AML
Wouters *et al* (2009)*	598	41 (7)	13 (32)	28 (68)	OS increased in univariate and multivariate analysis for *CEBPA*^{DM} only	*CEBPA*^{DM} had distinct gene expression profile
Renneville *et al* (2009)	638	53 (8)	29 (55)	24 (45)	*CEBPA*^{DM} had trend towards better survival over *CEBPA*^{SM}. *CEBPA*^{MUT} better OS if FLT3-ITD neg	

Table 3.1 Continued

Reference	Total no. of patients	*CEBPA*^{MUT} (%)	*CEBPA*^{SM} (% of *CEBPA*^{MUT})	*CEBPA*^{DM} (% of *CEBPA*^{MUT})	Prognostic impact of mutant compared to wild-type *CEBPA*	Notes
Pabst *et al* (2009)	224	19 (8)	7 (37)	12 (63)	OS increased in univariate and multivariate analysis for *CEBPA*^{DM} only	
Hou *et al* (2009)	543	71 (13)	24 (34)	47 (66)	OS increased in univariate and multivariate analysis for *CEBPA*^{DM} only	
Dufour *et al* (2010)*	467	38 (8)	18 (47)	20 (53)	OS increased in univariate and multivariate analysis for *CEBPA*^{DM} only	CN-AML
Green *et al* (2010b)	1427	107 (7)	48 (45)	59 (55)	OS increased in univariate and multivariate analysis for *CEBPA*^{DM} only	
Taskesen *et al* (2011)	1182	151 (13)	60 (40)	91 (60)	OS increased in univariate analysis for *CEBPA*^{MUT} and multivariate analysis for *CEBPA*^{DM} only	CN-AML
Dufour *et al* (2012)	663	59 (9)	28 (47)	31 (53)	*CEBPA*^{SM} associated with a good prognosis in *NPM1*^{MUT} subgroup	
Fasan et al (2014)	2296	244 (11)	140 (57)	104 (43)	*CEBPA*^{DM} associated with good prognosis in univariate and multivariate analysis	
TOTAL	**9098**	**881 (10)**	**420 (48)**	**461 (52)**		

*****These studies are excluded from total numbers as subsequent studies included the same patients

Abbreviations, CN-AML, Cytogenetically normal AML; CR, complete remission; OS overall survival.

the N terminus are nonsense or frameshift mutations, hereafter called “classic N mutations”, and are predicted to lead to absence of the full-length p42 C/EBPα protein but, as they occur upstream of the internal start site, retain p30 isoform expression. At the C terminus, in-frame insertions or deletions in the DBD or LZD predominate, which are presumed to disrupt DNA binding or C/EBPα dimerisation, hereafter termed “classic C mutations”. Approximately 50% of the patients with *CEBPA* mutations have two mutations, *CEBPA* double mutant (*CEBPA*^{DM}) (Green *et al*, 2010b). In most cases this consists of a classic N mutation on one allele and a classic C mutation on the other allele, i.e. biallelic, hereafter named a classic double mutation, which does not produce any fully functional C/EBPα but retains p30 expression.

When considering the impact of *CEBPA* status on outcome, initially all mutated *CEBPA* (*CEBPA*^{MUT}) cases were considered together, and they were shown to have a more favourable outcome than *CEBPA*^{WT} cases (Table 3.1). However, more recent studies have indicated that the favourable outcome is restricted to *CEBPA*^{DM} and in most studies the outcome for patients with a single *CEBPA* mutation, *CEBPA*^{SM}, is more similar to patients who are *CEBPA*^{WT} (Dufour *et al*, 2010; Green *et al*, 2010b; Hou *et al*, 2009; Wouters *et al*, 2009). These findings have led to suggestions that patients with AML who have *CEBPA*^{DM} at diagnosis should be considered in the favourable prognostic category, and therefore should not be recommended for a stem cell transplant in first remission as the risks would outweigh the benefits (Cornelissen *et al*, 2012). Furthermore, studies which have analysed *CEBPA*^{DM} separately from *CEBPA*^{SM} have confirmed that *CEBPA*^{DM} have other characteristics that are distinct from both *CEBPA*^{WT} and *CEBPA*^{SM}. For example, the rate of co-incidence of other mutations differs as both *NPM1* and *FLT3*/ITD mutations are inversely correlated with *CEBPA*^{DM} but not *CEBPA*^{SM} (Green *et al*, 2010b), and *CEBPA*^{DM} samples have a distinct gene expression profile (Dufour *et al*, 2010; van Vliet *et al*, 2013; Wouters *et al*, 2009) and DNA methylation profile (Figueroa *et al*, 2010b).

3.1.3 Types of *CEBPA* mutations

Although the majority of *CEBPA*^{DM} have a frameshift mutation in the N terminus of one allele and an in-frame insertion or deletion in the C terminus of the other allele, approximately one quarter of patients with *CEBPA*^{DM} have either non-classical mutations or a non-classical combination of mutations (Table 3.2). All *CEBPA*^{DM} are included in published studies that look at outcome and thus are considered as favourable prognostic alterations, however there is no definitive evidence that this is the case for non-classical *CEBPA*^{DM}. The non-classical mutations are too few in number and also too varied to truly assess their impact on outcome compared to other AML samples.

In vitro functional studies of *CEBPA* mutations have included transactivation (TA) assays, which assess the ability of C/EBPα to bind and activate target promoters by coupling the promoters to luciferase assays. Insertion of a single classic N or classic C mutant *CEBPA* construct led to a reduction in TA activity compared to wild-type (Gombart *et al*, 2002; Kato *et al*, 2011; Pabst *et al*, 2001b). Furthermore, when both a classic N and classic C construct were transfected simultaneously, to mimic a classic double mutant case, this led to a marked reduction of TA activity to the equivalent of vector alone (Pabst *et al*, 2009). However, only one of these studies investigated three non-classical mutations. Pabst *et al* (2001b) showed that a missense mutation in the N terminus led to the same TA activity as the wild-type construct, a frameshift in the C terminus (V351fs) caused a decrease in TA activity, whereas a frameshift in the middle of the gene (R165fs) resulted in an increase in TA activity when compared to wild-type.

Several groups have explored the effects of classical *CEBPA* mutations in mouse models. Bereshchenko *et al* (2009) competitively transplanted fetal liver cells from knock-in mice with wild-type, homozygous classic N, homozygous classic C or classic double compound heterozygous mutations into sublethally irradiated mice together with wild-type competitor bone marrow cells. All mice that were transplanted with mutant *cebpa* developed leukaemia. Kato *et al* (2011) transduced murine bone marrow mononuclear cells with retroviral constructs expressing either a classic N mutation, a classic C mutation or both, then transplanted the cells into irradiated syngeneic mice. In this model, all mice

Table 3.2 Published studies that list type of *CEBPA*^{DM} mutations in AML

Reference	Total no. of patients	No. of *CEBPA*^{DM} (%)	No. of classic *CEBPA*^{DM} (% of all *CEBPA*^{DM})	No. of homozygous classic *CEBPA*^{DM} (% of all *CEBPA*^{DM})	No. of other atypical *CEBPA*^{DM} (% of all *CEBPA*^{DM})
Pabst *et al* (2001b)	137	1 (1)	1 (100)	0 (0)	0 (0)
Gombart *et al* (2002)	78	1 (1)	1 (100)	0 (0)	0 (0)
Preudhomme *et al* (2002)	135	8 (6)	5 (62)	1 (13)	2 (25)
Barjesteh van Waalwijk van Doorn-Khosrovani *et al* (2003)	277	12 (4)	11 (92)	1 (8)	0 (0)
Snaddon *et al* (2003)	99	2 (2)	2 (100)	0 (0)	0 (0)
Frohling *et al* (2004)	236	17 (7)	15 (88)	0 (0)	2 (12)
Bienz *et al* (2005)	67	7 (10)	4 (57)	0 (0)	3 (43)
(Lin *et al*, 2005)	104	14 (13)	11 (79)	1 (7)	2 (14)
Frohling *et al* (2005)	166	15 (9)	13 (87)	0 (0)	2 (13)
Shih *et al* (2006)	149	20 (13)	18 (90)	1 (5)	1 (5)
Fuchs *et al* (2008)	152	4 (3)	2 (50)	0 (0)	2 (50)
Juhl-Christensen *et al* (2008)	485	5 (1)	4 (80)	0 (0)	1 (20)
Benthaus *et al* (2008)	469	20 (4)	14 (70)	0 (0)	6 (30)
Wouters *et al* (2009)	598	28 (5)	18 (64)	4 (14)	6 (22)
Pabst *et al* (2009)	224	12 (5)	11 (92)	0 (0)	1 (8)
Dufour *et al* (2010)	467	20 (4)	19 (95)	0 (0)	1 (5)
Green *et al* (2010b)	1427	59 (4)	46 (78)	1 (2)	12 (20)
Wen *et al* (2014)	233	11 (5)	6 (55)	1 (9)	4 (36)
Fasan *et al* (2014)	2296	104 (5)	60 (58)	0 (0)	44 (42)
Behdad *et al* (2015)	2393	74 (3)	43 (58)	11 (15)	20 (27)
TOTAL	**10192**	**434 (4)**	**304 (70)**	**21 (5)**	**65 (25)**

that were transplanted with a classic C mutant or both N and C mutants developed leukaemia. The latter studies were subsequently extended to look at the effect of different C mutants (Togami *et al*, 2015). They showed that two different classic C mutants (S299_K304dup and K313dup) and also a C terminal missense mutant (N321D) all induced AML in the mice. However, disease latency with the N321D mutation was much shorter than with the classical mutations at 107 days compared to 151 or 298 days, perhaps indicating that the pathogenesis differs between the different types of mutations.

CEBPA^{DM} samples have a distinct gene expression profile compared to *CEBPA*^{SM} or *CEBPA*^{WT} (Wouters *et al*, 2009). In this study, unsupervised principal component analysis of only *CEBPA*^{MUT} samples revealed a separation of *CEBPA*^{SM} from *CEBPA*^{DM} (Figure 3.2). The first principal component separated the three homozygous C *CEBPA*^{DM} samples from the classic *CEBPA*^{DM}. Four further *CEBPA*^{DM} samples were not located within the remaining cluster of 19 samples by both prinicipal components one and two. All these samples were non-classic *CEBPA*^{DM}, two had a classic N mutation with a missense mutation in the C terminus, and the other two had a classic N mutation coupled with a frameshift mutation in the middle of the gene. Thus there is a suggestion that not all non-classical *CEBPA*^{DM} form part of the distinctive *CEBPA*^{DM} group and this may have implications when grouping patients into prognostic categories.

3.1.4 DNA methylation in AML

Whole genome sequencing of numerous AML samples has revealed surprisingly few recurrent mutations compared to other malignancies (2013; Ley *et al*, 2008; Stratton, 2011; Welch *et al*, 2012). Furthermore, many of the mutations recently identified do not fit into one of the proposed classes of genetic mutations affecting proliferation or differentiation, which has led to questioning of the original model. Several of the mutated genes are known to have epigenetic functions, e.g. *DNMT3A*, *TET2*, *IDH1*, *IDH2*, *WT1* and *ASXL1* (Table 1.3). An aberrant epigenetic state has thus been postulated to play a role in the pathogenesis of AML. The most studied epigenetic alteration is DNA methylation. This is due not only to the stability of the additional methyl group

(COPYRIGHT PROTECTED IMAGE REMOVED)

Figure 3.2 Unsupervised components analysis of gene expression data from 38 *CEBPA*^{MUT} samples.

Figure adapted from Wouters et al (2009). *CEBPA*^{DM} samples are represented by red boxes and *CEBPA*^{SM} by blue boxes. Non-classical *CEBPA*^{DM} samples have been circled. Homozygous C mutant samples are circled in green, samples with a classic N mutation and missense mutation in the C terminus are in yellow, and those with a classic N mutation and frameshift in the middle or C terminus of the gene are in purple.

on nucleotide bases, but also because of the relative ease with which this can be investigated in primary DNA samples compared to histone modifications.

Initially the DNA methylation analysis was at the single gene level, for example methylation of the estrogen receptor was found in a subset of patients using methylation-sensitive restriction enzyme digestion followed by PCR (Li *et al*, 1999). In this study, 261 samples were analysed, of which 61% had a methylated CGI within the promoter of the estrogen receptor (defined as >15% methylation as quantified by Southern blot). As the number of genes shown to be aberrantly methylated in AML increased, many studies investigated the methylation status of several genes to see whether there was a specific "methylation phenotype" in a subset of samples. For example, the methylation of CGIs associated with eight genes previously shown to be methylated in leukaemia (including calcitonin, E-cadherin, *p15*, *p16*, Retinoblastoma) was analysed using bisulfite genomic sequencing in 20 AML and nine normal bone marrow samples (Melki *et al*, 1999). Apart from one normal bone marrow that had methylated calcitonin, all the other normal samples were completely unmethylated at all eight CGIs. Conversely, the AML samples all had at least one methylated CGI, and 75% of them had at least two methylated CGIs. The authors concluded that the aberrant methylation was therefore suggestive of general dysregulation of methylation mechanisms rather than targeted to specific genes.

With advances in the technology, more regions were simultaneously analysed in each sample. For example, Bullinger *et al* (2010) used bisulfite-converted DNA and mass spectrometry, MALDI-TOF, to quantify methylation at approximately 2000 CpG sites in 92 genomic regions in 256 samples. They found that the majority of CpG sites analysed showed very little variability in methylation levels across all samples. Most CpG sites were hypomethylated (median methylation level 10%) and a second smaller group of CpGs were predominantly hypermethylated (median level 70%). Unsupervised cluster analysis of the methylation profiles segregated the samples into several clusters, some of which correlated with underlying cytogenetic abnormalities, for example, most inv(16) samples clustered together, similarly t(15;17) and t(8;21) samples mainly clustered within individual groups.

This link between genetic abnormalities seen in AML and the DNA methylation profile has been corroborated in three studies that have investigated the "methylome" of AML in large numbers of unselected patients. Figueroa *et al* (2010b) used the HELP assay to interrogate over 50,000 CpG sites contained within approximately 14,000 genes in 344 patients. Deneberg *et al* (2011) used the Illumina Infinium Methylation 27K array in samples from 118 cytogenetically normal AML patients. Most recently, the Cancer Genome Atlas Research Network (2013) published a comprehensive genetic and DNA methylation analysis of 190 patients that were analysed using the Illumina Infinium Methylation 450K array. All three papers identified clusters of samples with unique methylation profiles. In the two largest studies which both looked at unselected patients, the DNA methylation profile could distinguish samples with inv(16), t(8;21) or t(15;17) from those without these aberrations with a high degree of accuracy, both in terms of sensitivity and specificity. Furthermore, all three studies highlighted that unsupervised cluster analysis led to several clusters with different methylation patterns, and when these were correlated with the molecular status of the samples, some clusters were enriched for certain mutations. For example, samples with *NPM1* mutations were enriched in particular clusters in all three studies. However, the findings between these studies were not all consistent, for example, *CEBPA* mutations were identified predominantly in just 2 of 16 clusters by Figueroa *et al* (2010b), but Deneberg *et al* (2011) did not find that they were enriched in any particular cluster.

3.1.5 Significance of aberrant DNA methylation in AML

Given that analysis of DNA methylation profiles has repeatedly been shown to differentiate AML from normal bone marrow (Figueroa *et al*, 2010b) and other malignancies (Hansen *et al*, 2011), and also differentiates some cytogenetic and molecular subcategories of AML (2013; Figueroa *et al*, 2010b), it is unlikely that aberrant methylation in AML is simply a random feature of general epigenetic dysregulation. DNA methylation of CGIs associated with tumour suppressor genes has been considered to be an alternative to loss-of-function mutations as a silencing mechanism that would lead to the same phenotype, which would be consistent with the fact that aberrant hypermethylation has

often been seen in promoters of genes known to be recurrently mutated in AML, e.g. *CEBPA* and *DNMT3A* (Hackanson *et al*, 2008; Jost *et al*, 2014). There is, however, evidence that aberrant methylation of CGIs occurs in genes which are already silenced in the cell of origin. By analysing the methylation of gene promoters in 19 breast cancer cell lines and comparing the results with previously published transcriptome data, Sproul *et al* (2011) demonstrated that genes methylated in the cancer cell lines were not expressed in the normal tissue of origin. Furthermore, demethylation of these cell lines using 5 aza cytidine led to derepression and expression of only 10% of the genes shown to be methylated. They subsequently extended this study to examine published methylation array and RNA-sequencing data in over 1000 malignancies in seven different tissue types, including AML, and found that genes that were prone to hypermethylation were not constitutively expressed (Sproul *et al*, 2012). This suggests that aberrant DNA may be a passenger event rather than driving the disease.

There are several reasons why it is important to ascertain whether DNA methylation is pathogenic or not. Firstly, targeting DNA methylation as part of the treatment for AML is an attractive option as methylation is considered to be “reversible” (Kelly *et al*, 2010). Indeed, there has been some success in treatment of MDS and AML with DNMT inhibitors, but it is unclear if the activity of these drugs is due to their demethylating ability (Silverman *et al*, 2006). Secondly, knowledge of the functional role of specific DNA methylation patterns might aid in categorising patients who lack a mutation in a particular gene but have a methylation profile similar to those patients who do have the mutation. For example, Figueroa *et al* (2010b) found that in one cluster of 31 patients, 22 had t(8;21), which is associated with a favourable prognosis, and the remaining nine patients had neither the translocation nor a cryptic *AML1-ETO* fusion gene, but their survival curve was indistinguishable from the 22 t(8;21) patients.

Thirdly, given that clustering of patients based on their methylation profile can be associated with specific cytogenetic and molecular abnormalities, some of which have prognostic impact, the methylation profiles themselves could potentially aid in prognostic stratification. This would be especially useful in AML as many patients fall into the heterogeneous intermediate risk prognostic

category, of which approximately 20% have no known mutations with prognostic value. Two studies that looked at the “methylome” in 118 cytogenetically normal and 344 unselected patients used supervised principal components analysis to create methylation profiles that would predict for prognosis (Deneberg *et al*, 2011; Figueroa *et al*, 2010b). Both groups created methylation predictor signatures consisting of approximately 300 CpG sites and 18 *Hpa*II fragments respectively to classify patients as good or poor outcome. They validated the predictors in an independent set of patients. Both studies showed that the methylation profile retained its prognostic significance in multivariate analysis when age, white cell count*, NPM1* and *FLT3*/ITD status (Deneberg *et al*, 2011) or age, cytogenetic risk, *NPM1*, *CEBPA* and *FLT3*/ITD status (Figueroa *et al*, 2010b) were taken into account. Of note, both the methods used to quantify the methylation and the CpG sites analysed differed between these groups and the corollary was that the two signatures were completely different from one another, with no overlapping regions.

The preliminary aim of the studies presented in this chapter was to ascertain whether prognosis could be associated with the methylation pattern of samples taken at diagnosis. Therefore, rather than screening a large number of unselected patients, as in previously published studies, samples were used from two highly selected cohorts with known outcome, each of 21 patients, that were at either extreme of clinical response, either chemosensitive or chemoresistant. Using the information available at the time of the commencement of the studies, common cytogenetic alterations and mutations known to correlate with outcome were excluded so that all samples analysed were from patients with a normal karyotype (NK) and WT for *NPM1*, *FLT3*/ITD and *FLT3*/TKD. Initial analyses identified a specific methylation profile that, with the availability of additional molecular characterisation, was found to be associated with *CEBPA* mutations. This then became the focus of further investigations.

3.2 Patients, Materials and Methods

Preliminary data was available from the methylation analysis of a cohort of 42 samples that formed the starting point of the studies presented here. Our initial

collaborators, Dr Andy Feber and Dr Andrew Teschendorff, UCL Cancer Institute, UK, prepared the samples for the arrays and performed preliminary cluster analysis respectively. All subsequent cluster analysis was performed by Dr Duncan Sproul, Institute of Genetics and Molecular Medicine, University of Edinburgh, UK.

3.2.1 Sample selection

The initial cohort of 42 DNA samples were selected from AML patients entered onto the UK MRC AML10 and AML12 trials and analysed using the Illumina Infinium 27K HumanMethylation array. Samples were available from a DNA Biobank held in the Department of Haematology, UCL Cancer Institute. Ethical approval for the use of the samples was obtained from the Multi-Centre Research Committee of Wales. Informed consent was obtained in accordance with the Declaration of Helsinki. The samples were from patients known to have an NK, and to be WT for three common mutations, *NPM1*, *FLT3*/ITD and *FLT3*/TKD (Gale *et al*, 2008; Mead *et al*, 2007). To ascertain whether clinical outcome was associated with a particular methylation profile, patients were selected to be at either extreme of clinical response. Half of the patients chosen were chemosensitive, defined as in continuous complete remission for at least 3 years, and the other half were chemoresistant, defined as either failure or slow to respond to chemotherapy or early relapse within 4 months of completing chemotherapy. Two follow-up cohorts of 48 samples each were also selected for analysis from the UK MRC AML trials' DNA bank. Further details of these patients are provided in the results section.

During the course of the project, mutation status for the *CEBPA*, *GATA2*, *IDH1*, *IDH2*, *DNMT3A*, *TET2* and *WT1* genes was determined (Gale *et al*, 2015; Green *et al*, 2010a; Green *et al*, 2011; Green *et al*, 2010b; Green *et al*, 2013; Virappane *et al*, 2008).

3.2.2 Sample analysis on the Illumina Infinium Methylation Arrays

For samples to be analysed on the Illumina Infinium Methylation array (Illumina inc, California, USA), 500ng DNA was bisulfite-converted, as described in section 2.2.2. To assess the quality of the conversion, random samples from

each converted batch were subjected to methylation-specific PCR. Two PCRs of the HLA-B gene were performed per sample assessed, one that would only amplify bisulfite-converted DNA, and the other that would only amplify unconverted DNA. Samples were considered to be successfully bisulfite-converted if they had a PCR product with the primers for the bisulfite-converted DNA but no product with the primers for the unconverted DNA. Details of the conditions for the PCRs are given in section 2.2.3.

Bisulfite-converted samples were sent to UCL Genomics for analysis on the Infinium Methylation array according to the manufacturer's protocol. Briefly, each bisulfite-converted DNA sample was whole genome amplified, enzymatically fragmented and precipitated, then hybridised to the BeadChip. Two different arrays were used during the course of this project. The first cohort was analysed with the HumanMethylation 27K BeadChip, which examines over 27,000 CpG sites predominantly located in CpG islands. Each CpG locus analysed is represented by two bead types, one corresponding to the methylated allele and the other to the non-methylated allele. Both bead types have 50mer probes attached that differ only in their last base which corresponds to the cytosine under investigation, i.e. one bead has a guanine as the last base to bind to methylated cytosines, the other has adenine to bind to thymine for non-methylated cytosines, or C and A respectively if the complementary allele is being interrogated. Labelled nucleotides are then added for single base extension of the correctly hybridised oligonucleotides (Figure 3.3). The proportion of the labelled nucleotides is measured indirectly by multi-layer immunohistochemical staining and laser excitation and the level of light emitted recorded.

The HumanMethylation 450K BeadChip, which was used to examine cohorts 2 and 3, analyses over 450,000 CpG sites. It covers 94% of the CpGs analysed on the 27K BeadChip, 99% of the RefSeq genes, as well as intergenic regions. Not only is there greater coverage of the genome but also more CpGs are analysed per gene, on average 17 probes compared with two for the 27K BeadChip. The limiting factor for the number of CpGs that can be assessed with the array is the number of beads that can be assembled on each BeadChip. To accommodate the extra beads required to analyse over 485,000 CpG sites, the

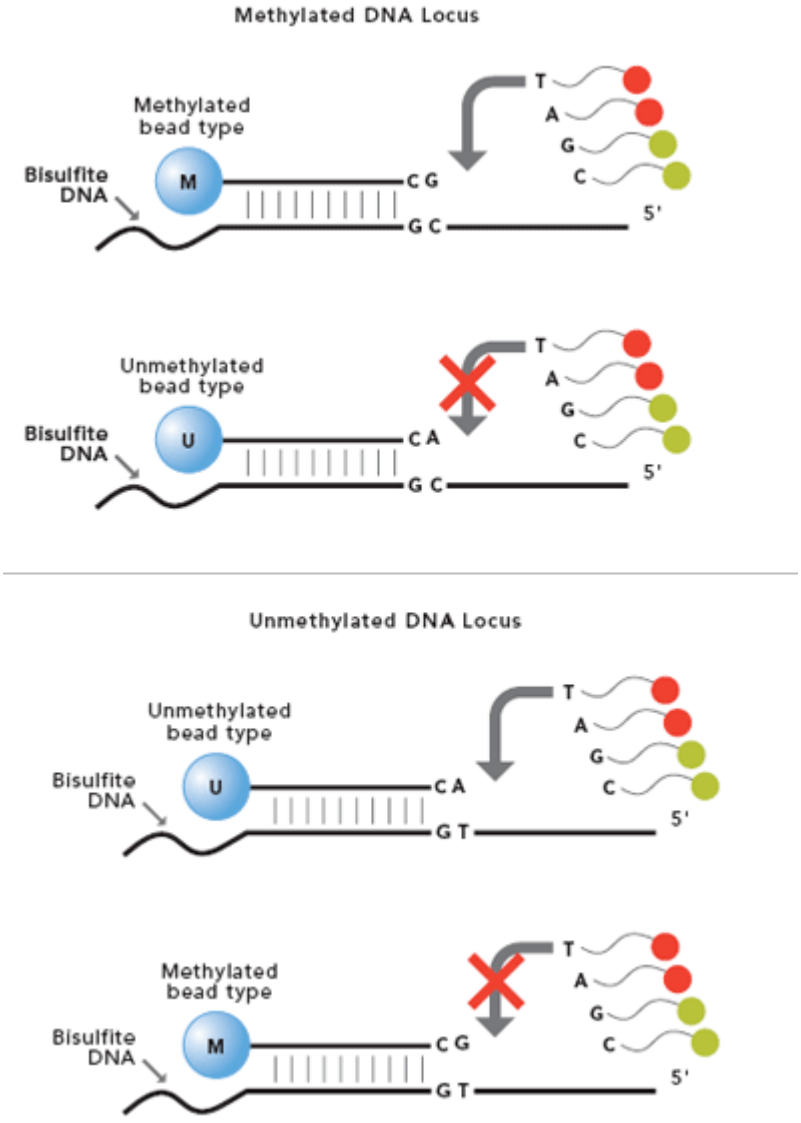


Figure 3.3 Illumina Infinium I assay. The assay uses an unmethylated (U) and methylated (M) bead type for each CpG analysed. In the top panel, the CpG site being analysed is methylated and thus binds to the methylated bead type enabling single base extension and detection, however this will not bind with the unmethylated bead type due to the sequence mismatch. The reverse situation is shown in the bottom panel. (Figure taken from http://www.illumina.com/technology/beadarray-technology/infinium-methylation-assay.ilmn.)

450K BeadChip uses two different assays. Approximately 28% of the CpG sites are analysed by the Infinium I assay, which is the technology employed with the 27K BeadChip. The remaining 72% of CpG sites are analysed by the Infinium II assay that has a single bead type per CpG locus that allows the complementary DNA fragment to hybridise, regardless of the methylation status of the CpG under investigation, rather than an unmethylated and methylated bead type. Methylated and non-methylated alleles are differentiated from one another by generating green and red coloured signals respectively and thus can be quantified separately.

For each CpG site interrogated, the fluorescence intensities measured for the methylated and unmethylated probes were converted to a raw β value corresponding to the methylated signal over the total (methylated plus unmethylated) signal for that CpG site. Hence, the β value was between 0 and 1, with 0 being fully unmethylated and 1 fully methylated. CpG sites analysed using the Infinium II assay were normalised to account for the slight bias in β values produced by this assay from the differently coloured signals (Dedeurwaerder *et al*, 2011). All CpG sites analysed were subjected to filtering to exclude those with a low signal-to-background ratio, based on the detection *p*-value >0.01. Probes interrogating CpG sites on the X and Y chromosome were excluded from the cluster analyses, as were any other probes displaying gender-specific biases, defined as those with Wilcoxon test p-values <0.05 between genders. For display in figures, β values were converted to estimated percentage methylation levels by multiplying them by 100.

3.2.3 Methylation quantification using pyrosequencing

Pyrosequencing was performed as outlined in section 2.2.4. The primers for each assay were designed using the PyroMark Assay Design Software 2.0 (Qiagen, Germany), and initially tested using titration standards. Titration standards containing 0%, 10%, 25%, 50%, 75%, 90% and 100% methylated DNA were prepared by two methods. One method used in-house mixes of fully unmethylated DNA, created by whole genome amplification of normal genomic DNA, and fully methylated DNA, created by using the enzyme SssI (New England Biolabs, USA) to methylate DNA, followed by bisulfite conversion of the

individual mixes. The other method mixed purchased bisulfite-converted fully unmethylated and fully methylated Epitect DNA standards (Qiagen, Crawley, UK). If the mixes showed methylation bias with the known standards, new primers were designed to improve the accuracy of the analysis. Once the assays produced by the titration curves were considered to be sufficiently accurate, aliquots of patient samples were bisulfite-converted and analysed in duplicate.

Pyrosequencing assays were created for four CpG sites to validate the results from the methylation array. The mean of the duplicate values for each sample were compared to the corresponding β value. Analysis of correlation data was performed using GraphPad Prism (Version 6.01 for windows, GraphPad, USA). Three further pyrosequencing assays were created to analyse the methylation levels of CpG sites in AML samples with favourable prognostic cytogenetic or molecular changes.

3.3 Results

At the commencement of these studies, preliminary data was available from the initial cohort of 42 patients (21 chemosensitive and 21 chemoresistant) that had been processed by Dr Andy Feber and analysed by Dr Andrew Teschendorff. β values for each CpG site for all patients was provided, as well as the top 100 CpG sites that showed the largest difference between the chemosensitive patients and the chemoresistant patients. Four of these CpG sites were chosen to verify the results of the methylation array using pyrosequencing.

3.3.1 Validation of array results using pyrosequencing

Pyrosequencing assays were designed and titration curves created for four CpG sites associated with the *SOCS2*, *WNT1* and *PRF1* genes, as detailed in the Materials and Methods. Fresh patient samples were bisulfite-converted and tested in duplicate.

Suppressor of cytokine signalling 2, SOCS2, is a negative regulator of the JAK-STAT pathway (Krebs & Hilton, 2001), and has been identified as being upregulated in murine leukaemia stem cells transfected with a *FLT3*/ ITD

(Mizuki *et al*, 2003). The *SOCS2* CpG site evaluated, CG04797323, is located in the intron between exons 2 and 3. The coding sequence for this gene begins in exon 2. Representative examples of the pyrograms obtained with this assay are shown in Figure 3.4. The titration curve obtained using the DNA mixes showed a good correlation between the expected and observed levels of methylated alleles, r^2= 0.99 ($p<0.0001$) (Figure 3.5A). In the patient samples, the range of methylation values was 3%-89%, mean 33% and median 25% (Figure 3.6A). There was a highly significant correlation between the methylation levels predicted by pyrosequencing and those predicted by the Illumina methylation array, r^2 =0.93 ($p<0.0001$) (Figure 3.7A). Comparing the pyrosequencing results of the chemosensitive and chemoresistant patients, there was a significant difference in median levels of methylation between the two groups, 40% and 9% respectively (p=0.02).

Wingless type 2, WNT2, is a protein implicated in the WNT- β catenin signalling pathway that is required for self-renewal of leukaemic stem cells (Wang *et al*, 2010). The CpG site evaluated, CG018302894, is located 149 bases upstream of the transcription start site (TSS) for WNT2. The titration curve showed a good correlation between observed and expected values, r^2 =0.98 ($p<0.0001$) (Figure 3.5B). The patient samples had methylation values ranging from 8%-91%, with mean 43% and median 46% (Figure 3.6B). Comparing the pyrosequencing results with the β values from the methylation array, a highly significant correlation was seen, r^2= 0.93 ($p<0.0001$) (Figure 3.7B). The chemosensitive and chemoresistant cohorts had significantly different median levels of methylation, 49% versus 29% respectively (p=0.01).

Perforin 1, PRF1, is a major component of cytolytic vesicles and a key effector of natural-killer cell-mediated cytolysis (Yanai *et al*, 2003). The gene has been reported to be significantly differentially methylated in DNA samples from patients with de novo and secondary AML (Figueroa *et al*, 2009a). Two CpG sites were investigated for this gene, CG02374486 (PRF1A), 222 bases upstream, and CG09914304 (PRF1B), 298 bases downstream of the TSS. Although both had titration curves with significant correlation co-efficients, r^2 =0.79 ($p<0.0001$) and r^2 =0.81 ($p<0.0001$) respectively (Figure 3.5C and D), there was evidence of preferential amplification of the methylated alleles at both

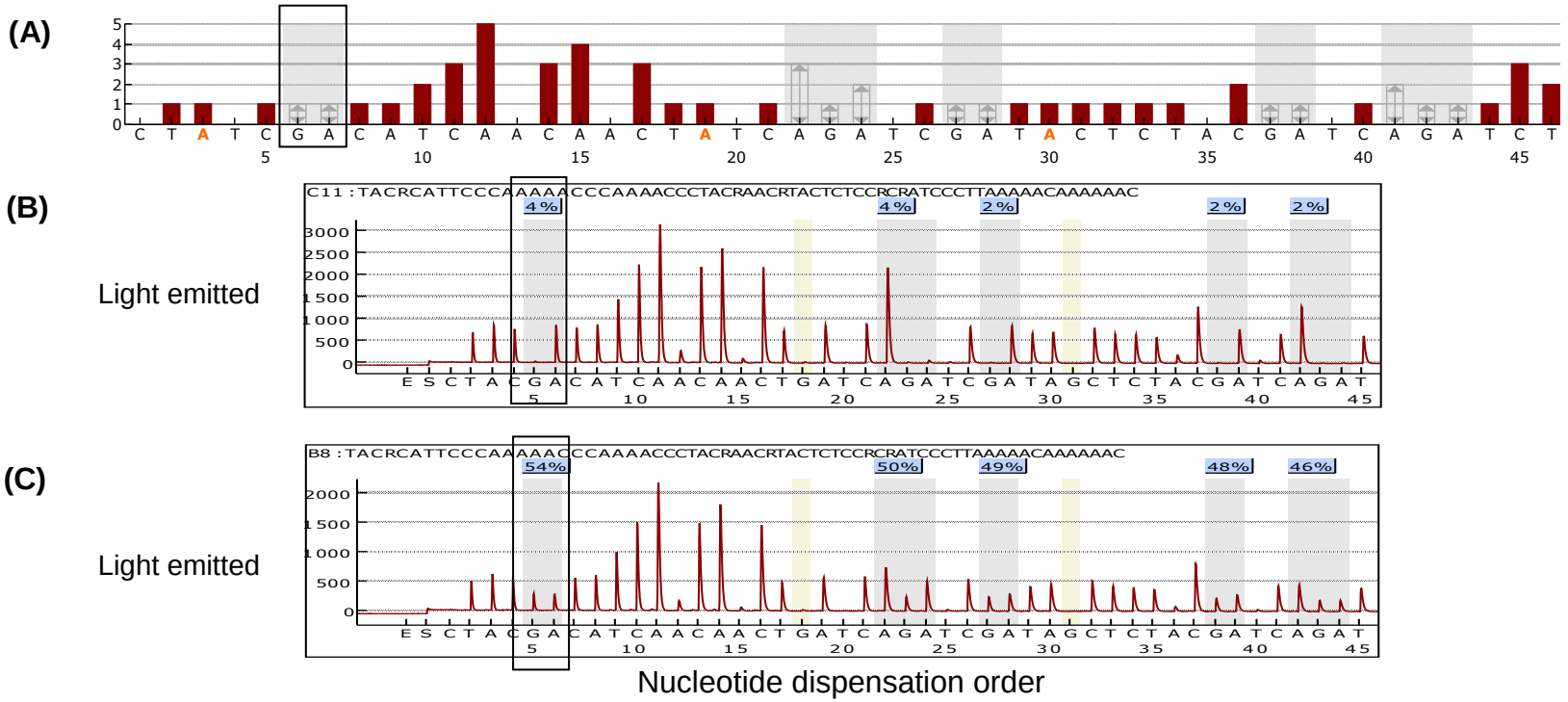


Figure 3.4 Pyrosequencing assay design and representative pyrograms for the SOCS2 probe. (A) Sequence with the individual bars showing the expected relative peaks for each nucleotide. Negative controls are included in the design, as well as controls to ensure that complete bisulfite conversion has occurred, i.e. non-CpG Cs (or Gs if the complementary strand is being examined) that should be completely converted to T (or A). The shaded areas represent the CpG sites analysed in the assay, the one corresponding to the CpG site analysed in the Illumina methylation array is boxed. (B) and (C) are examples of pyrograms produced with this assay. The percentage methylation is given for each CpG site analysed. E and S represent addition of enzyme and substrate mix respectively.

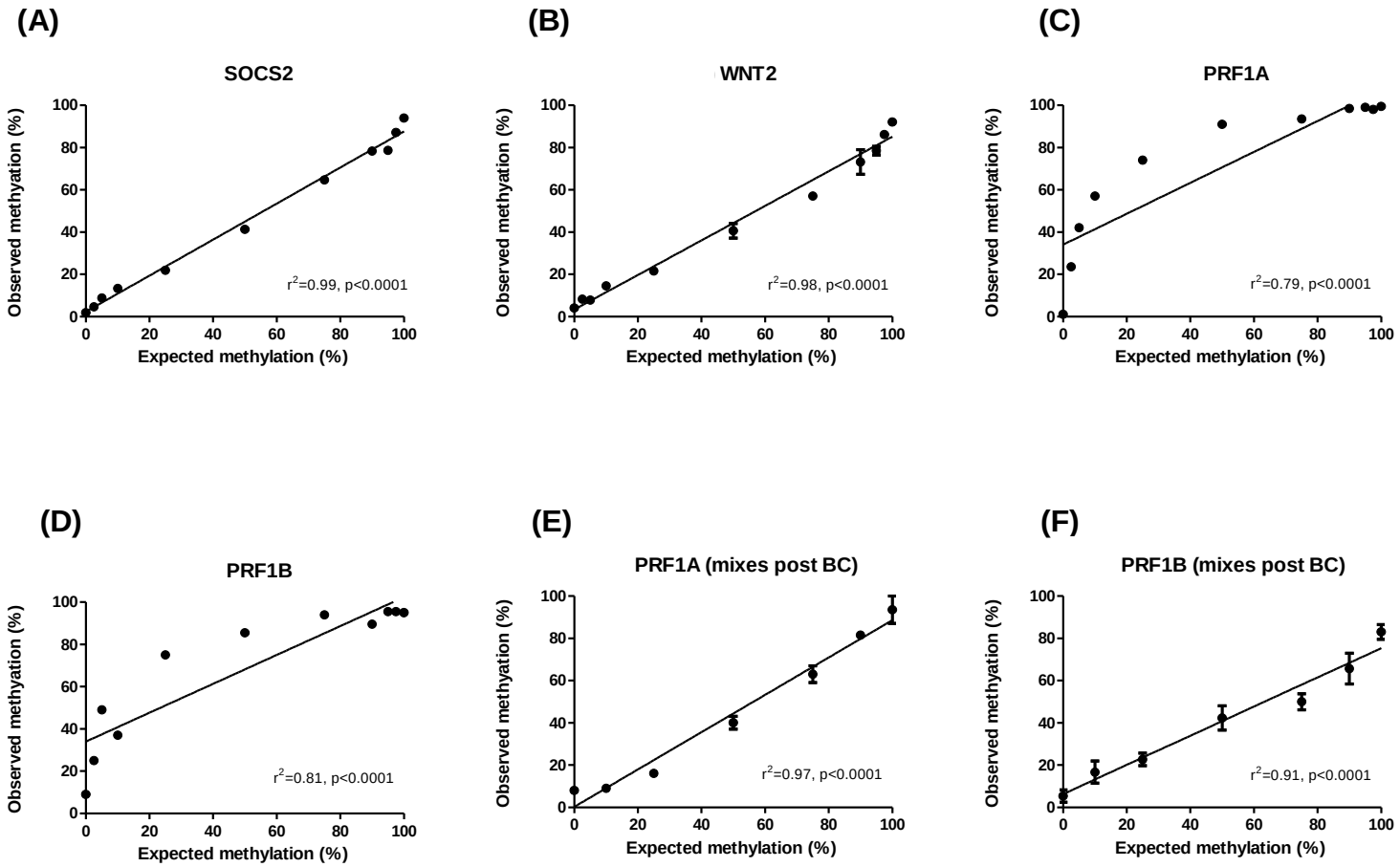


Figure 3.5 Titration curves for the percentage methylation obtained using standards in the pyrosequencing assays. For (A) – (D), mixes for titration standards were made pre-bisulfite conversion (BC). For (E) – (F), fully methylated and fully unmethylated DNA were bisulfite-converted and then mixed to create standards.

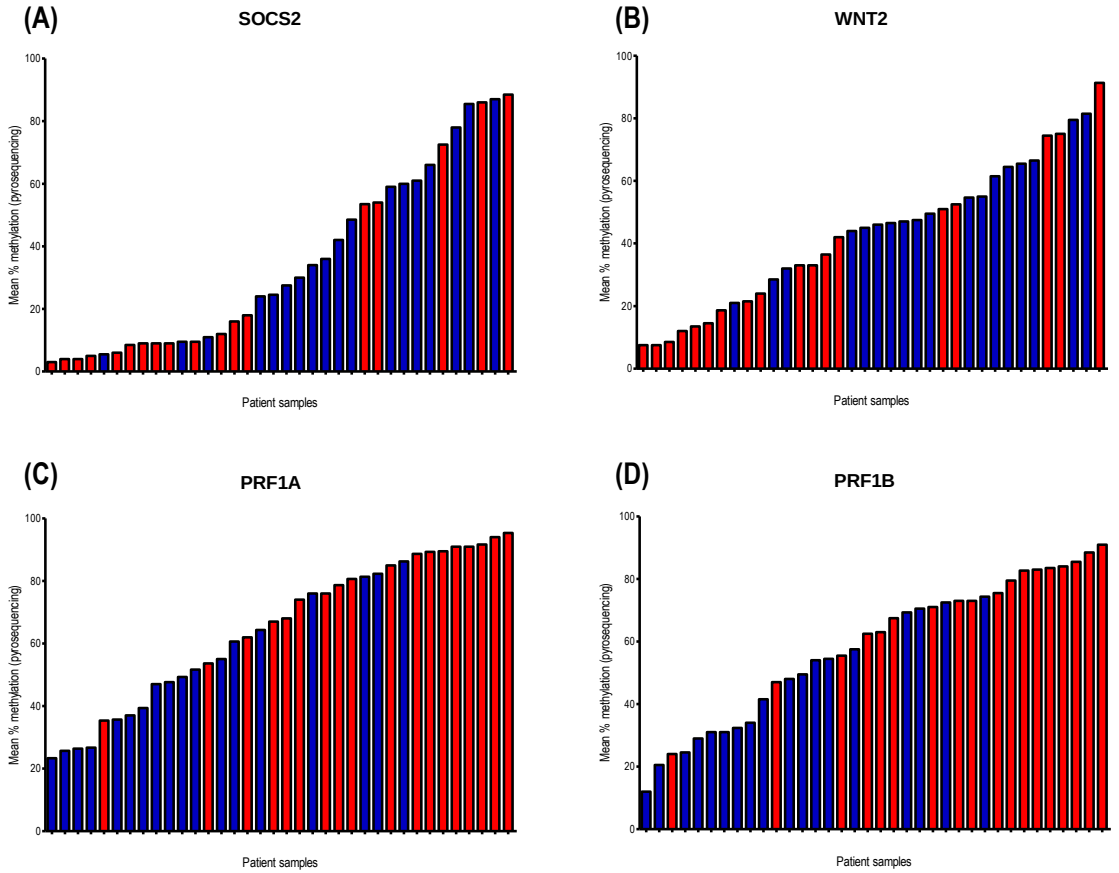


Figure 3.6 Pyrosequencing results for the patient samples as analysed at the selected CpG sites.

(A) SOCS2, (B) WNT2, (C) PRF1A, (D) PRF1B. Each column represents a patient sample, blue columns represent samples from patients with chemosensitive disease and red columns those with chemoresistant disease.

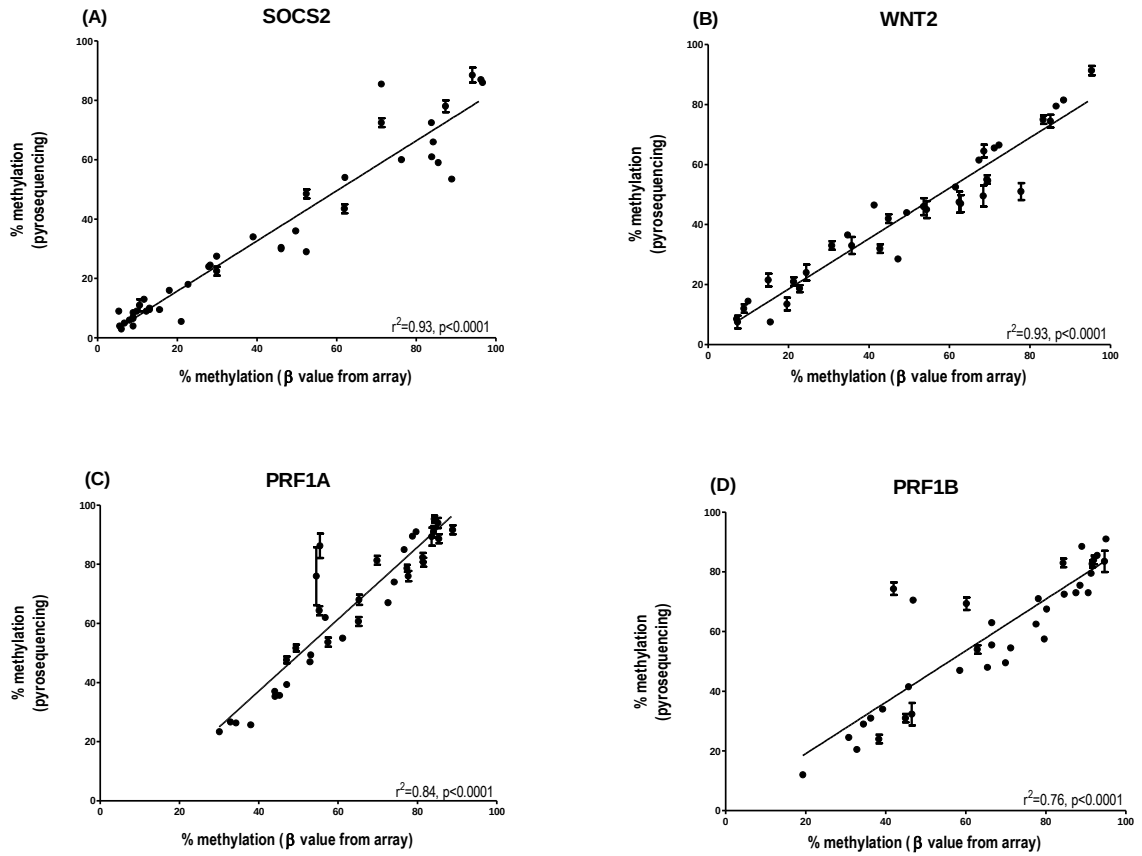


Figure 3.7 Comparison of the methylation values obtained by pyrosequencing with beta values from the methylation array.

sites, as the data points would best fit a second order curve rather than a line. This is a recognised complication of PCR of bisulfite-converted DNA, particularly in CpG islands where the final GC content will vary greatly in methylated versus unmethylated DNA strands (Warnecke *et al*, 1997). However, an alternative explanation was that the DNA mixes created were not accurate for this particular chromosomal region, or that the bisulfite conversion itself was the source of bias. By mixing purchased bisulfite-converted fully methylated and unmethylated DNA, an improvement in the correlation between observed and expected was seen (Figure 3.5E and F). In patient samples, the range of methylation values was 23%-95% and 12%-91% for the two sites respectively, means 65% and 58%, and medians 68% and 63% (Figure 3.6C and D). Despite the preferential amplification of methylated alleles, both CpG sites had a significant correlation between the methylation levels predicted by pyrosequencing and those predicted by the methylation array, r^2= 0.84 ($p<0.0001$) and 0.76 ($p<0.0001$) respectively (Figure 3.7C and D). Comparing the chemosensitive to the chemoresistant cohort, there was a significant difference in median levels of methylation between the two groups, 49% versus 83% ($p=0.0004$) and 45% versus 74% ($p=0.0003$) respectively.

Bland-Altman plots showing the difference between the quantification by pyrosequencing and array for each sample compared to the mean result from both methods were produced for each probe (Figure 3.8). A negative point indicated that the pyrosequencing level was lower than that measured by the array. Overall, there was a good correlation between methylation levels measured by the two methods. In three of the four probes, there was a consistent bias in that pyrosequencing quantified the level to be approximately 10% less than the array (consistent with the titration curves). In PRF1A there was a proportional bias, with pyrosequencing giving lower methylation levels in largely unmethylated samples, but higher levels than the array in more methylated samples. However, the standard deviation of the difference between the two methods was low for all four probes (range 7-11%). Furthermore, biologically a difference of 10% methylation is not thought to be significant.

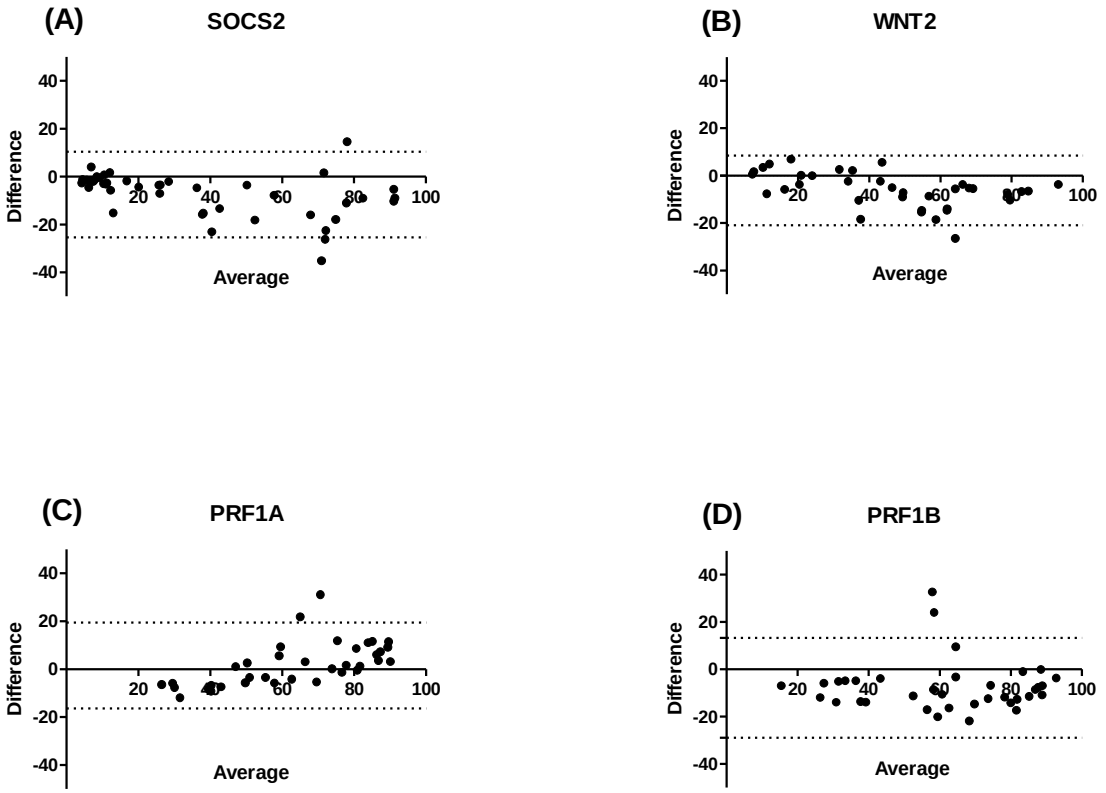
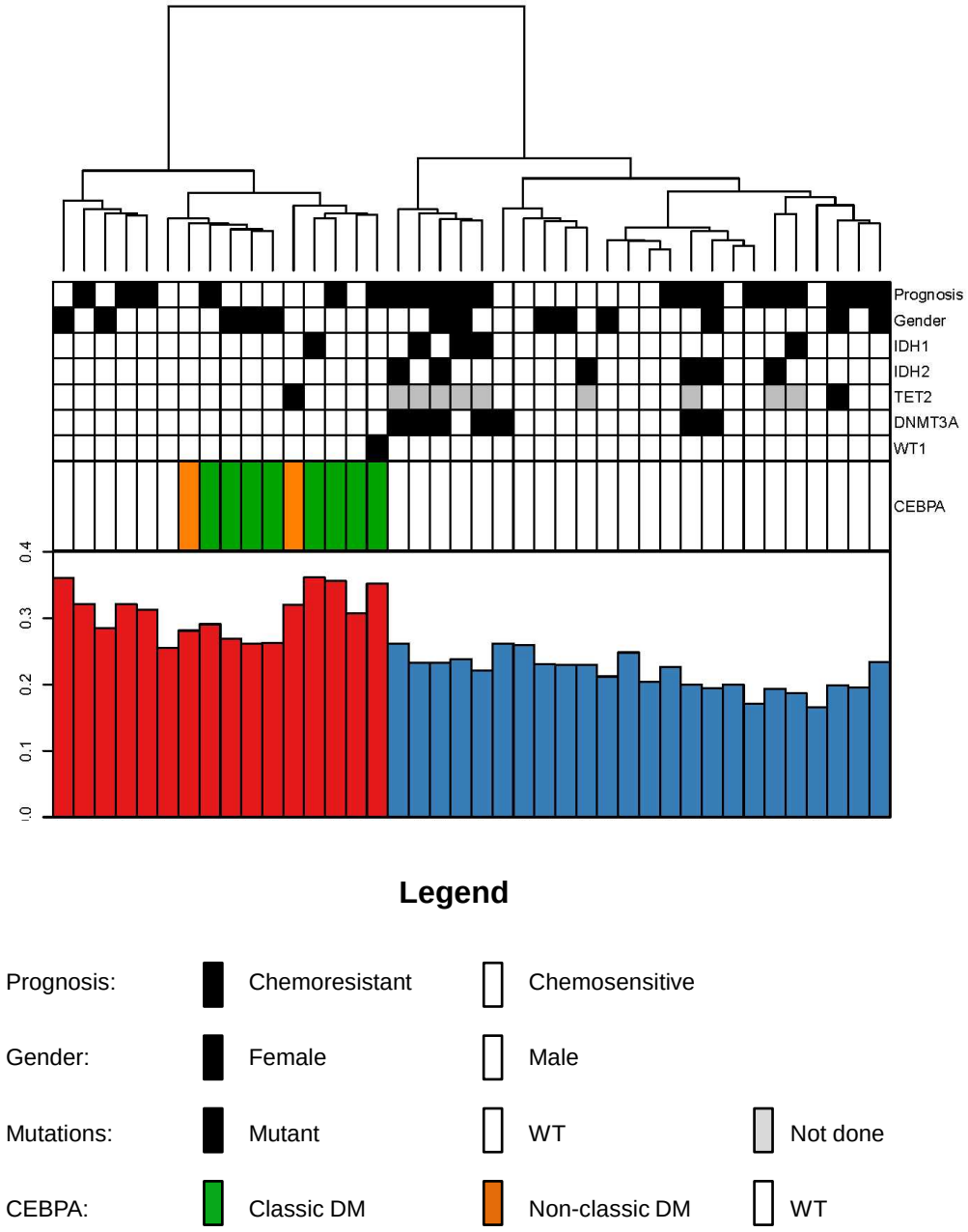


Figure 3.8 Bland-Altman plots to compare the pyrosequencing and methylation array quantification results. The mean result for each sample as calculated by each method is plotted against the difference in values (pyrosequencing result minus the array result). A result below zero on the x axis signifies that the array estimates the methylation level to be higher than the level measured by pyrosequencing. The dotted lines represent the 95% confidence intervals.

3.3.2 Cluster analysis of samples based on methylation array results

The data from the methylation array was analysed further by Dr Duncan Sproul, using both unsupervised and supervised cluster analysis of CpGs located within CGIs. Two samples were excluded from the analysis as they failed quality control measures, leaving 20 patients in each cohort. The majority of CpG sites analysed showed little variation in methylation levels across the whole cohort, however a proportion of the CpG sites had a wide range of methylation values. Samples were therefore grouped according to the methylation levels of the most differential probes using unsupervised cluster analysis, and methylation levels of CGIs (as defined in (Illingworth *et al*, 2010) were derived by calculating the mean β value of probes at these locations (Figure 3.9). Two main groups were found showing either a predominantly hypermethylated (n=16) or hypomethylated (n=24) profile. These clusters did not correlate with outcome, six of the patients in the hypermethylated group were chemoresistant and ten were chemosensitive, whereas in the hypomethylated group 14 were chemoresistant compared to 10 chemosensitive (p=0.20).

Although not known at the commencement of the study, by the time analysis of the methylation array was performed, the mutation status for the *IDH1*, *IDH2*, *CEBPA*, *WT1*, *TET2* and *DNMT3A* genes had been determined for all samples. All samples from *CEBPA*^{MUT} patients (n=10) had a “hypermethylated phenotype”. Further investigation showed that all these *CEBPA*^{MUT} samples were *CEBPA*^{DM}; eight of them (80%) were classic DM, one had a homozygous missense mutation in the C terminus and the other was predicted to produce p30 only as it had a classic N mutation with a second frameshift mutation after the second ATG site (Table 3.3 and Appendix Table 2). The mean CpG island methylation level for the classic *CEBPA*^{DM} samples was significantly higher than that of the *CEBPA*^{WT} samples (Figure 3.10). Conversely, all patients with a *DNMT3A* (n=7), *IDH2* (n=6) and *IDH1* (n=5) mutation were in the hypomethylated group, apart from one patient who had both *IDH1* and *CEBPA* mutations. One patient with a *WT1* mutation and one of two with a *TET2* mutation were in the hypermethylated group, however these patients also had a *CEBPA* mutation.



Legend

Prognosis:	Chemoresistant	Chemosensitive	
Gender:	Female	Male	
Mutations:	Mutant	WT	Not done
CEBPA:	Classic DM	Non-classic DM	WT

Figure 3.9 Unsupervised cluster analysis based on the most differentially methylated CpG sites. CpG sites on the X chromosome were excluded from the analysis. The median CpG β value for each sample is shown in the bottom panel and the clustering in the top panel. The red and blue columns indicate samples with a predominantly hypermethylated or hypomethylated profile respectively. Patient characteristics and genotype are given in the top panel.

Table 3.3 *CEBPA* genotype of samples investigated using methylation arrays

Cohort (n)	*CEBPA* genotype*	Mutation type	n	Predicted functional consequence
1 (40)	DM	Classic N + C	8	p30 + C-LOF
		Classic C + C-frameshift	1	C-LOF
		Homozygous C-missense	1	C-LOF
	WT		30	WT
2 (48)	DM	Classic N + C	13	p30 + C-LOF
		Classic N + mid-frameshift	2	p30
		Homozygous C-missense	1	C-LOF
	SM	Classic N	2	p30 + WT
		Classic C	3	C-LOF + WT
		Mid-frameshift	2	Null + WT
		C-missense	1	C-LOF + WT
	WT		24	WT
3 (47)	DM	Classic N + C	4	p30 + C-LOF
		Classic N + C-missense	3	p30 + C-LOF
		Homozygous classic C	2	C-LOF
		Homozygous C-missense	1	C-LOF
		Classic C + C-missense	1	C-LOF
		Classic N + mid-frameshift	3	p30
		Classic N + C-frameshift	1	p30
	SM	Classic N	7	p30 + WT
		Classic C	2	C-LOF + WT
		Mid-indel	3	UNK + WT
		Mid-frameshift	7	Null + WT
		Mid-missense	5	UNK + WT
		C-frameshift	2	Null + WT
		C-missense	4	C-LOF + WT
	WT		2	WT

*Details of the specific mutations are given in Appendix 2.

Abbreviations: C, C-terminal mutation; C-LOF, C-terminal loss-of-function; indel, in-frame insertion and/or deletion; N, N-terminal mutation; n, number of patients; UNK, unknown; WT, wild-type

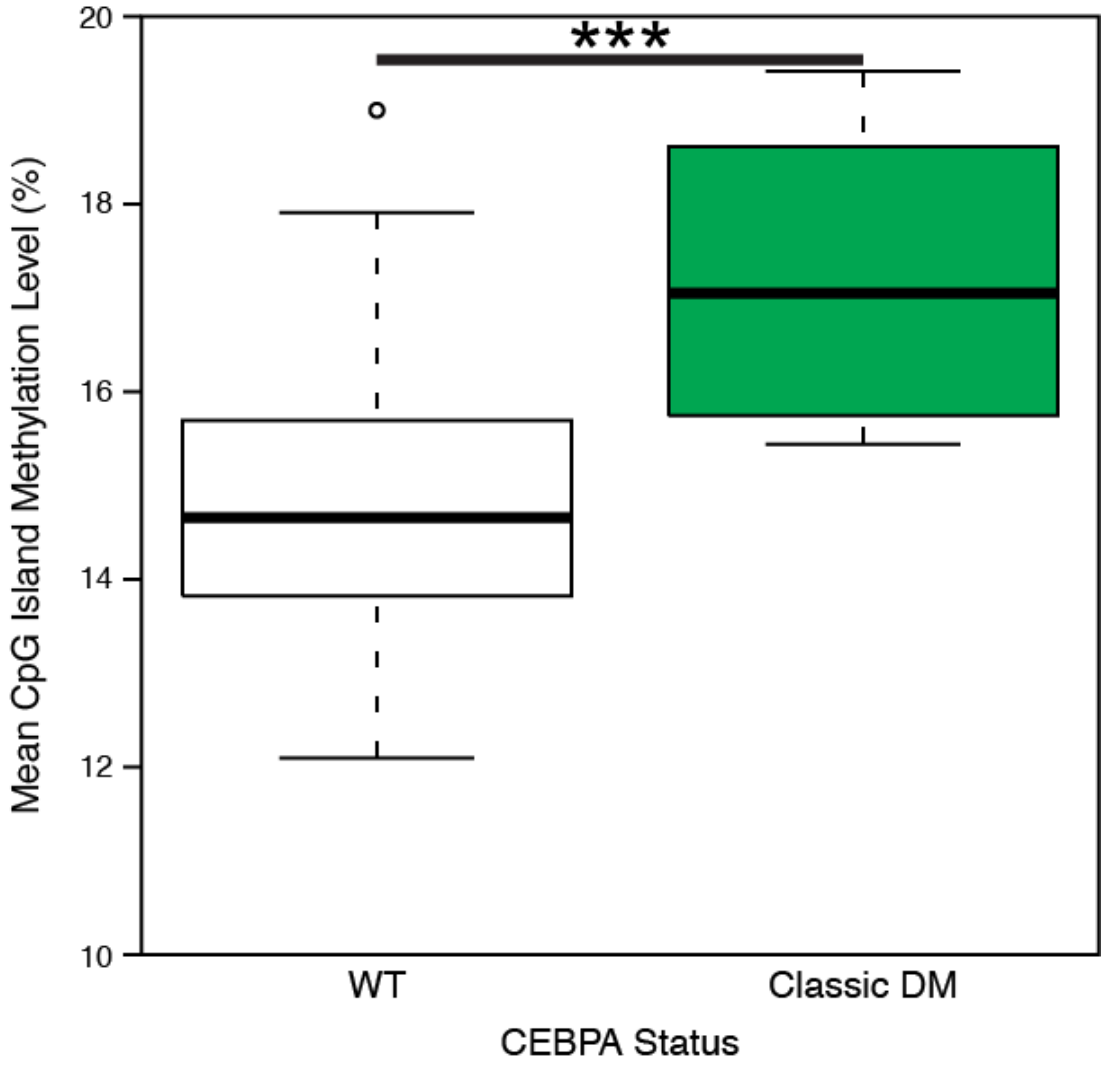


Figure 3.10 Boxplot showing the mean CpG Island methylation level in classic *CEBPA*^{DM}.and *CEBPA*^{WT} samples. β values were derived from the Illumina Infinium Methylation 27K array analysis of the 40 samples in cohort 1, median methylation levels for each CGI were determined and a mean overall level for all probes was then calculated.

3.3.3 Creation of *CEBPA* methylation signatures

Given the strong segregation of *CEBPA*^{MUT} from *CEBPA*^{WT} in the cluster analysis, Dr Sproul used a supervised approach to create a classic *CEBPA*^{DM} and a *CEBPA*^{WT} methylation signature, the two non-classic *CEBPA*^{DM} were not considered in this analysis. In a given sample, probes were defined as methylated if their beta value was >0.3 and unmethylated if beta was ≤0.3. CpG probes were considered differentially methylated if there was at least one unmethylated and one methylated sample. The signature comprised the top 25 most differentially methylated CpG sites (Table 3.4). The function of the 20 genes associated with these 25 sites varies and none of them are known to be C/EBPα targets. Six of these genes have been reported to have infrequent mutations in AML samples (*TMEM125*, *LTBP3*, *AHNAK*, *GRHL3*, *NDFIP1* and *LAMA4*) (Table 3.5). Furthermore, five of the genes (*RAB34*, *KHNYN*, *LTBP3*, *NDFIP1* and *ARPP21*) were differentially expressed in *CEBPA*^{DM} AML samples compared to other AML samples, including three which were included as gene expression predictors to identify *CEBPA*^{DM} samples (Hollink *et al*, 2011; Taskesen *et al*, 2011; Wouters *et al*, 2009). Promoter methylation in AML samples has been reported in three of the genes (*GNMT*, *TMEM125* and *ARPP21*). Two genes encode T cell proteins, *CD52* and *LY9*, and aberrant expression of T cell markers has been reported in some AML samples (Lewis *et al*, 2007; Wouters *et al*, 2007).

Comparison of the derived signatures with published methylation data of different blood cell types (Calvanese *et al*, 2012) showed that the methylation profile of both normal CD34+ cells and neutrophils were more similar to the *CEBPA*^{WT} signature than the classic *CEBPA*^{DM} signature, suggesting that the changes seen in mutated samples were not simply due to a change in the predominant cell type within each sample (Figure 3.11).

The patient samples were then ordered according to their similarity to the classic *CEBPA*^{DM} signature, and two distance scores were calculated for each sample based on the Euclidian distance between their methylation levels at these signature probes and the median profile of classic *CEBPA*^{DM} and *CEBPA*^{WT} samples (Figure 3.12). By comparing the two scores for each sample

Table 3.4 List of the genes or regions associated with the 25 differentially methylated CpG sites in the *CEBPA* methylation signature.

Genomic Location							Signature Details			
Probe ID	Chr	Position	Dist to TSS	ENSEMBL Gene ID	Symbol	Description	In CpG Island	WT Median Beta	Classic DM Median Beta	Median Δ
cg21237418	17	24069170	-157	ENSG00000109113	RAB34	RAB34, member RAS oncogene family	FALSE	0.153	0.928	0.775
cg14338887	6	43036478	0	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.170	0.918	0.748
cg17186163	10	44794323	7	ENSG00000165507	C10orf10	Chromosome 10 open reading frame 10	FALSE	0.155	0.907	0.752
cg24101359	6	43036473	5	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.199	0.904	0.705
cg13105904	14	23969884	-903	ENSG00000100441	KHNYN	KH and NYN domain containing	TRUE	0.306	0.889	0.583
cg01274660	7	100303561	-675	ENSG00000087077	TRIP6	Thyroid hormone receptor interactor 6	FALSE	0.317	0.888	0.572
cg04355435	1	43508877	117	ENSG00000179178	TMEM125	Transmembrane protein 125	FALSE	0.349	0.879	0.531
cg10056627	6	43036751	-273	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.249	0.877	0.629
cg27588902	6	43036129	349	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.265	0.854	0.589
cg25651505	2	85665534	-492	ENSG00000168899	VAMP5	Vesicle associated membrane protein 5	TRUE	0.347	0.837	0.489
cg23696834	6	43036323	155	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.102	0.822	0.720
cg24081819	8	27404857	-295	ENSG00000120915	EPHX2	Epoxide hydrolase 2, cytoplasmic	TRUE	0.243	0.817	0.574
cg08965235	11	65081734	541	ENSG00000168056	LTBP3	Latent TGF beta binding protein 3	TRUE	0.266	0.804	0.538
cg16068833	1	26517102	-104	ENSG00000169442	CD52	CD52 molecule	FALSE	0.227	0.763	0.536
cg19764555	11	62071695	-787	ENSG00000124942	AHNAK	AHNAK nucleoprotein	TRUE	0.272	0.763	0.491
cg00350296	11	65841417	-343	ENSG00000174807	CD248	CD248 molecule, endosialin	FALSE	0.253	0.738	0.485
cg10798171	7	8268826	-59	ENSG00000003147	ICA1	Islet cell autoantigen 1	TRUE	0.249	0.715	0.466
cg15032239	15	20443395	709	ENSG00000068793	CYFIP1	Cytoplasmic FMR1 interacting protein 1	TRUE	0.195	0.708	0.513

Table 3.4 continued

Genomic Location							Signature Details			
Probe ID	Chr	Position	Dist to TSS	ENSEMBL Gene ID	Symbol	Description	In CpG Island	WT Median Beta	Classic DM Median Beta	Median Δ
cg16155382	1	24518778	-135	ENSG00000158055	GRHL3	Grainyhead-like transcription factor 3	FALSE	0.101	0.678	0.577
cg13490971	5	141468305	203	ENSG00000131507	NDFIP1	Nedd4 family interacting protein 1	TRUE	0.203	0.653	0.450
cg21697134	17	78287128	-331	ENSG00000167363	FN3K	Fructosamine 3 kinase	FALSE	0.090	0.614	0.524
cg08897388	6	112682398	44	ENSG00000112769	LAMA4	Laminin subunit alpha 4	TRUE	0.135	0.575	0.440
cg12417466	3	35658823	30	ENSG00000172995	ARPP21	CAMP regulated phosphoprotein 21kDa	FALSE	0.720	0.179	-0.541
cg05615150	3	35658819	34	ENSG00000172995	ARPP21	CAMP regulated phosphoprotein 21kDa	FALSE	0.668	0.113	-0.555
cg18920397	1	159032429	123	ENSG00000122224	LY9	Lymphocyte antigen 9	FALSE	0.625	0.071	-0.554

[1]CpG ID as numbered by Illumina Methylation BeadChip.

Abbreviations: Chr, chromosome; Dist, distance; DM, double mutant; TSS, transcription start site; WT, wild-type

Table 3.5 Published associations between genes in the *CEBPA* signature and leukaemia.

Gene[1]	Mutations seen in AML[2]	Changes in expression in AML	Evidence of methylation in AML	References
RAB34		Top 50 most significantly differentially expressed genes comparing the *CEBPA* cluster (↓) vs other AML samples		Hollink *et al* (2011)
RAB34		Top 25 differentially expressed genes comparing $CEBPA^{DM}$ (↓) vs other AML samples		Taskesen *et al* (2011) Wouters *et al* (2009)
GNMT			Promoter methylation seen in AML Associated with improved survival in univariate analysis	Wilop *et al* (2011)
KHNYN		Downregulated expression in $CEBPA^{MUT}$		Marcucci *et al* (2008)
TMEM125			Hypermethylated in transformed PV/ MPN compared to non-transformed.	Perez *et al* (2013)
TMEM125	1 in germline CBL^{MUT}			Becker *et al* (2014)
EPHX2		Differential expression associated with promoter methylation seen in childhood T-ALL		Borssen *et al* (2013)
LTBP3	1/193			COSMIC COSU377 Cancer.sanger.ac.uk Forbes *et al* (2015)
LTBP3		Increased expression in R172 IDH2 mutant AML samples compared to IDH1/2 WT		Marcucci *et al* (2010)
CD52		Lymphocyte differentiation Ag expressed in approx. 36% AML samples, associated with high EVI1 expression.		Saito *et al* (2011)
CD52		Decreased expression in $CEBPA^{DM}$ compared to other AML samples		Dufour *et al* (2010)

Table 3.5 continued

Gene[1]	Mutations seen in AML[2]	Changes in expression in AML	Evidence of methylation in AML	References
AHNAK	2/182			COSMIC COSU544 Cancer.sanger.ac.uk Forbes *et al* (2015)
GRHL3	1/182			COSMIC COSU544 Cancer.sanger.ac.uk Forbes *et al* (2015)
NDFIP1	1/182			COSMIC COSU544 Cancer.sanger.ac.uk Forbes *et al* (2015)
NDFIP1		Differential expression in *CEBPA*^{DM} (↓) compared to other AML samples		Dufour *et al* (2010)
NDFIP1		Part of 25 probe set differentiating *CEBPA*^{DM} (↓ expression) from other AML samples		Taskesen *et al* (2011) Wouters *et al* (2009)
LAMA4	2/182			COSMIC COSU544 Cancer.sanger.ac.uk Forbes *et al* (2015)
ARPP21		Top 50 most significantly differentially expressed genes comparing the *CEBPA* cluster (↑) vs other AML samples		Hollink *et al* (2011)
ARPP21		Used as part of a gene expression predictor to identify silenced *CEBPA* samples (↑ in methylated *CEBPA*)		Wouters *et al* (2007)

[1]All CpG sites in the signature associated with the genes listed in the table are hypermethylated in the classic *CEBPA*^{DM} apart from ARPP21 which is hypomethylated.

[2]number of samples with mutations detected/ number of samples analysed

↑= Increased expression; ↓= decreased expression

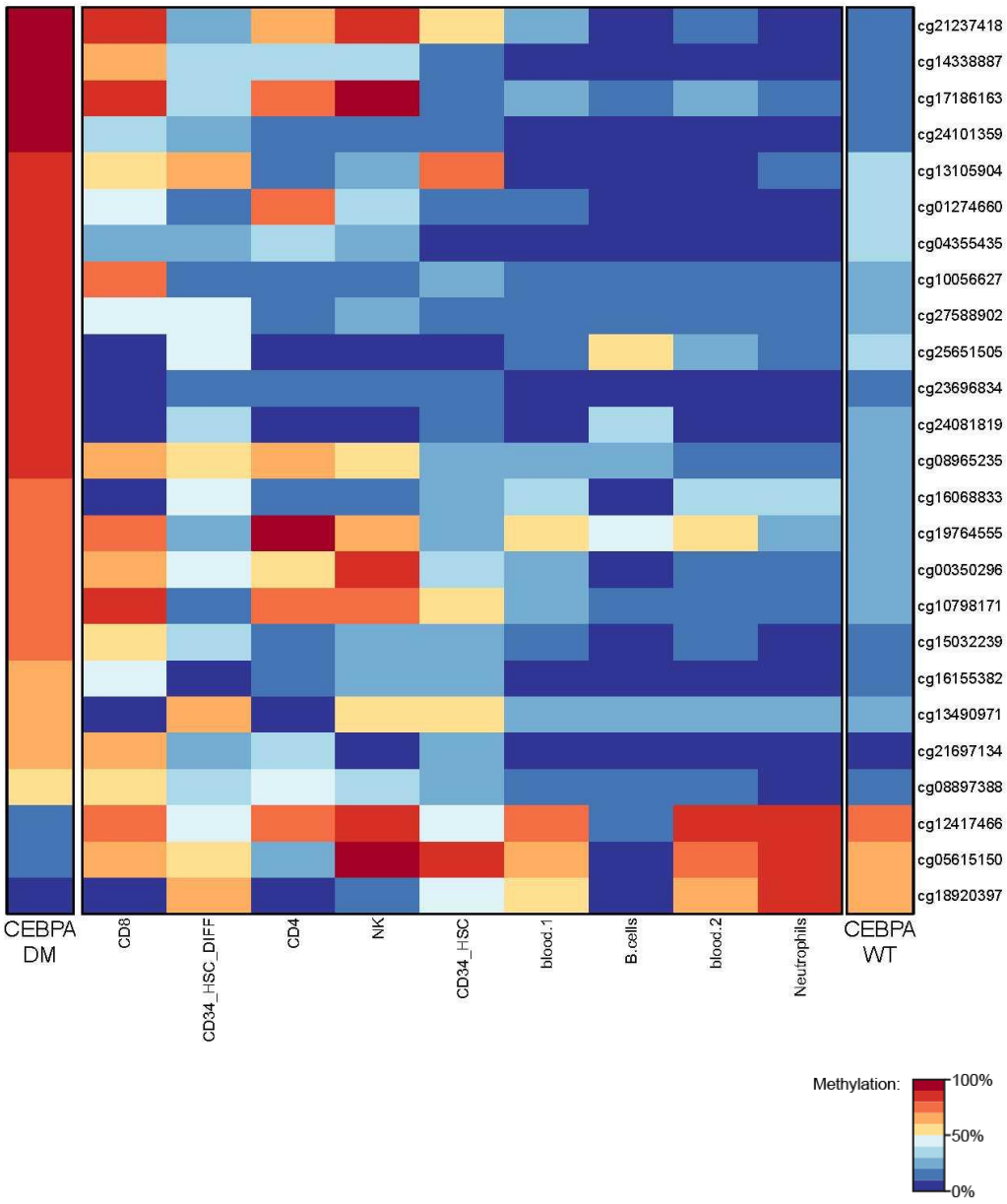


Figure 3.11 Comparison of the classic CEBPA^{DM} and CEBPA^{WT} methylation signatures to the methylation levels at these 25 CpG sites in normal bone marrow and blood cells. β values were derived from published methylation levels (Calvanese *et al*, 2012). Cell subtypes had been purified using magnetic bead separation and FACS sorting. Blood1 and blood2 are peripheral blood samples, i.e interrogating all peripheral mononuclear blood cells; NK, natural killer cells; CD34_HSC, CD34 selected haematopoietic stem cells; CD34_DIFF, CD34 cells differentiated *in vitro*.

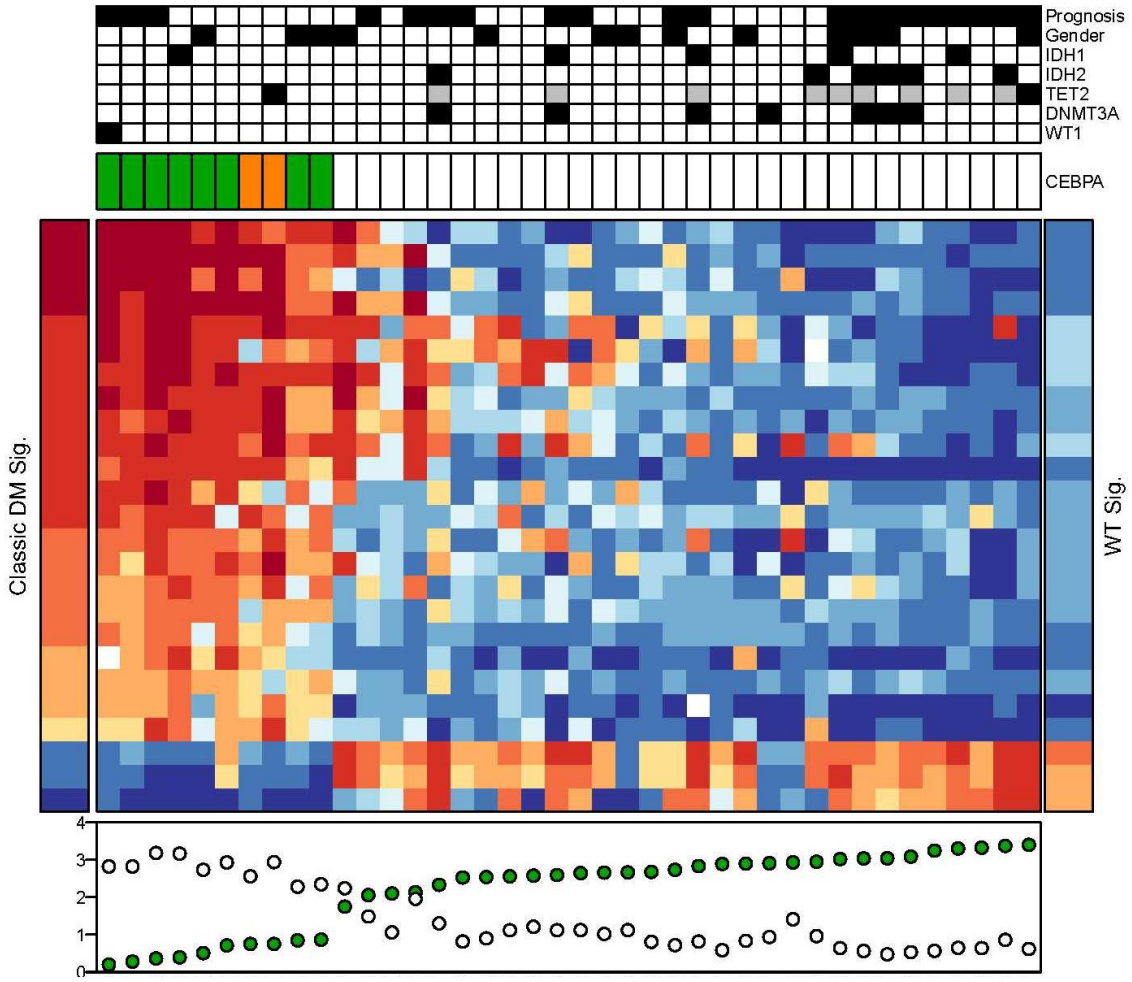


Figure 3.12 Heatmap showing methylation values for each sample at the 25 CpG sites in the *CEBPA* methylation signature. Each column represents a different patient. Patient characteristics and genotype are given at the top of the diagram (key as in Figure 3.9). Methylation levels are as in Figure 3.11. The classic CEBPA^{DM} signature is shown to the left of the heatmap, and the CEBPA^{WT} signature to the right. Samples are ordered according to which signature they are most alike. The bottom panel shows how alike each patient sample is to the two signatures. The green circles represent the similarity of the methylation levels of the sample to the CEBPA^{DM} signature, and the white circles to the CEBPA^{WT} signature. The lower the y axis position of the circle, the more closely the patient sample matches that particular signature.

and also the difference between the two scores, a cluster of 10 samples was identified that consisted of all 10 *CEBPA*^{DM} samples and no *CEBPA*^{WT} samples (Figure 3.13). This clustering did not correlate with prognosis as three out of the ten samples were chemoresistant.

3.3.4 Validation of results in two further cohorts of samples

A further two cohorts of samples from a total of 96 patients were selected and analysed as previously except using the Illumina Infinium Methylation 450K BeadChip array. For cohort 2, 48 samples were chosen with the same criteria as the first cohort, namely NK, WT for *NPM1*, *FLT3*/ITD and *FLT3*/TKD. However, following on from the results of the first cohort, samples with known *CEBPA*^{MUT} genotype were targeted, and 16 were *CEBPA*^{DM}, 8 *CEBPA*^{SM} and 24 *CEBPA*^{WT} (Table 3.3 and Appendix Table 2). Overall, 20 were known to be chemosensitive and 18 chemoresistant. Outcome was not available in 10 cases (3 *CEBPA*^{DM}, 7 *CEBPA*^{SM}). Cohort 3 consisted of samples from 48 patients that were specifically selected to analyse the profiles of different types of *CEBPA* mutations, in particular non-classic mutations. One sample was subsequently excluded from analysis due to low signal to noise ratios for all CpG sites. Of the remaining 47 samples, 15 were *CEBPA*^{DM}, of which 11 were non-classic DM; 30 were *CEBPA*^{SM}, of which 21 were non-classic, seven were classic N and two were classic C mutations; and two were *CEBPA*^{WT} (Table 3.3 and Appendix Table 2). It was not possible to select samples in the latter cohort that were all NK and WT for *NPM1* and *FLT3* and only eight patients fulfilled all these criteria. Of the other 39 patients, 14 were NK, nine had an abnormal karyotype of intermediate prognostic significance, five had an adverse karyotype, and karyotype was unknown in 11; 16 patients were WT for all three mutations, 14 had an *NPM1* mutation, nine a *FLT3*/ITD and seven a *FLT3*/TKD. Overall therefore in these two cohorts, 31 samples (33%) were *CEBPA*^{DM}, 17 (18%) with classic mutations and 14 (15%) with non-classic mutations, 38 (40%) were *CEBPA*^{SM} and 26 (27%) *CEBPA*^{WT}.

In view of the increased depth of coverage for each gene analysed in cohorts 2 and 3 using the 450K HumanMethylation BeadChip, all CpG sites relating to four of the 20 genes in the *CEBPA* methylation signature were examined to

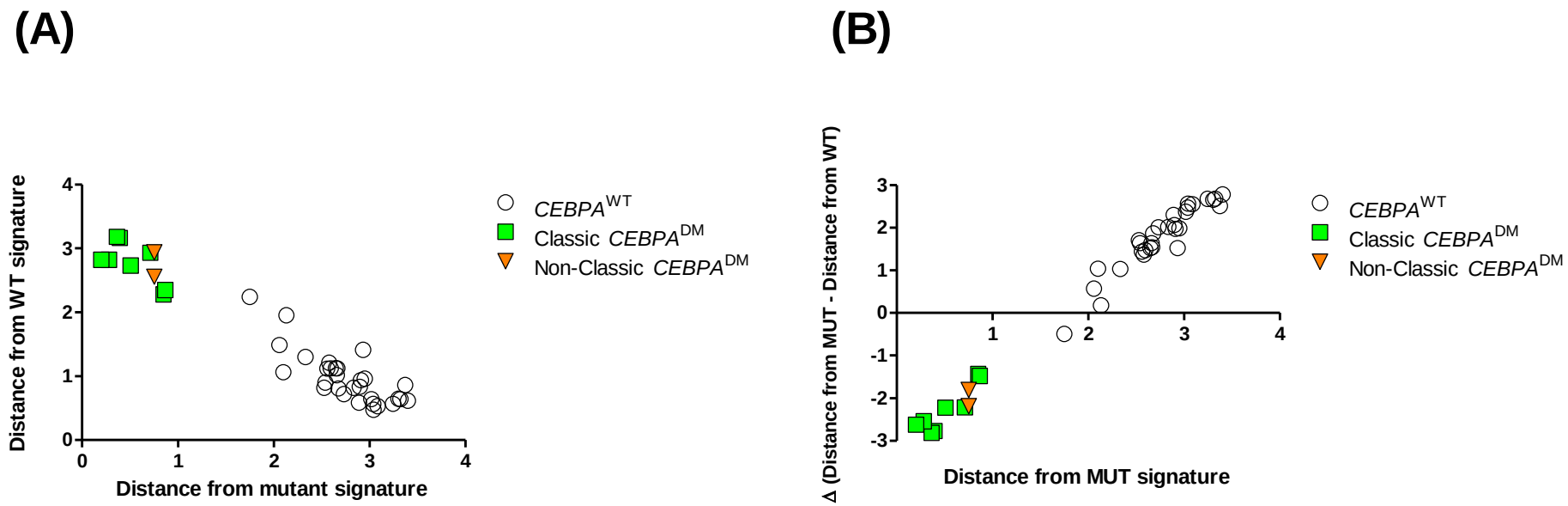


Figure 3.13 Distance from the classic *CEBPA*^{DM} and *CEBPA*^{WT} signatures for the cohort 1 samples. (A) Comparison of the distance from each of the two signatures. (B) Distance from the classic *CEBPA*^{DM} signature compared to the difference between the distance from each signature.

check that differential methylation was maintained at the specific CpGs used in the signature, and to assess whether neighbouring sites also showed the same variation. The sites selected, *GNMT*, *KHNYN*, *VAMP5* and *LY9*, all showed a marked difference in β values between *CEBPA*^{DM} and *CEBPA*^{WT} samples in cohorts 2 and 3, which was consistent with the results from cohort 1. Three were more hypermethylated (*GNMT*, *KHNYN* and *VAMP5)* and *LY9* was more hypomethylated in the *CEBPA*^{DM} samples (Figure 3.14). They were associated with 36, 17, 11 and 21 CpG sites respectively on the 450K array.

A heatmap was created to show the methylation levels of the validation cohorts at all CpG sites associated with the four genes (Figure 3.15). Of the 36 CpG sites associated with *GNMT*, 27 were differentially methylated (as defined in Section 3.3.3). Of these, 25 showed a difference in median β value of >0.2 between classic *CEBPA*^{DM} and *CEBPA*^{WT} samples, and 12 were >0.5. Similarly, nine of the 17 CpG sites related to *KHNYN* were variably methylated, but only one site which was included in the signature showed a difference in median β value between classic *CEBPA*^{DM} and *CEBPA*^{WT} samples of 0.4. The *LY9* gene had five differentially methylated CpG sites, two of which showed a difference in mean β value between classic *CEBPA*^{DM} and *CEBPA*^{WT}. *VAMP5* had ten variably methylated CpG sites, of which four showed a difference in mean β value of >0.2. Thus the variability in methylation seen between *CEBPA*^{DM} and *CEBPA*^{WT} samples was restricted to small regions associated with particular genes rather than the whole gene, in three of the four cases this was within or on the edge of a CGI, for *LY9*, where the closest CGI is approximately 5000 bases downstream of the TSS, the differentially methylated CpGs were close to the TSS.

Unsupervised cluster analysis was performed on the two follow-up cohorts using the same method as for cohort 1. The majority of samples in cohort 2 had been selected based on their outcome. When assessing whether unsupervised cluster analysis of this cohort separated samples based on prognosis, for the 38 samples with known outcome, 12 of the 20 chemosensitive samples clustered together in the high methylation cluster and none of the 18 chemoresistant samples were seen in this cluster. However, this was linked to *CEBPA* genotype as all 12 samples were *CEBPA*^{DM}. Of the remaining eight

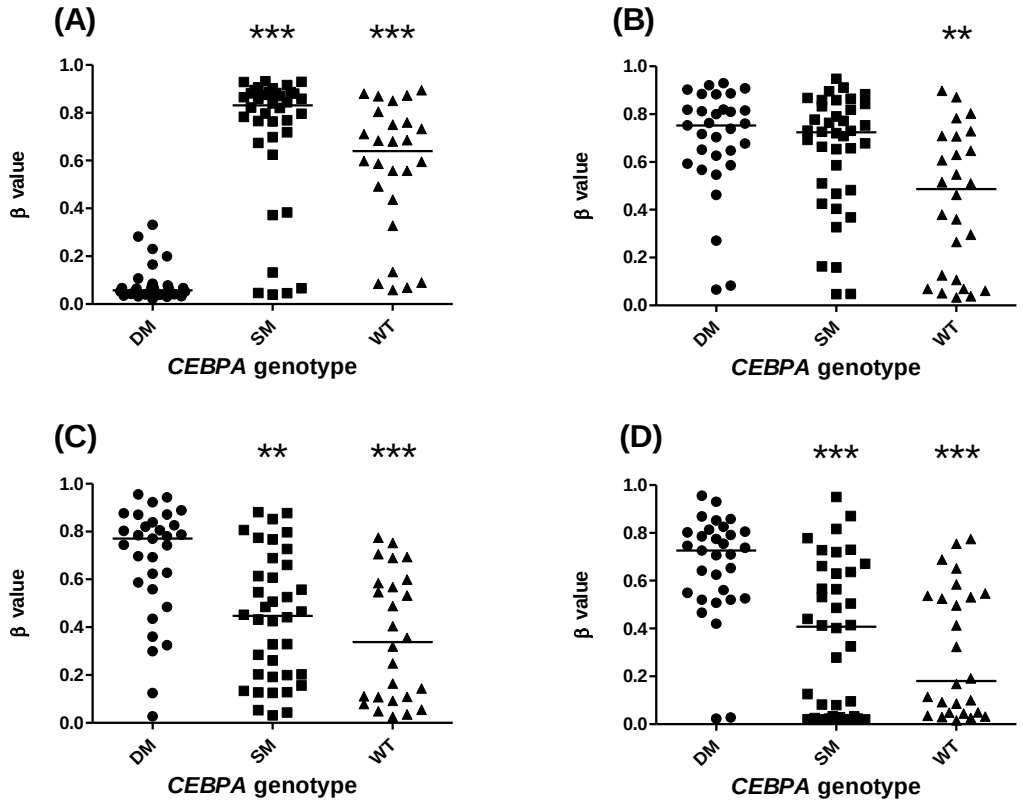


Figure 3.14 β values from patients studied in cohorts 2 and 3 with the 450K array at four of the sites in the CEBPA methylation signature. Results have been grouped according to *CEBPA* genotype. Samples highlighted in red are non-classic *CEBPA*^{DM}. Medians for each cohort are given and significance compared to *CEBPA*^{DM}, **P*<0.05; ***P*<0.01; ****P*<0.001. (A) LY9, (B) VAMP5, (C) KHNYN, (D) GNMT.

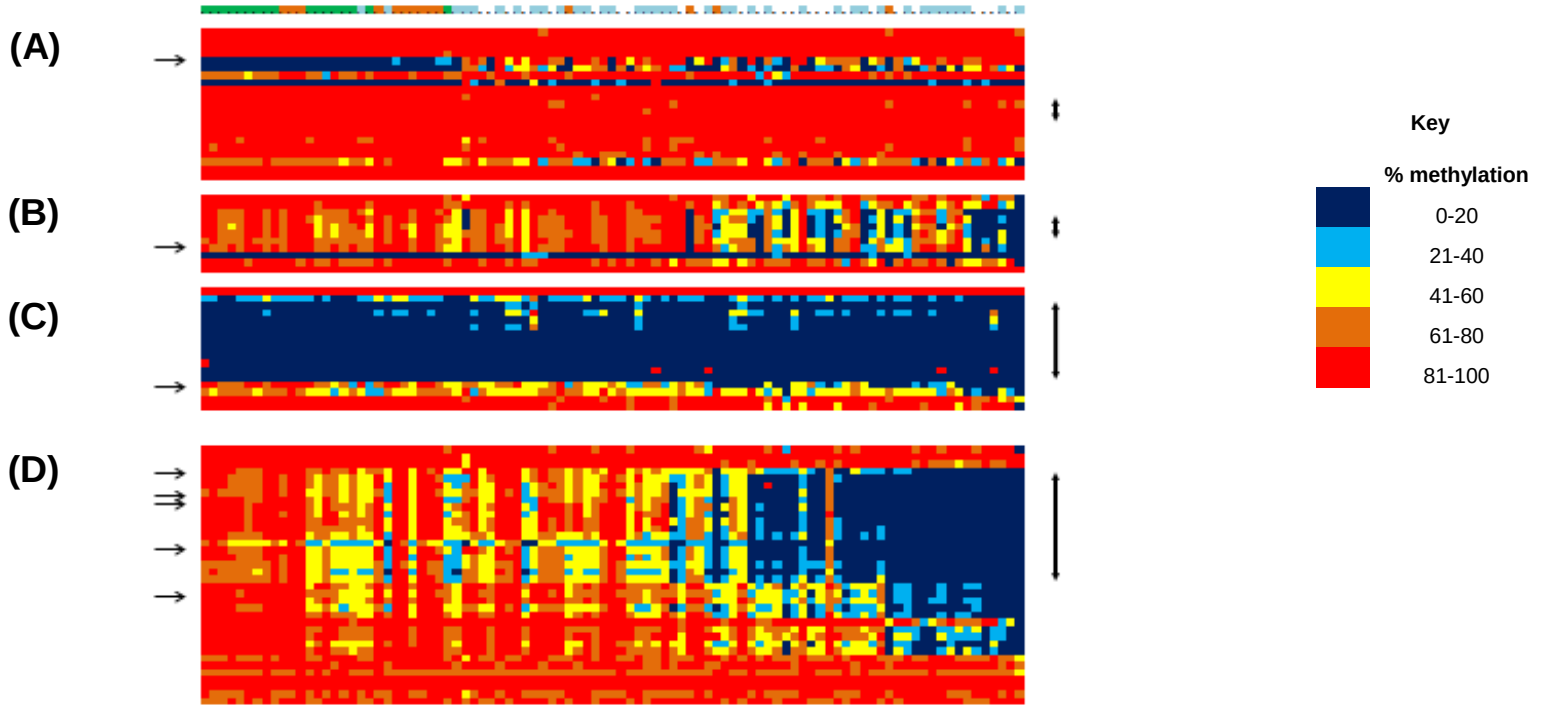


Figure 3.15 Methylation levels at all the CpG sites investigated on the 450K arrays that relate to 4 specific genes in the methylation signature. (A) LY9 (21 CpG sites), (B) VAMP5 (11 CpG sites), (C) KHNYN (17 CpG sites) and (D) GNMT (36 CpG sites). Each row represents a different CpG site and they have been arranged in sequential order. The arrows to the right of the diagram indicate the location of the CpG islands. The arrows on the left indicate the CpGs assayed on the 27K array that were included in the methylation signature. Patient samples (columns) are arranged in the same order as Figure 3.12 with CEBPA status indicated at the top, classic DM, ■; non-classic DM, ■; SM, ■ and WT, □.

chemosensitive samples, seven were *CEBPA*^{WT}. Further analyses therefore focussed on the *CEBPA* genotype, regardless of outcome. Overall, in both follow-up cohorts, 25 of the 31 *CEBPA*^{DM} samples analysed (81%) clustered together and, as before, they had a relatively hypermethylated profile compared to the other samples (Figure 3.16). Of the six *CEBPA*^{DM} samples that did not fall in this cluster, five were non-classic DM samples. The majority of the *CEBPA*^{SM} samples (31 of 38, 82%) clustered with the *CEBPA*^{WT} samples.

Given the confirmation from these follow-up cohorts that *CEBPA*^{DM} samples have a methylation profile that is distinct from non-DM samples, the 25 CpG site *CEBPA* methylation signature created from cohort 1 was assessed in these samples. As before, samples were ordered according to how close to the classic *CEBPA*^{DM} signature they were (Figure 3.17). Seventeen of these cases were classic DM. When the score for the distance to the mutant signature was plotted against the difference between the scores for the mutant and wild-type signatures, all classic *CEBPA*^{DM} except one formed a cluster equivalent to that observed in cohort 1 (Figure 3.18A). The remaining non-classic *CEBPA*^{DM} and *CEBPA*^{SM} cases will be considered further in chapter 4.

3.3.5 Definition of criteria for a classic *CEBPA*^{DM} methylation profile

In total, 25 cases in the three cohorts were classic *CEBPA*^{DM} and 56 were *CEBPA*^{WT}. From the cluster analysis (Figure 3.18A), one *CEBPA*^{DM} case in cohort 2 was clearly an outlier. Possible reasons for this are presented in chapter 4, and it was excluded from further analyses. Data from the remaining 24 classic *CEBPA*^{DM} cases were then combined and used to define criteria for a classic *CEBPA*^{DM} methylation profile. The mean score ± 2SD for the distance to the *CEBPA*^{DM} signature was 0.65 ± 0.44, and the mean difference between the distance to the classic *CEBPA*^{DM} and the *CEBPA*^{WT} signatures was -2.14 ±0.90. Together the upper limits of these scores were used as cut-offs to define a classic *CEBPA*^{DM} quadrant (Figure 3.18B). All classic *CEBPA*^{DM} cases fell in this quadrant, although one was borderline for the difference between the two distance scores. No *CEBPA*^{WT} samples were located within the *CEBPA*^{DM} quadrant. These criteria were then used to examine the non-classic *CEBPA*^{DM} samples and this data is presented in the next chapter.

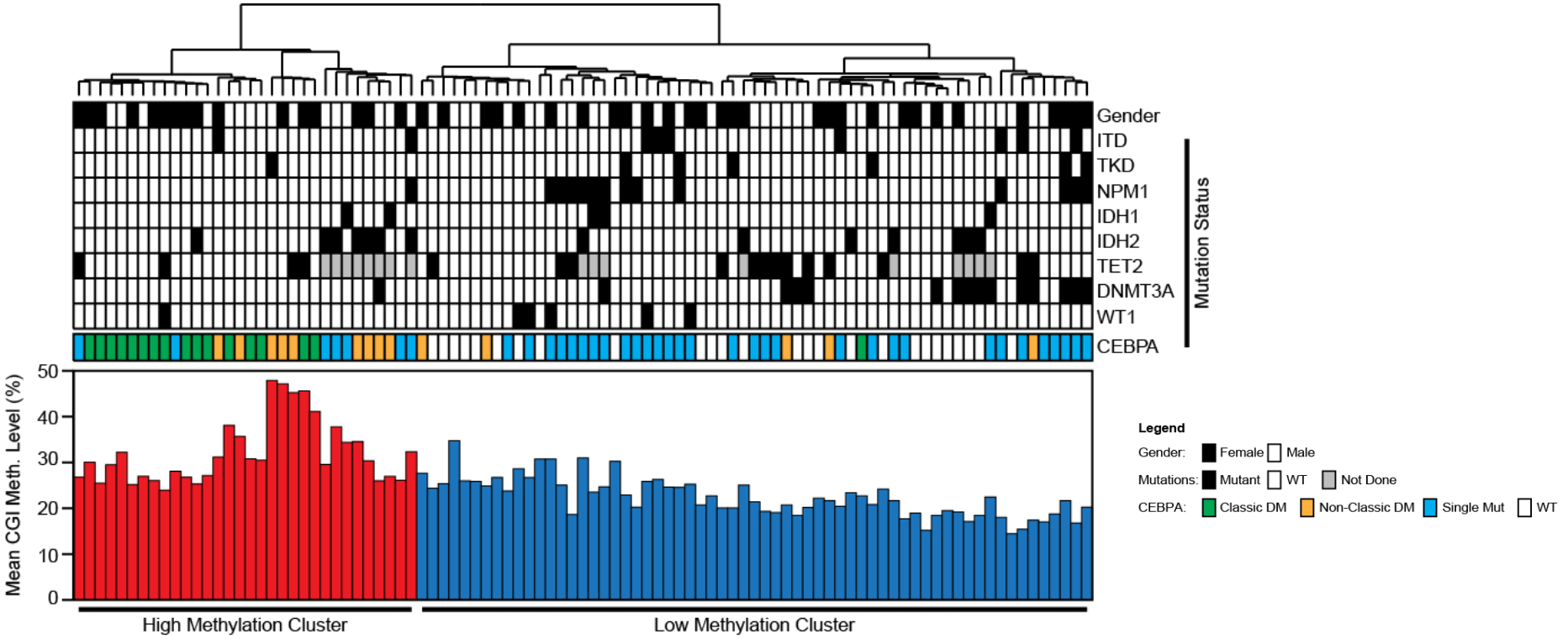


Figure 3.16 Unsupervised cluster analysis of cohorts 2 and 3. The top panel shows the clustering. The middle panel shows the karyotype and molecular status for the specified genes. The bottom panel displays the mean methylation level for all CpG sites assayed on the arrays and located within CpG islands for each patient.

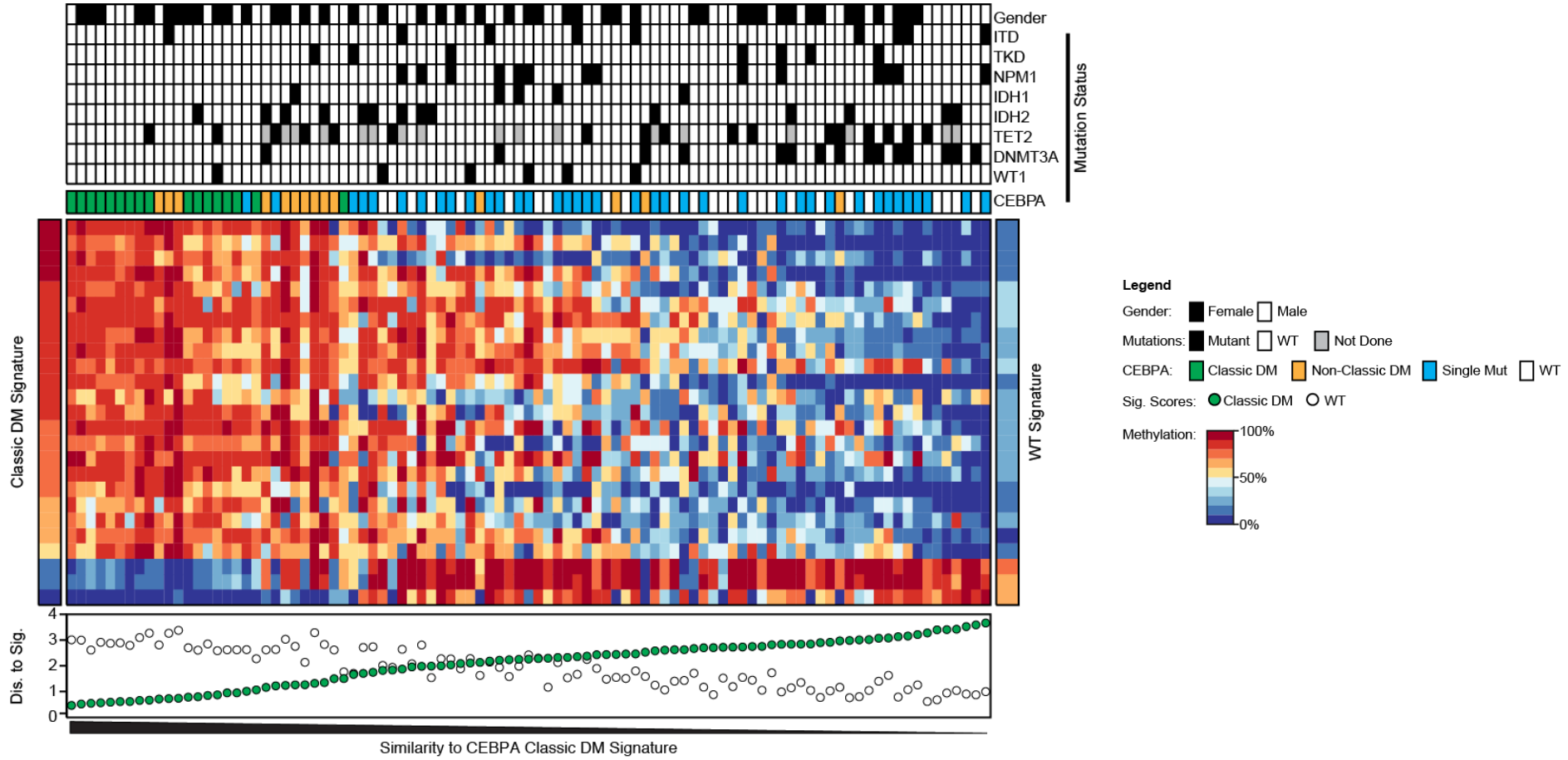


Figure 3.17 Heatmap showing methylation values for the CEBPA signature loci in the 95 patients analysed in the follow-up cohorts. The ordering of the samples and the distance to the signatures for each sample is plotted as given for Figure 3.12..

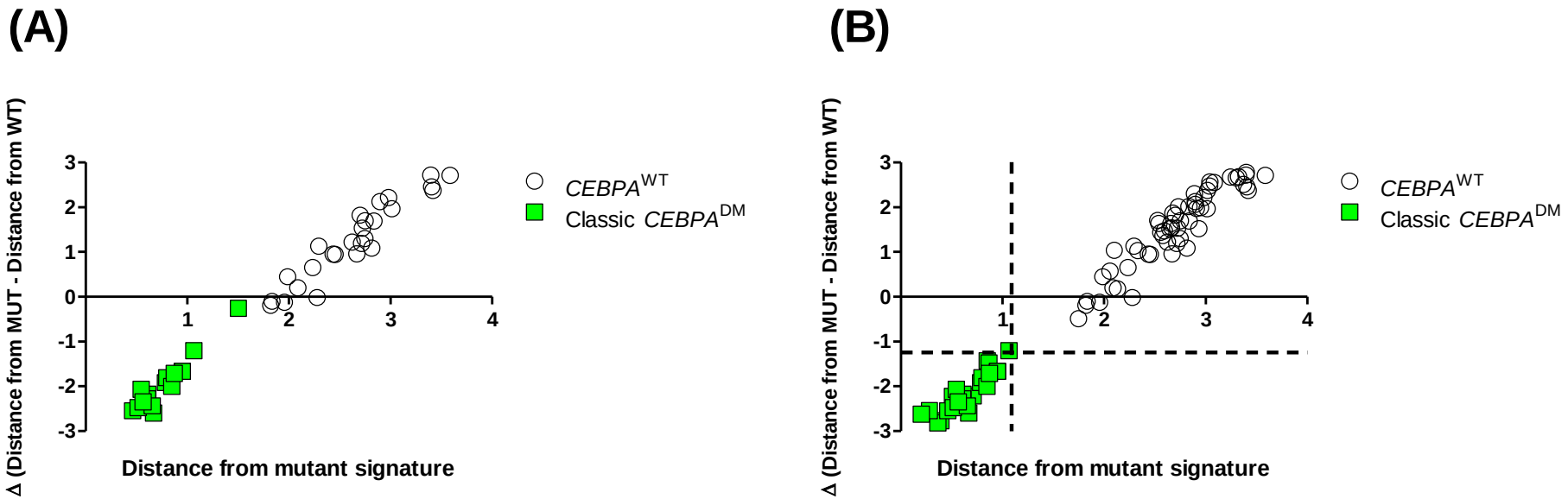


Figure 3.18 Comparison of the distance from the CEBPA^{MUT} signature with the difference in the distance from each of the signatures in classic CEBPA^{DM} cases. (A) Classic *CEBPA*^{DM} cases in the follow-up cohorts. (B) The 24 classic *CEBPA*^{DM} cases used to define a classic *CEBPA*^{DM} quadrant.

3.3.6 Correlation of methylation array results with other mutations

Although the majority of the samples in cohort 2 and all samples in cohort 3 were selected based on their *CEBPA* genotype, clustering associated with other mutations was also examined. There was a suggestion that *IDH*1/2-mutated samples clustered together. Within cohorts 2 and 3, eight of the 18 *IDH*^{MUT} samples were located in one cluster within the hypermethylated group, of which four were non-classic *CEBPA*^{DM} and four were *CEBPA*^{SM} (Figure 3.16). Five of the 10 *IDH*^{MUT} samples that were within the hypomethylated group were *CEBPA*^{WT} and the other five were *CEBPA*^{SM}. Cohort 1 contained 11 *IDH*^{MUT} samples. One also had a *CEBPA*^{DM} and was in the hypermethylated group; the other 10 samples were all located within the hypomethylated group, and five of these clustered together (Figure 3.9). In cohorts 2 and 3, four of the 17 *TET2*^{MUT} samples were in the hypermethylated group, three associated with *CEBPA*^{DM} and one with *CEBPA*^{SM}. A *TET2* sub-group could be seen within the hypomethylated cluster which consisted of nine samples including six *TET2*^{MUT} and one *IDH2*^{MUT} sample. The remaining 11 *TET2*^{MUT} samples did not cluster based on their methylation profile. Similarly, the six *WT1*^{MUT} samples in cohorts 2 and 3 did not cluster.

Overall, there were 22 samples with *DNMT3A* mutations, 21 of which were within the hypomethylated cluster. In cohort 1, four of the seven *DNMT3A*^{MUT} were tightly clustered together although these samples also all had *IDH*^{MUT}. In the follow-up cohorts one subgroup of seven samples had five with *DNMT3A*^{MUT}. Overall, 14 samples with *NPM1* mutations were studied all of which also were *CEBPA*^{SM}; six clustered together within the hypomethylated group on unsupervised analysis.

3.3.7 Assessment of other samples within the good-risk prognostic category

To assess whether the methylation signature created was specific to *CEBPA*^{DM} or whether other mutations and cytogenetic alterations that correlate with good-risk prognosis also had a similar profile, the methylation level at three differentially methylated CpG sites was quantified in samples that had either inv(16) (n=21) or t(8;21) (n=19), or were *NPM1*^{MUT}*FLT3*^{WT} (n=42).

Pyrosequencing assays were created and titration curves prepared as described in section 1.2.3. Good correlations between the observed and expected results were obtained for all three sites ($r^2 \geq 0.96$) (Figure 3.19). PCRs were also performed on seven patient samples that had been investigated using the Illumina methylation arrays, to confirm the accuracy of the assays. These samples were selected as they were known to have a range of methylation values. The results of the pyrosequencing and arrays were highly comparable in all three assays ($r^2 \geq 0.98$) (Figure 3.20). The test samples were then bisulfite converted and analysed in duplicate using all three assays, and the mean of the replicates for each sample compared to the beta values multiplied by 100 for the 24 classic *CEBPA*^{DM} and 56 *CEBPA*^{WT} samples.

Samples from patients with core-binding factor leukaemias had similar results to the *CEBPA*^{DM} samples for *LY9* but were significantly different for *VAMP5* and *KHNYN* (Figure 3.21). The *NPM1*^{MUT}*FLT3*^{WT} samples were significantly different from the *CEBPA*^{DM} samples for *KHNYN* and *LY9*. These results indicate that the methylation levels observed in the *CEBPA*^{DM} samples were not due to a “good prognostic signature”. They also did not simply reflect reduced C/EBPα activity, which is a recognised feature of core-binding factor leukaemias (Helbling *et al*, 2005; Pabst *et al*, 2001a). The Euclidian distance between these methylation scores and the median for *CEBPA*^{DM} was calculated for each sample. All three subgroups were significantly different from *CEBPA*^{DM} (Figure 3.22).

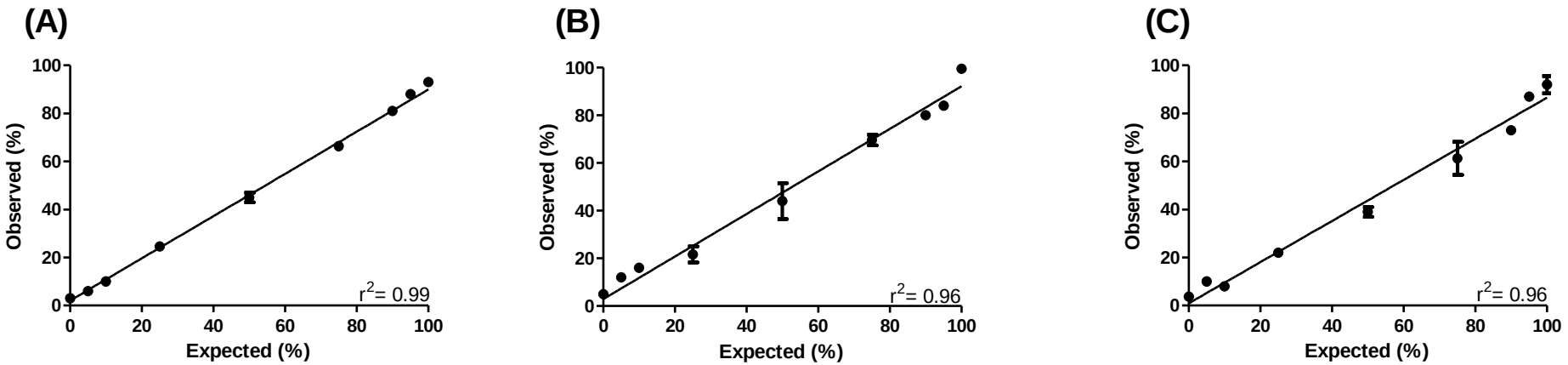


Figure 3.19 Titration curves for the percentage methylation obtained using standards in the pyrosequencing assays.

Results shown are the mean of duplicates. (A) LY9, (B) VAMP5, (C) KHNYN.

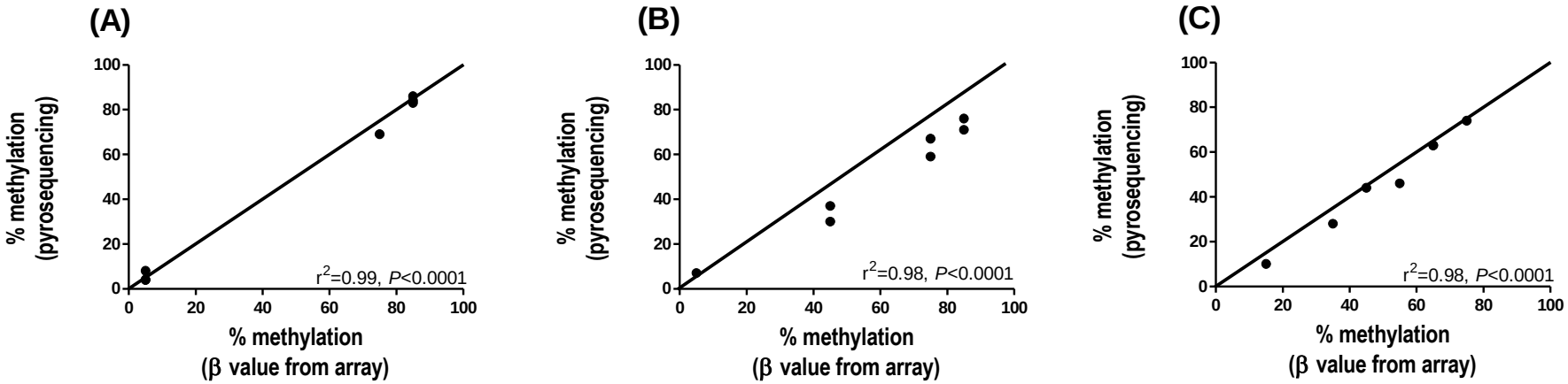


Figure 3.20 Comparison of the pyrosequencing results and array beta values for patient samples. The expected line for 100% concordance is indicated. (A) LY9, (B) VAMP5, (C) KHNYN.

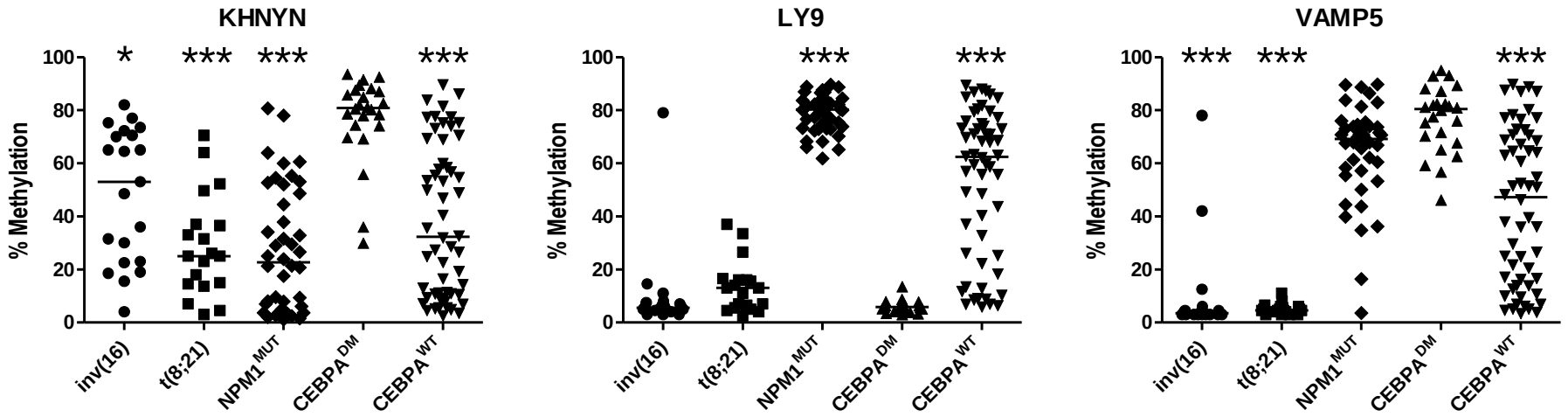


Figure 3.21 Pyrosequencing results of samples from good-risk prognostic patients at three of the 25 CpG sites in the *CEBPA*^{MUT} signature. Mean of duplicate results from 21 inv(16), 19 t(8;21) and 42 *NPM1*^{MUT}*FLT3*^{WT} patients. The results of the 56 *CEBPA*^{WT} and 24 classic *CEBPA*^{DM} samples are the beta values x100 from the methylation arrays. Medians for each cohort are given and significance compared to *CEBPA*^{DM}, **P*<0.05; ***P*<0.01; ****P*<0.001.

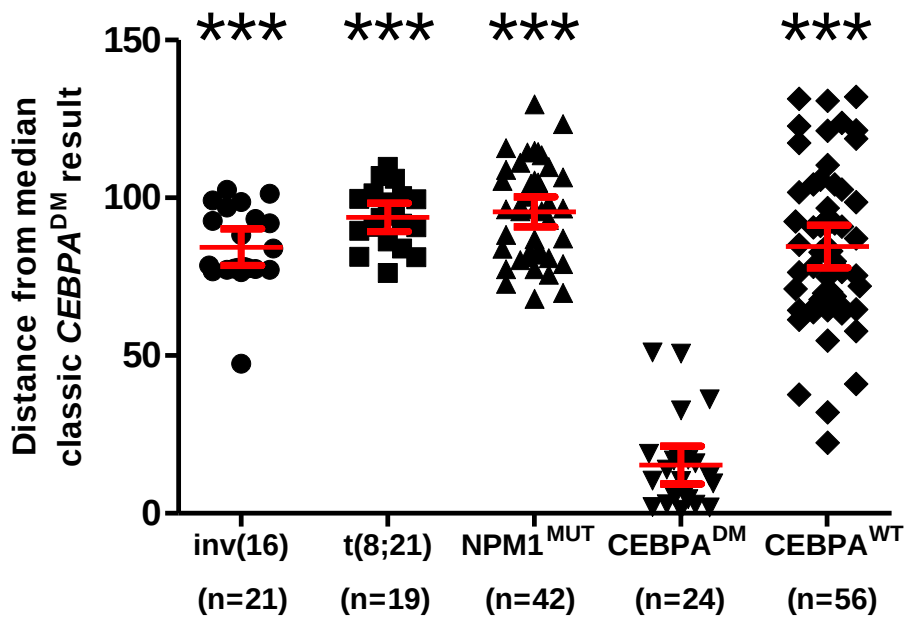


Figure 3.22 Euclidian distance of samples from patients in the good risk prognostic groups from the median profile of *CEBPA*^{DM}. The distances have been calculated from the results shown in Figure 3.21. The mean and 95% confidence interval for each group is shown and significance is compared to *CEBPA*^{DM}.

3.4 Discussion

Specific cytogenetic abnormalities and mutations in AML patients correlate with particular prognostic groups that have marked differences in predicted 5-year survival rate ranging from approximately 5% to 60% (Grimwade *et al*, 2010). This has led to changes in management of these patients based on their prognostic category. However, approximately 50% of newly diagnosed patients fall into the intermediate prognostic category, thus there is a need to find new features at diagnosis that will aid in stratifying this cohort further. Previously published studies have assessed the methylome in large numbers of unselected patients and have found that those samples with the same karyotype or certain mutations in genes such as *CEBPA* and *NPM1* have similar methylation patterns. Only two groups have created signatures predicting prognosis based on methylation status of defined CpG sites (Deneberg *et al*, 2010; Figueroa *et al*, 2010b). However, there was no overlap in the CpG sites analysed in the predictors developed by these groups, which probably reflects the different methodologies used as well as the different cohorts assessed. In the studies presented here, the approach used was to assess whether prognosis was associated with the methylation patterns of 40 highly selected AML patients that were at the extremes of clinical outcome, either chemosensitive or chemoresistant, and lacked common cytogenetic or molecular abnormalities that were associated with prognosis, i.e. they all had an NK and were WT for *NPM1*, *FLT3*/ITD and *FLT3*/TKD. This approach was only possible due to the availability of over 1000 well-characterised samples in the departmental AML DNA Biobank, as only a small proportion of these patients fulfilled all the clinical and molecular criteria.

Pyrosequencing was used to verify the β values of four of the CpG sites that showed the greatest variability in methylation between the chemosensitive and chemoresistant patients on the array. There was a highly significant correlation between the methylation level as quantified by the Illumina methylation array and pyrosequencing at all four CpG sites, with correlation coefficients ranging between 0.76 and 0.93 ($p<0.0001$ for all sites). Thus the percentage of methylated alleles at a particular site, as measured by the Illumina methylation

array, was reproducible using another method of analysis. However, both techniques used bisulfite conversion, and so would not pick up any biases in the conversion process. Furthermore, neither would be able to differentiate between an unmethylated C that had been converted to a T during bisulfite conversion and C/T SNPs, which theoretically could be an issue as 5-methylcytosine conversion to thymine is the most common evolutionary mutation seen on a genome-wide level (Bird, 1980). It has been reported that this may affect up to 8% of CpG sites analysed (Morris & Lowe, 2012).

By unsupervised cluster analysis of methylation levels of CpGs within CGIs, the 40 samples included in the initial cohort could be divided into two groups, one with a hypermethylated phenotype and the other a hypomethylated phenotype. There was no evidence that these cohorts correlated with prognosis, as 38% of the hypermethylated group and 58% of the hypomethylated group were chemoresistant. However, when the cluster analysis was correlated with the mutant status of other recurrently mutated genes that had been investigated since the project started, there was a striking association between *CEBPA*^{MUT} and the hypermethylated group. All ten *CEBPA*^{MUT} samples in cohort 1 were in this group, and all these samples had double rather than single mutations. The incidence of *CEBPA*^{MUT} in this cohort (25%) was higher than the incidence seen in published studies (9%), but this was due to the selection criteria used given that an NK, WT *NPM1* and *FLT3*, and a favourable outcome are all known to be associated with *CEBPA*^{DM} (Green *et al*, 2010b). Of note however, three of the 10 *CEBPA*^{DM} samples (30%) were from patients in the chemoresistant group, hence the methylation profile did not correlate with outcome despite the known association between *CEBPA*^{DM} and prognosis.

A second cohort of samples was chosen to validate these results suggesting that prognosis per se did not correlate with methylation profile but *CEBPA*^{MUT} status did, and to further assess whether the difference was restricted to *CEBPA*^{DM} or related to all *CEBPA*^{MUT}. A third cohort was subsequently chosen based on the various types of *CEBPA* mutations to assess whether *CEBPA*^{DM} with classic and non-classic mutations had different methylation profiles. Both of these cohorts were analysed using the 450K BeadChip array as this had superseded the 27K BeadChip array. As the Infinium II assay used for the 450K

BeadChip has been shown to lead to some bias and to be less sensitive to either extreme of methylation (Dedeurwaerder *et al*, 2011), a statistical normalisation was performed by Dr Sproul to account for this within-batch variation for cohorts 2 and 3.

Unsupervised cluster analysis of cohort 2 confirmed that there was no association between overall prognosis and methylation profile. Only 12 of the 20 chemosensitive samples (60%) clustered together and all of them were *CEBPA*^{DM}. Further analysis therefore focussed on *CEBPA* genotype. Analysis of cohorts 2 and 3 confirmed that *CEBPA*^{DM} had a distinctive methylation profile, with 81% of the *CEBPA*^{DM} samples in the hypermethylated cluster. Only seven of the 38 *CEBPA*^{SM} (18%) and none of the 26 *CEBPA*^{WT} samples were in the hypermethylated group. These results are similar to those in the study published by Figueroa *et al* (2010b) where 14 of the 24 patients with *CEBPA*^{DM} (58%) were in a unique cluster with a hypermethylated profile. They also found a second cluster of nine patients, five *CEBPA*^{DM}, two *CEBPA*^{SM} and two *CEBPA*^{WT}, which had a predominantly hypomethylated phenotype. Conversely, using the same Illumina methylation array as the one used in the studies presented here, Deneberg *et al* (2011) did not find an association between *CEBPA*^{MUT} and methylation profile, but only six of the 118 patients they analysed had *CEBPA*^{MUT}, and they did not state whether they were single or double mutations. This clustering is supported by data from published gene expression arrays, which also show that *CEBPA*^{DM} samples form unique clusters (Grossmann *et al*, 2013; Taskesen *et al*, 2011; Valk *et al*, 2004; van Vliet *et al*, 2013; Wouters *et al*, 2009). Together with other factors, including for example the inverse correlation with *NPM1* mutations and the improved overall survival seen in patients with *CEBPA*^{DM} compared to *CEBPA*^{WT} or *CEBPA*^{SM} (Dufour *et al*, 2010; Green *et al*, 2010b; Pabst *et al*, 2009; Wouters *et al*, 2009), this data provides further evidence that *CEBPA*^{DM} is a distinct biological entity.

Only limited conclusions can be drawn about the correlation of methylation profiles with other mutant genes as the selection of a high proportion of *CEBPA*^{MUT} samples would have influenced the cluster analysis. However, there was a suggestion that *IDH*1/2-mutated samples clustered together based on their methylation profile. Samples with *IDH*^{MUT} have been reported to have a

hypermethylated profile (Deneberg *et al*, 2010; Figueroa *et al*, 2010a), although in the analysis presented here, only ten of the 30 *IDH*^{MUT} clustered in the hypermethylated group. A hypomethylated cluster of nine samples included one *IDH2*^{MUT} and six *TET2*^{MUT} samples, suggesting that these mutations, which are known to be mutually exclusive, can be associated with similar methylation profiles. Mutations in *DNMT3A* are known to be inversely correlated with *CEBPA*^{DM}, and in these cohorts only three of the 22 *DNMT3A*^{MUT} samples also had *CEBPA*^{DM} (Gale *et al*, 2015). All bar one of the samples with *DNMT3A*^{MUT} clustered within the hypomethylated group. *DNMT3A*^{MUT} have been associated with hypomethylation in other studies, which is consistent with the mutations being loss-of-function and dominant-negative, thus affecting the protein's ability to catalyse *de novo* methylation of CpG sites (2013; Hajkova *et al*, 2012; Qu *et al*, 2014; Russler-Germain *et al*, 2014). Overall, 14 samples with *NPM1* mutations were studied; six clustered together on unsupervised analysis, which has also been seen in other studies (2013; Figueroa *et al*, 2010b). Of note, all 14 also had a *CEBPA*^{SM}, which may have influenced the analysis. Together these studies suggest that the clustering of samples by their methylation profiles can be linked with the underlying molecular status of the samples, but there is still variability in these profiles, even in samples with the same mutations, which may partly be explained by co-incident mutations.

A *CEBPA* methylation signature was created based on the 25 most differentially methylated CpG sites between classic *CEBPA*^{DM} and *CEBPA*^{WT} samples investigated in the first cohort. None of the genes connected to these CpG sites were known to be associated with *CEBPA*, and none of them were included in the published methylation signatures that were related to prognosis (Bullinger *et al*, 2010; Figueroa *et al*, 2010b). The distance to the mutant signature and the difference between the distance to the wild type and the mutant signature for each sample was plotted. All ten *CEBPA*^{DM} and no *CEBPA*^{WT} samples were located within this cluster. The *CEBPA* methylation signature was then validated using 17 additional classic *CEBPA*^{DM} and 26 *CEBPA*^{WT} cases. Sixteen of the 17 classic *CEBPA*^{DM} had a methylation profile closest to the *CEBPA*^{MUT} signature, and conversely all 26 *CEBPA*^{WT} samples had a signature closest to the *CEBPA*^{WT} signature. Criteria to define a classic *CEBPA*^{DM} quadrant were then

derived from the distance scores for the 24 classic *CEBPA*^{DM} samples, excluding the outlier in the follow-up cohort.

Pyrosequencing assays of a subset of differentially methylated CpG sites were used to assess the methylation levels of samples from patients with other categories of good-risk AML. The results showed that samples from patients with core binding factor leukaemia had very similar methylation levels at two of the three CpG sites analysed, and from patients with *NPM1*^{MUT}/ *FLT3*^{WT} at one of the sites. However, the levels were significantly different between *CEBPA*^{DM} and the core-binding factor leukaemias at two of the three probes, and at two of the three probes when compared to *NPM1*^{MUT}/ *FLT3*^{WT} samples. This suggests that the methylation profile produced is not due to reduced C/EBPα activity alone or good-risk prognostic AML, but is specific to *CEBPA*^{DM} samples. This is in agreement with published data that shows that these subtypes of AML cluster separately based on their methylation profiles (Figueroa *et al*, 2010b).

In the next chapter, the methylation signature created that is associated with classic *CEBPA*^{DM} samples is explored further to examine whether non-classic *CEBPA*^{DM} and *CEBPA*^{SM} also have a similar methylation profile.

CHAPTER 4: INVESTIGATION OF NON-CLASSIC *CEBPA*^{DM} AND *CEBPA*^{SM} SAMPLES AND METHYLATION OF THE *CEBPA* PROMOTER

4.1 Introduction

As discussed in the previous chapter, approximately 75% of patients with a *CEBPA*^{DM} have a frameshift or nonsense mutation in the N terminus coupled with an in-frame insertion and/or deletion in the C terminus on the other allele. The N-terminal mutation occurs between amino acids 1-119, leading to increased translation from an internal ATG start site at amino acid 120 and production of a truncated p30 protein lacking the first transactivation domain (TAD) (Figure 3.1). The C-terminal mutations, occurring between amino acids 278-358 that encode the bZIP DNA binding domain and leucine zipper domain, are predicted to lead to a non-functional protein with impaired ability to bind DNA or dimerise, C-loss of function protein (C-LOF) (Figure 3.1). However, the remaining 25% non-classic *CEBPA*^{DM} are varied, with differing predicted consequences of the mutations. Approximately 5% of patients with *CEBPA*^{DM} have either a homozygous classic N mutation, predicted to lead to p30 protein only, or a homozygous classic C mutation, predicted to lead to C-LOF protein only (Table 3.2). Many non-classic mutations have also been reported, including missense mutations in the C terminus predicted to also give rise to C-LOF protein, and frameshift or nonsense mutations located after the first TAD that would lead to a truncated protein and are likely to be associated with nonsense-mediated decay and haploinsufficiency (Frischmeyer & Dietz, 1999). As discussed in section 3.1.3, functional work with non-classic mutations is limited, however there is a suggestion that they do not behave as classical mutations from gene expression studies and mouse models (Togami *et al*, 2015; Wouters *et al*, 2009).

The methylation array data presented in the previous chapter and published by other groups highlight that AML samples with particular mutations, e.g. *CEBPA*^{DM} or *DNMT3A*^{MUT}, often have similar DNA methylation profiles. This chapter explores whether the methylation profiles of non-classic *CEBPA*^{DM} and *CEBPA*^{SM} samples are similar to that of classic *CEBPA*^{DM}; why some *CEBPA*^{DM}

might have a methylation profile more like wild-type; and finally whether methylation of the *CEBPA* promoter could provide an explanation for some *CEBPA*^{SM} samples having a methylation profile like classic *CEBPA*^{DM}. Given that patients with *CEBPA*^{DM} lacking a *FLT3*/ITD are classified as good-risk and are therefore not usually considered for a stem cell transplant in first remission (Cornelissen *et al*, 2012), the methylation profile of non-classic *CEBPA*^{DM} samples may improve our understanding of the prognostic significance of these mutations and whether they should also be considered good-risk.

4.1.1 Allelic status of *CEBPA*^{DM} and *CEBPA* mutant level

The two mutations in *CEBPA*^{DM} samples are presumed to be biallelic by most groups, leading to a complete lack of WT allele. This is supported by competitive transplantation models of *cebpa* mutations in mice, as only mice receiving cells with two mutant alleles subsequently developed leukaemia (Bereshchenko *et al*, 2009). Only a few groups have investigated patient samples by cloning the entire *CEBPA* coding sequence and sequencing the clones to determine whether these mutations are indeed biallelic or if they are both located on the same allele. The distinction is an important one to make as, if the mutations are monoallelic, the impact would be predicted to be more akin to a *CEBPA*^{SM} than a *CEBPA*^{DM} genotype and this would have prognostic implications. Overall, seven studies have reported data from 87 *CEBPA*^{DM} patients that showed only three *CEBPA*^{DM} cases (3%) were likely to be monoallelic (Barjesteh van Waalwijk van Doorn-Khosrovani *et al*, 2003; Dufour *et al*, 2010; Frohling *et al*, 2004; Green *et al*, 2010b; Lin *et al*, 2005; Preudhomme *et al*, 2002; Shih *et al*, 2006).

Knowledge that the two mutations found in *CEBPA*^{DM} samples are located on different alleles does not, however, prove that both mutations are located within the same cell. In theory, they could represent two separate AML subclones containing different *CEBPA* mutations, which would require single cell analysis for confirmation. Quantifying the level of the mutants may, however, provide some indication as to whether the mutations are likely to be within the same cell as the level of the two mutations would be expected to be equivalent. Mutant level has also been associated with prognosis, for example, in the case of *FLT3*/ITD, a higher mutant level is associated with a worse overall prognosis

(Gale *et al*, 2008). Lastly, assuming that the mutation is in the dominant leukaemic clone, mutant levels may provide an estimate of the presence of non-leukaemic cells in a sample, which would be relevant to other analyses including the methylation profile. Only one paper provided details of mutant level of *CEBPA* in AML patients (Green *et al*, 2010b). The median level for 102 mutations was 44% (range 9-97%), and the level was at least 35% in 87% of the samples, suggesting that the mutation is likely to be acquired early in the disease pathogenesis. Furthermore, the mutant level of paired mutations in 26 $CEBPA^{DM}$ cases was highly correlated, suggesting that they were likely to be found within the same cell.

4.1.2 Promoter hypermethylation in AML

When aberrant methylation was detected in CGIs of cancer cells, initially it was thought that this could be an alternative mechanism for silencing genes and was coined an “epimutation” (Esteller, 2002; Herman & Baylin, 2003). Indeed, many of the genes that are recurrently mutated in AML are also found to have promoter hypermethylation in a subset of samples, e.g. *DNMT3A* and *CEBPA* (Fasan *et al*, 2013a; Jost *et al*, 2014). Furthermore, those samples with mutations and those with epimutations seem to be mutually exclusive (Shen & Laird, 2013). However, contrary to the hypothesis that methylation is causing the genes to be silenced, many of these genes, although unmethylated, are not expressed in the normal cell of origin (Sproul *et al*, 2012). This has led to the proposal that DNA methylation stabilises long-term repression rather than initiates it (Feldman *et al*, 2006), or that DNA methylation in promoter regions is a consequence of lack of transcription factor binding to the gene promoter rather than the cause of repression (Gebhard *et al*, 2010).

4.1.3 Promoter hypermethylation of *CEBPA*

Given the role that C/EBPα plays in myeloid differentiation, and the fact that its expression is reduced in certain sub-types of AML, Chim *et al* (2002) investigated whether this was due to aberrant hypermethylation of the *CEBPA* promoter. Analysing bisulfite-converted DNA from 70 AML samples using methylation-specific PCR, they found that two of the samples showed promoter methylation (2.8%). The primers used covered regions around the transcription

start site (TSS), the so-called “core promoter region”. Following this, it was noted that methylation of the *CEBPA* promoter also occurred in other malignancies such as head and neck, and lung cancer (Bennett *et al*, 2007; Tada *et al*, 2006). However, in these cases, the methylation was seen approximately 800-1000 bases upstream from the TSS, “the proximal and distal regions”. This prompted several groups to examine methylation of the *CEBPA* promoter in these regions in AML, as well as the core region (Table 4.1). In these studies, the frequency of methylation in the core region in AML ranged from 1-16% and in the distal region from 13-51%. The variable frequencies seen by the different groups may partly reflect the different methods used to assess methylation levels, and also how methylation or hypermethylation has been defined, as there is no set standard.

The published data on whether methylation of the *CEBPA* promoter is associated with *CEBPA* silencing in AML varies. Although APML is associated with lower levels of *CEBPA* expression, Santana-Lemos *et al* (2011) found no direct correlation between gene expression and aberrant methylation of the core or distal promoter in these samples. Hackanson *et al* (2008) also found no association between *CEBPA* gene expression and distal region methylation in unselected AML samples. However, some studies have reported a link between mRNA expression and core promoter methylation (Hollink *et al*, 2011; Szankasi *et al*, 2011), and others a link with distal methylation (Fasan *et al*, 2013a; Lin *et al*, 2011; Musialik *et al*, 2014).

When the gene expression profile of 285 unselected AML samples was analysed using an unsupervised cluster approach, two distinct clusters were identified that were comprised primarily of *CEBPA*^{MUT} cases but in one of these clusters, six of the 15 samples did not have mutations in *CEBPA* (Wouters *et al*, 2007). Further analysis of the data revealed that they did, however, have minimal or absent *CEBPA* expression. Bisulfite sequencing showed that of these six samples, four had methylation in the core region of the *CEBPA* promoter and the authors suggested that methylation may indeed have a similar impact on gene expression profile of AML samples as mutations in the gene. This group of *CEBPA*-silenced samples also had abnormal expression of T cell markers and were associated with activating mutations in *NOTCH*, a gene that encodes a transmembrane receptor.

Table 4.1 Frequency of *CEBPA* promoter methylation reported in diagnostic AML samples

Reference	Total no. of patients	No. with *CEBPA* methylation[1] (%)	Method of analysis	Region of analysis[2]	Prognostic impact of *CEBPA* methylation	Association of *CEBPA* methylation with other molecular markers
Chim *et al* (2002)	70	2 (3)	MSP	Core	Not assessed	
Agrawal *et al* (2007)	81	9 (11)	MALDI TOF	Proximal and core	Not assessed	
Wouters *et al* (2007)	285	4 (1)	Bisulfite sequencing	Proximal	Not assessed	
Hackanson *et al* (2008)	39	20 (51)	COBRA/ bisulfite sequencing	Distal, proximal and core	Not assessed	More frequent in inv(16) and t(15;17) cytogenetic subgroups
Jost *et al* (2009)	80	10 (13)	MSP/ bisulfite sequencing	Core	Not assessed	
Griffiths *et al* (2010)	169	27 (16)	MSP	Core	Not assessed	No association with cytogenetic risk group or *FLT3* or *NPM1* status
Lu *et al* (2010)	53	7 (13)	MSP	Core	Not assessed	Inverse association with *FLT3*/ITD and *NPM1*^{MUT} *CEBPA*^{METH} and *CEBPA*^{MUT} mutually exclusive
Szankasi *et al* (2011)	102	5 (5)	Pyrosequencing	Core	Not assessed	*CEBPA*^{METH} and *CEBPA*^{MUT} mutually exclusive All 5 samples had CD7 expression
Hollink *et al* (2011)	237[2]	3 (1)	MSP	Core	Not assessed	
Lin *et al* (2011)	193	28 (15)	MALDI TOF	Distal, proximal and core	Higher methylation associated with better OS	Mutual exclusion with *NPM1*^{MUT}
Santana-Lemos *et al* (2011)	39[3]	17 (13)	MSP/ bisulfite sequencing	Distal and core	Not assessed	

Table 4.1 continued

Reference	Total no. of patients	No. with *CEBPA* methylation[1] (%)	Method of analysis	Region of analysis[2]	Prognostic impact of *CEBPA* methylation	Association of *CEBPA* methylation with other molecular markers
Fasan *et al* (2013a)	623	238 (38)	MSP/ bisulfite sequencing	Distal, proximal and core	No association between methylation and prognosis	Inverse association with $NPM1^{MUT}$ $CEBPA^{METH}$ and $CEBPA^{MUT}$ mutually exclusive
Musialik *et al* (2014)	76	28 (37)	qMSP	Distal, proximal and core	Not assessed	Associated with good cytogenetic risk group

[1]The definition of *CEBPA* methylation varied between papers. [2]Paediatric AML samples analysed only. [3]Acute promyelocytic leukaemia samples analysed only.

COBRA, combined bisulfite restriction analysis; MALDI TOF, matrix-assisted laser desorption/ ionisation- time of flight; METH, methylated promoter as defined by authors; MSP, Methylation-specific PCR; MUT, mutant gene; OS, overall survival; qMSP, quantitative methylation-specific PCR

Comparing the methylome of the *CEBPA*-silenced cases with that of *CEBPA*^{MUT} showed that the two groups were epigenetically distinct, with the silenced cases showing marked hypermethylation at over 90% of the variable sites compared to the mutant cases (Figueroa *et al*, 2009b). Furthermore, there were biological differences between the two groups, with the *CEBPA*-silenced patients exhibiting significantly worse survival.

The aims of this chapter were to further characterise the *CEBPA*^{DM} samples, including assessing the allelic status and the level of the mutations, and then to use the *CEBPA* methylation signature to examine the non-classic *CEBPA*^{DM} and *CEBPA*^{SM} samples. Finally, methylation of the *CEBPA* promoter itself was analysed.

4.2 Materials and Methods

4.2.1 Allelic status of *CEBPA*^{DM} samples

To assess the allelic status of *CEBPA*^{DM} samples, PCR products spanning the entire coding region were prepared, cloned, and the clones then analysed to determine whether they contained one, both or neither of the mutations. Amplicons were prepared using primers *CEBPA* 1F and 3R (Appendix Table 1) with standard BIOTAQ polymerase mix plus 5% DMSO (Chapter 2, section 2.1.2). The time for each denaturation, annealing and extension step was extended to 60, 60 and 90 seconds respectively, and the final extension step was 10 minutes. Once the presence of products had been confirmed on an agarose gel, aliquots of each product were cloned as described in Chapter 2 (section 2.1.9). At least 20 colonies per sample were picked, seeded into 96 well plates and incubated overnight at 37°C.

Each clone was assessed with the two relevant PCRs to examine whether they had one, both or neither of the known mutations for each sample. The method of mutation detection varied depending on the type of mutation being analysed (Table 4.2). If the mutation resulted in a size change of ≤2 nucleotides, a restriction digest was performed to differentiate mutant from WT clones. Standard BIOTAQ PCRs with 5% DMSO were performed with 1µl of bacterial culture. An initial hotstart at 95°C for 5 minutes preceded all reactions to ensure

Table 4.2 Methods used to differentiate between WT and mutant alleles in *CEBPA* mutation quantification and clone characterisation

Patient No.	Mutation	Forward Primer[1]	Reverse Primer[1]	Method used post PCR	Expected size of WT fragment(s)	Expected size of mutant fragment(s)
27	334_335insGC	1F*	1R3	RED AscI	219+329*	550*
	912_913InsTTG	3F*	3R	CEQ	424*	427*
53	113 delG	1F*	1R3	RED Eco109I	37+54+141+142*+174	37+54+141+315*
	938_939insTA	3F*	3R	CEQ	424*	426*
63	198_201dup	1F*	1R3	CEQ	548*	552*
	890G>C	3F*	3R	RED FspI	84+142*+198	198+226*
76	232delC	1F*	1R3	RED BstNI	103+183+262*	103+444*
	899G>C	3F*	3R	RED BanII	126+298*	126+147+151*

[1]Primers as detailed in Appendix 1. *Fluorescently labelled primers and fragments

PCR, polymerase chain reaction; RED, restriction enzyme digest.

lysis of the bacteria. The presence of products was confirmed by agarose gel and then 6µl of each product was digested with the appropriate enzyme. Presence or absence of a mutation in a particular colony was assessed by running the digested products on an agarose gel with a DNA ladder to assess the length of the digested fragments. If the mutation resulted in a size change of ≥3 nucleotides, or there was no restriction enzyme digest readily available for mutations with a size change of 1-2 bases, then clones were amplified with BIOTAQ PCRs using a fluorescently labelled primer and the products were analysed by size separation on the CEQ (section 2.1.7). To ensure accurate fragment sizing, each product was run twice, once unmixed and once mixed with known WT amplicons. Thus if two peaks were seen in the mixed run, and the size difference was as expected, then the clone carried the mutation under investigation.

4.2.2 Quantification of *CEBPA* mutant level

PCRs covering the mutations were performed using a fluorescently labelled primer. They were analysed by size separation on the CEQ as detailed in chapter 2 (section 2.1.7) and above for allelic status determination, with restriction enzyme digestion where required, but without mixing with WT amplicons. The area under the wild-type and mutant peaks was assessed by the instrument software and used to calculate the relative level of mutant alleles as a percentage of total alleles.

4.2.3 Analysis of *CEBPA* promoter methylation

Methylation levels in the *CEBPA* promoter region were initially analysed using data from the 27K and 450K arrays. To examine a greater number of CpG sites in this region, three PCRs were designed for bisulfite sequencing. The regions chosen were based on previously published data, namely, the core and distal regions (Lin *et al*, 2011). Primers were designed using the PyroMark Assay Design Software 2.0 (Qiagen, Germany) (Appendix Table 1). Amplicons were prepared using GoTaq DNA polymerase with a standard reaction mix (section 2.2.4). Products were purified and sent for DNA sequencing. Methylation levels were estimated by comparing the peak height of the “C” nucleotide in a CpG site over the total peak heights of the “C” and “T” nucleotides.

Fresh aliquots of bisulfite-converted DNA were prepared for all 135 samples analysed on the arrays. Samples from each batch were checked for conversion efficiency as detailed in Chapter 2 (section 2.2.3). Each bisulfite-converted sample was then subjected to three PCRs to examine the core region, and the distal region in both the forward and reverse direction. Products were sent for DNA sequencing and the methylation levels estimated as outlined above.

4.3 Results

4.3.1 Further investigation of the *CEBPA*^{DM} samples

In total, 31 *CEBPA*^{DM} cases were investigated in the two validation cohorts. When these cases were evaluated using the parameters for the classic *CEBPA*^{DM} quadrant, as defined in the previous chapter, 12 (39%) did not satisfy the criteria and one case was borderline (Figure 4.1). Only one of the latter cases was a classic *CEBPA*^{DM}.

4.3.1.1 Allelic status of the *CEBPA*^{DM} samples

Six of the samples investigated (all classic *CEBPA*^{DM}) that were within the classic *CEBPA*^{DM} quadrant had previously been shown to be biallelic (Green *et al*, 2010b). As it is known that a small proportion of *CEBPA*^{DM} samples may be monoallelic, and thus may behave more like *CEBPA*^{SM}, the distribution of the two mutations in some of the 12 *CEBPA*^{DM} that were not located in the classic *CEBPA*^{DM} quadrant was investigated. Two were homozygous C-terminal mutations and thus had to be biallelic. Full-length amplicons were cloned from one classic *CEBPA*^{DM} and five non-classic cases. No full-length PCR product could be obtained for one of these samples. At least 19 full-length *CEBPA* clones were analysed for the remaining five samples (range, 19-37). In three samples all clones had just one of the mutations, and in two samples, just one clone of the 24 and 19 clones contained both mutations, indicating that all five samples were biallelic, including the classic *CEBPA*^{DM} case that did not fall in the classic *CEBPA*^{DM} quadrant (Table 4.3).

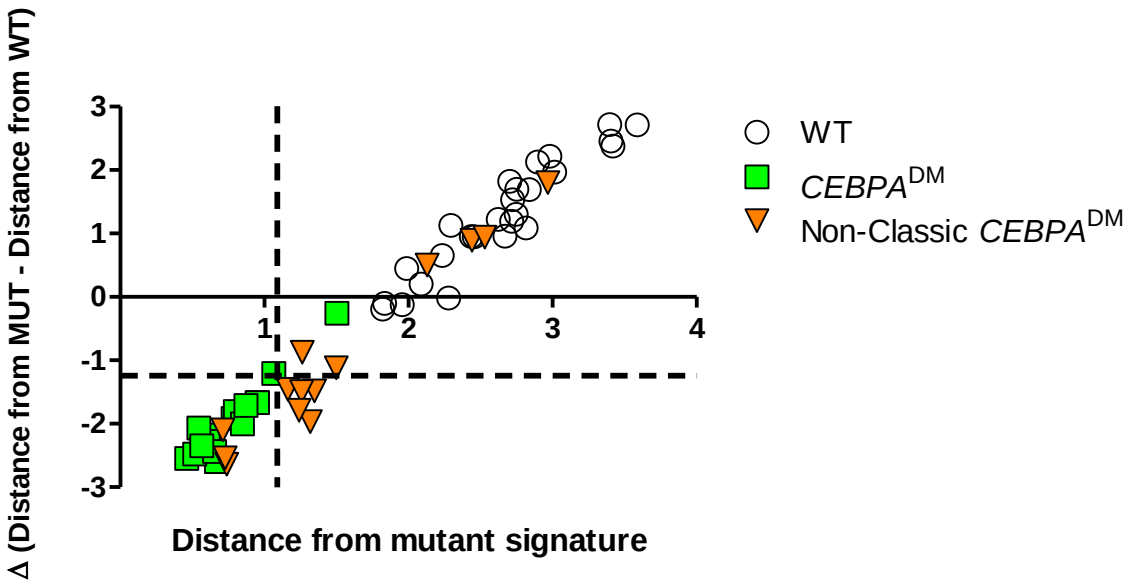


Figure 4.1 Distance scores for all *CEBPA*^{DM} and *CEBPA*^{WT} samples in the validation cohorts.

Table 4.3 Mutation screening of full length clones from *CEBPA*^{DM} samples

Patient No.	No. of clones analysed	No. of clones with 1 mutation	No. of clones with both mutations	No. of WT clones
12	19	15	1	3
27	37	36	0	7
49	27	27	0	0
53	24	18	1	5
76	21	16	0	5

4.3.1.2 Mutant level in *CEBPA*^{DM} samples

The relative mutant level of both mutants in 37 of the 41 *CEBPA*^{DM} cases in the three cohorts was already known (Green *et al*, 2010b) or had been quantified by other members of the department. This included 12 cases (16% of all mutations) that were estimated using peak heights in the sequence chromatogram, with the mean of at least five peaks, because the mutations did not lead to size changes and there were no restriction enzymes available to differentiate between the mutant and wild-type alleles. Mutant levels were measured in the remaining four cases (Table 4.4). Overall therefore, 77 mutations were quantified from the 41 *CEBPA*^{DM} samples analysed on the array. Assuming that the mutant allele level was half of the total mutant level for the five homozygous mutant samples, the mean mutant level was 44% (range, 24%-57%). Of note, the one classic *CEBPA*^{DM} case that did not fall within the classic quadrant had mutant levels of 24% and 31%, consistent with only half the cells in the sample carrying the mutations. Thus the methylation profile in this case could have been affected by the presence of a significant proportion of non-leukaemic cells. The remaining 24 classic *CEBPA*^{DM} cases had a mean mutant level of 44%, range 28%- 57%. The mean mutant level for the five non-classic cases that fell within the classic quadrant was 45% (range, 40%-50%), and it was 45% (range, 26%-54%) for the eleven cases that were outside the classic quadrant. One of the non-classic *CEBPA*^{DM} cases that did not fall into the classic quadrant group, patient number 69 had low mutant levels of 26%

and 38%. Apart from the two patient samples discussed, mutant level did not differ between samples within the classic quadrant from those outside of the quadrant.

Table 4.4 Mutant levels of compound heterozygous *CEBPA*^{DM} samples

Patient no.	Mutation 1	Level (%)	Mutation 2	Level (%)
27	334_335insGC	24	K304_Q305insL	31
53	113delG	46	938_939insTA	48
63	198_201dup	38	890G>C	54
76	232delC	36	899G>C	41

4.3.2 Methylation profile of non-classic *CEBPA*^{DM} cases

The initial cohort included two non-classic *CEBPA*^{DM} samples, both of which clustered with the classic *CEBPA*^{DM} samples on unsupervised cluster analysis (Figure 3.9). The follow-up cohorts contained 14 such cases, with a variety of different mutations, five of which fell in the relatively hypomethylated cluster (Figure 3.18). All these non-classic *CEBPA*^{DM} cases were therefore considered according to the predicted functional consequence of their mutations in order to assess their methylation profile.

Six cases were predicted to produce just p30 protein due to a classic N mutation on one allele and a frameshift mutation after the 2nd ATG site on the second allele (Figure 4.2A). Only one of these fulfilled the classic *CEBPA*^{DM} criteria, the remaining five were all more distant from the mutant signature. Three samples were predicted to produce only C-LOF protein; two had homozygous classic C mutations and the other one a classic C mutation coupled with a frameshift mutation after the 2nd ATG site. Two of the three cases fulfilled the classic *CEBPA*^{DM} criteria (Figure 4.2B). The remaining seven cases all had at least one missense mutation in the C terminus of unknown functional consequence (Figure 4.2C). Three of these samples had a classic N mutation with a missense mutation in the C terminus (p.A295P, p.R297P and p.R300P), thus if the missense mutation led to a C-LOF protein then these

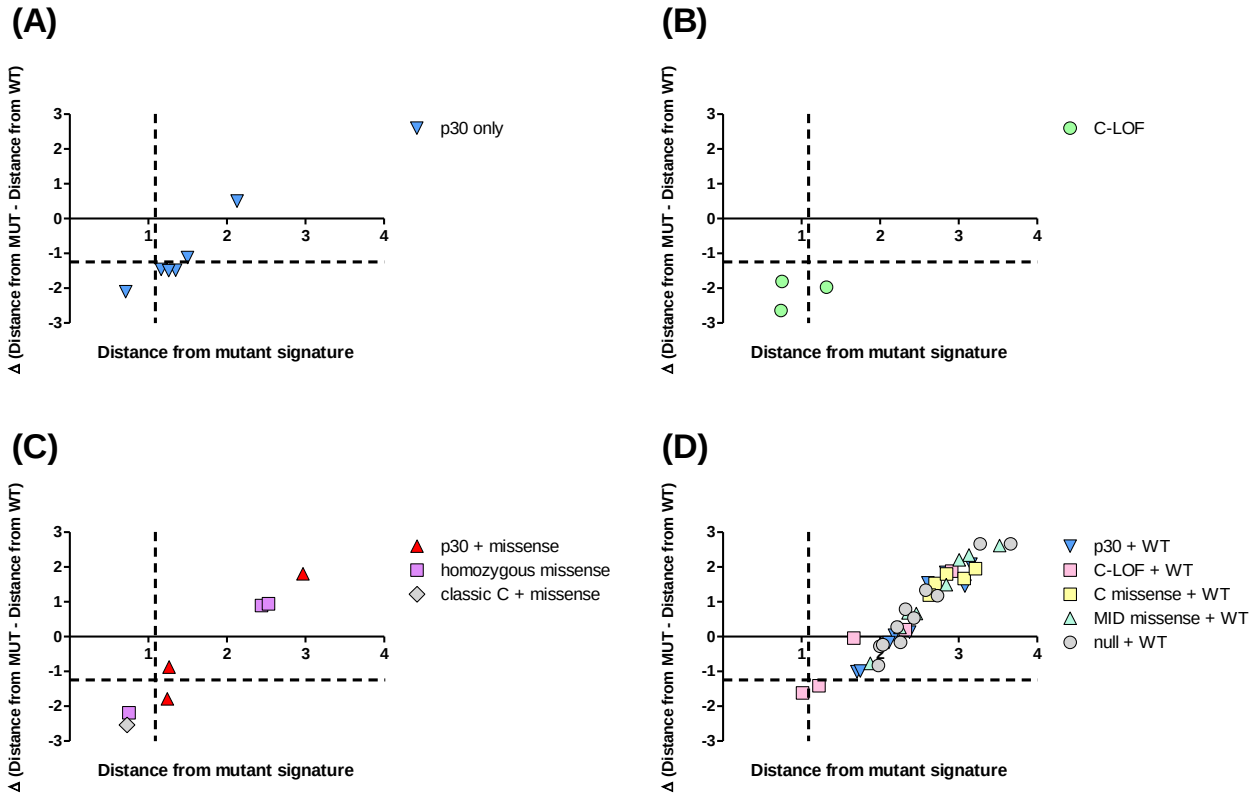


Figure 4.2 Distance scores for the non-classic *CEBPA*^{DM} and the *CEBPA*^{SM} cases. *CEBPA*^{DM} cases predicted to produce (A) p30 only, (B) C-LOF only, (C) with missense mutations. (D) *CEBPA*^{SM} cases. Null indicates a frameshift or nonsense mutation after the second ATG site.

Abbreviations: C-LOF, C-terminal loss of function mutation; MID-missense, missense mutation in the middle of the gene; MUT, mutation; p30, classic N mutation; WT, wild-type.

cases would be predicted to be equivalent to a classic *CEBPA*^{DM} case with p30 plus a C-LOF protein. All three samples were located outside the classic *CEBPA*^{DM} quadrant (Figure 4.2C). Three samples had homozygous missense mutations in the C terminus; one of these cases with a p.V314G mutation was in the first cohort. One further case had a classic C terminus mutation with a missense mutation. Assuming that all these mutations led to a C-LOF protein, these four cases would be predicted to produce just C-LOF protein without p30. Two of these samples satisfied the classic *CEBPA*^{DM} criteria (homozygous p.V314G and p.N321S/p.R306_V314dup) and two did not (homozygous p.N321S and homozygous p.L317Q).

4.3.3 Analysis of *CEBPA*^{SM} samples in the validation cohort

The follow-up cohorts included 38 *CEBPA*^{SM} cases, nine were classic N mutations and five classic C mutations, and the others were a range of non-classic mutations. On unsupervised analysis, seven (18%) clustered with the *CEBPA*^{DM} samples (Figure 3.16). By assessing the methylation levels using the *CEBPA* signature, only one satisfied the criteria for classic *CEBPA*^{DM} and was located within that quadrant and another was borderline; both of these samples had a classic C mutation (Figure 4.2D). One possible explanation is that the wild-type allele in these two cases was silenced leading to functional hemizygosity. Unfortunately, RNA was not available in these cases to examine relative expression of the mutant and WT alleles.

4.3.4 Methylation of the *CEBPA* promoter

4.3.4.1 Data from the arrays

It has been reported that the promoter region of the *CEBPA* gene can be methylated in AML and that this leads to a similar gene expression profile to *CEBPA*^{DM} (Wouters *et al*, 2007). To determine whether methylation of the *CEBPA* promoter, *CEBPA*^{METH}, could be influencing the methylation signatures of the samples investigated in these studies, the CpGs associated with the *CEBPA* gene that were analysed on the Illumina Methylation Arrays were assessed. The 27K HumanMethylation BeadChip analysed 2 CpG sites which were 516 bases upstream and 790 bases downstream from the TSS (Figure 4.3, probes 7 and 14).

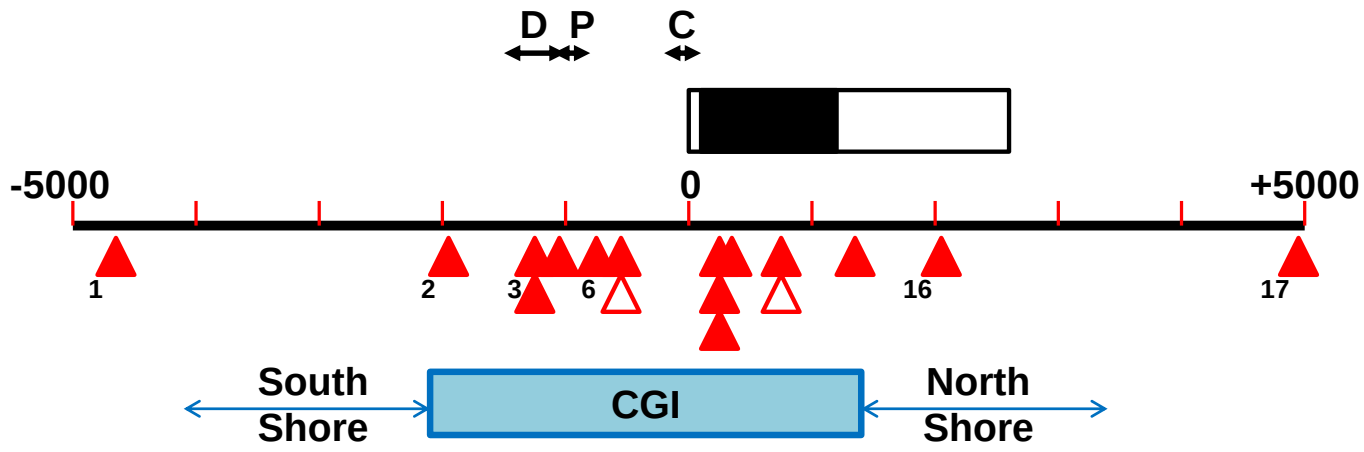


Figure 4.3 Location of probes on the Illumina methylation arrays that are associated with the *CEBPA* gene.

The black line represents the DNA, with the numbering depicting the distance from the transcription start site of *CEBPA*. The transcribed region is shown by the upper box, with the translated region shaded in black. The location of the CGI and shores is shown in the lower box. The unshaded triangles represent the location of the 2 CpG sites analysed on the 27K Illumina Methylation array. All 17 triangles represent the locations of the CpG sites examined by the 450K Illumina Methylation array. The three arrows indicate the distal (D), proximal (P) and core (C) regions of the *CEBPA* promoter as defined by Lin *et al* (2011).

The 450K HumanMethylation BeadChip analysed 17 CpG sites between the shelves from 4618 bases upstream to 4960 bases downstream of the TSS, including the two sites analysed on the 27K array.

The heatmap from the 450K array data of the samples from cohorts 2 and 3 is shown in Figure 4.4. Most CpG sites showed little variance in the methylation level across the whole cohort; all samples were unmethylated (β value <0.10) within the CGI (probes numbered 6-14 on Figure 4.4) and hypermethylated (β value >0.45) outside the shores (probes 1 and 17). The differential methylation within the cohort occurred at the outer boundaries of the CGI (probes 3-5 and 15) and shores (probes 2 and 16). For example probe 2 (cg21715751 on the BeadChip), 1919 bases upstream of the TSS was generally hypomethylated in all the *CEBPA*^{DM} samples, median β value 0.07 (range, 0.03-0.35) (Figure 4.5). Conversely, it was predominantly hypermethylated in the *CEBPA*^{SM} and *CEBPA*^{WT} samples, medians 0.6 (range, 0.04-0.92) and 0.79 (range, 0.08-0.92) respectively. The difference across all three groups was statistically significant (P<0.0001). Neither of the two *CEBPA*^{DM} samples that were outliers with β values >0.2 fulfilled the classic *CEBPA*^{DM} criteria based on the methylation signature. One of these was the classic *CEBPA*^{DM} sample that had low mutant levels and was thus thought to have a high proportion of non-leukaemic cells, the other had a homozygous missense mutation in the C terminus (p.N321S). Otherwise, no difference in methylation levels was observed in classic versus non-classic *CEBPA*^{DM} at this probe site. Although the median β value for all *CEBPA*^{SM} samples was significantly different from the median for *CEBPA*^{DM} samples, the standard deviation was also greater, 0.3 and 0.07 respectively. Seven of the 38 *CEBPA*^{SM} samples had methylation levels <0.2 at this probe, however this did not correlate with the *CEBPA* methylation signature. Only one of the two *CEBPA*^{SM} samples that was located in or at the border of the classic *CEBPA*^{DM} quadrant had a methylation level <0.2 (β values 0.38 and 0.1 respectively).

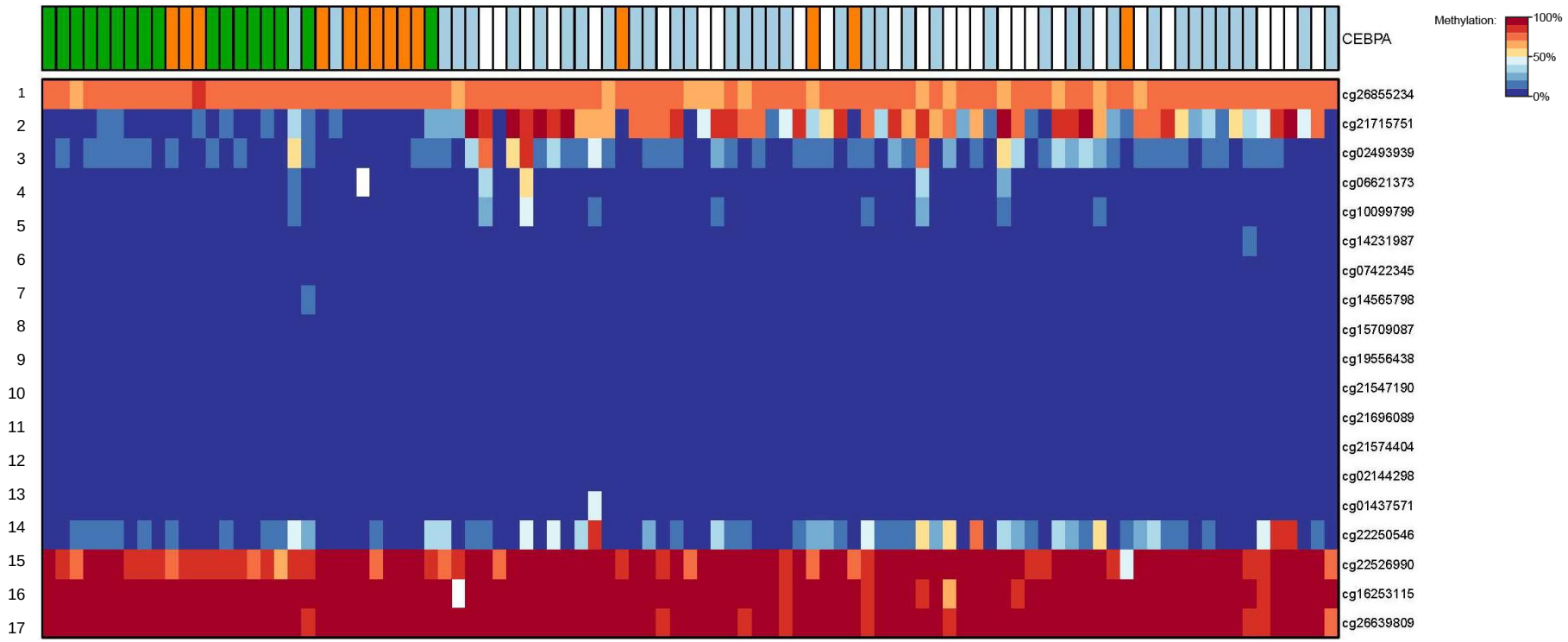


Figure 4.4 Heatmap showing the methylation values for probes closest to the *CEBPA* gene in the 95 patients in the follow-up cohorts, as analysed on the 450K Illumina Methylation Array. Each row represents a CpG probe, numbered as in Figure 4.3. Probes 3 and 4 fall within the distal region and probe 5 in the proximal promoter region as defined by Lin *et al* (2011). Each column represents a sample, ordered as in Figure 3.17, with the *CEBPA* genotype indicated above the heatmap, ■= classic *CEBPA*^{DM}, ■= non-classic *CEBPA*^{DM}, ■= *CEBPA*^{SM}, and □=*CEBPA*^{WT}.

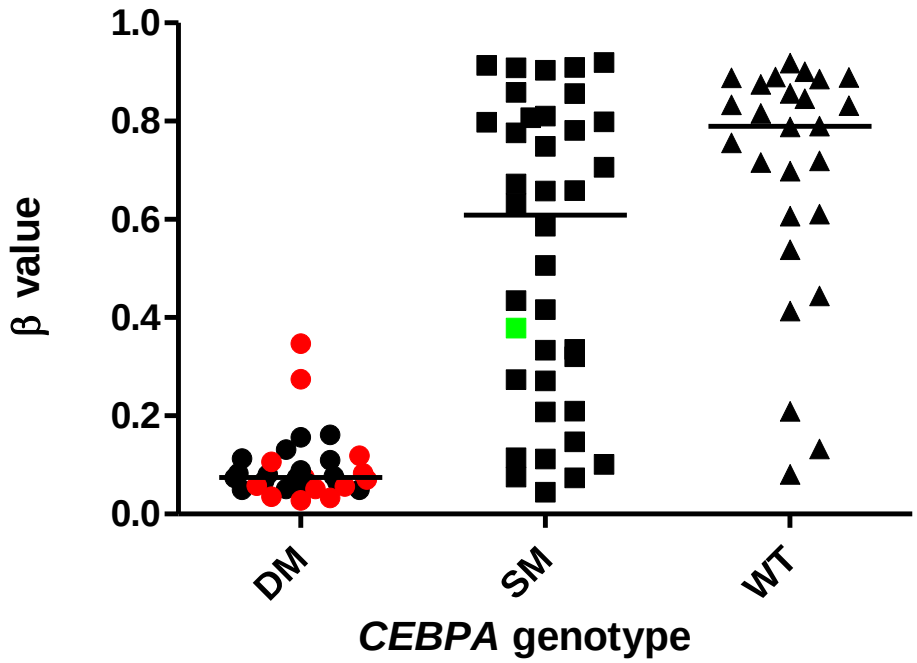


Figure 4.5 β values at probe 2, CG21715751 for the 95 samples in the follow-up cohorts, arranged according to *CEBPA* genotype.

The median β value is given. The samples highlighted in red in the *CEBPA*^{DM} column are those that did not fall within the classic *CEBPA*^{DM} quadrant. The sample highlighted in green in the *CEBPA*^{SM} column was located within the classic *CEBPA*^{DM} quadrant.

The CpG site cg02493939, probe 3, 1275 bases upstream of the TSS, located in the distal promoter region as defined by Lin *et al* (2011) and Fasan *et al* (2013a), was predominantly hypomethylated in all samples. In all *CEBPA*^{DM} samples, the maximum β value was 0.15. Of the 28 *CEBPA*^{SM} samples, two had a β value >0.4, one of which was located in the classic *CEBPA*^{DM} quadrant. Of the 26 *CEBPA*^{WT} samples, 17 (65%) had a β value >0.15, but only four (15%) were >0.5.

4.3.4.2 Analysis of the *CEBPA* promoter by bisulfite sequencing

Only two *CEBPA*-associated CpG sites interrogated as part of the Illumina Infinium Methylation 450K array were located in the distal region of the promoter (Probes 3 and 4, 1275 and 1219 bases respectively from the TSS) and one in the proximal region (Probe 5, 1085 bases from the TSS) as defined by Lin *et al* (2011). No probes were located in the core region. Thus three PCRs were designed to assess the methylation levels in the distal and core regions in more detail (Figure 4.6). Just one of 843 AML samples in published studies showed evidence of methylation in the proximal region, and thus this region was not examined (Fasan *et al*, 2013a; Lin *et al*, 2011; Musialik *et al*, 2014). Although bisulfite sequencing is not a truly quantitative method of analysing methylation levels in samples, an estimate of the methylation level was sought at each CpG site by measuring the height of the cytosine peak over the summed height of the cytosine and thymine peaks at that particular site (Jiang *et al*, 2009). To assess the accuracy of this approach, control samples with known levels of methylation were prepared by mixing bisulfite-converted fully methylated and unmethylated Epitect standards at defined ratios and then sequenced.

One PCR was created to examine the “core promoter region” defined as -11 bases upstream to +157 bases downstream of the TSS; in total 19 CpG sites were analysed between +15 and +98 (Figure 4.6). The PCR was assessed using standards of 0%, 50% and 100% methylation. The fully unmethylated standard showed 0% methylation at 17 of the 19 CpG sites, the remaining two sites had 30% and 11% methylation respectively. The fully methylated standard showed 100% methylation at all sites. The control sample that was 50% methylated had estimated levels of between 61-84% methylation across the

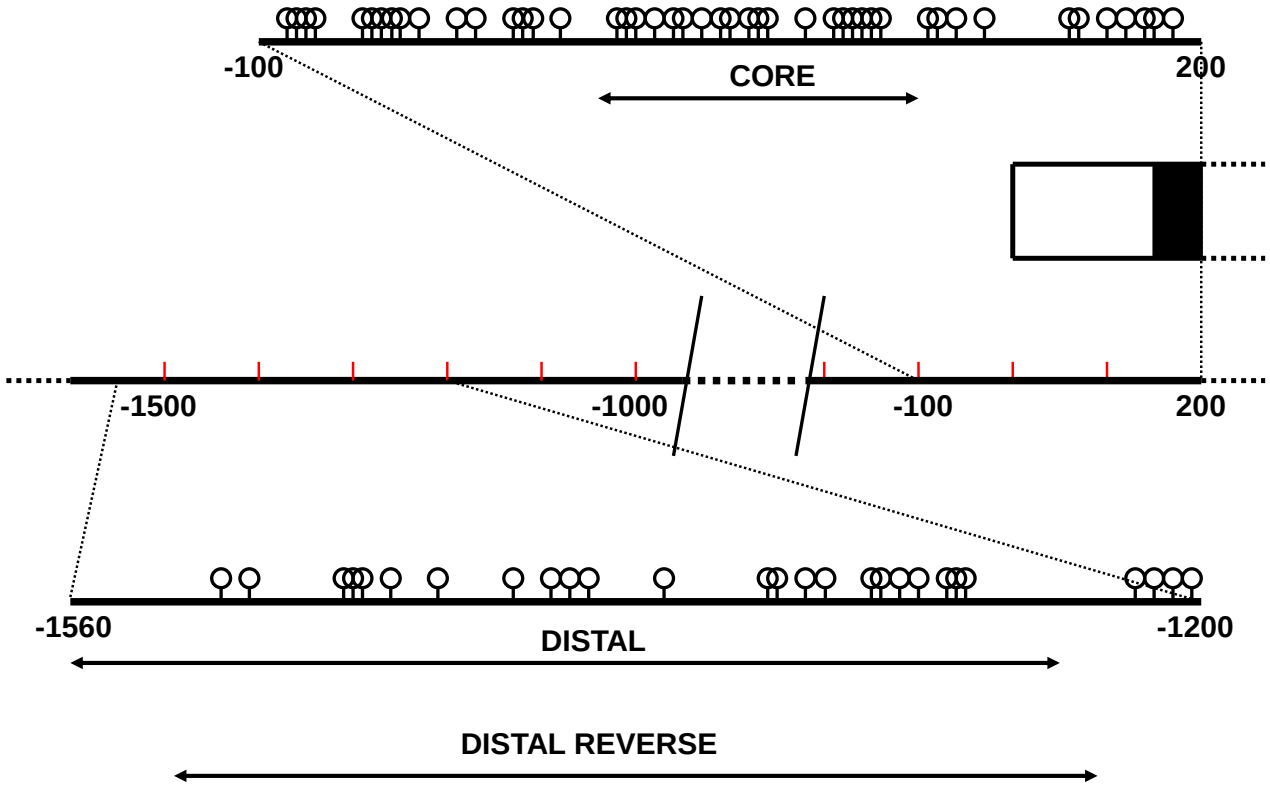


Figure 4.6 Location of the three PCRs designed to investigate methylation in the *CEBPA* promoter. The line in the centre represents the DNA, numbered according to distance from the transcription start site. The box denotes the *CEBPA* gene, the shaded part is the translated region. The expanded regions indicate all CpG sites in the core and distal regions of the promoter

sites. In spite of this marked bias in percentage methylation, all 135 samples analysed had 0% methylation at all 19 sites, indicating that there was no evidence of methylation at the core promoter region, irrespective of *CEBPA* genotype.

As most differential methylation occurred in the distal part of the distal region at approximately 1400 bases upstream of the TSS in the 238 cases examined by Fasan *et al* (2013a), and this area was not covered by the probes on the array, two PCRs were designed specifically for this part of the distal promoter. One analysed the forward strand from 1559 to 1245 bases upstream of the TSS covering 23 CpG sites, and the other the reverse strand from 1500 to 1238 bases upstream of the TSS covering 21 CpG sites (Figure 4.6). Although the two strands overlapped with 21 CpG sites in common, both were analysed in order to investigate possible strand-specific methylation, to validate the level of methylation at each CpG, and also to avoid miscalling C/T SNPs as unmethylated CpG sites. The forward strand was sequenced from the forward primer and thus the first two CpG sites were not seen as they were located within the first 30 bases of the amplicon. Similarly, the reverse strand was sequenced from the reverse primer, and five CpG sites were close to this primer and thus not assessed. Hence, 21 CpG sites were analysed in the forward direction (between -1474 and -1277), and 16 of these were also analysed in the reverse direction (between -1474 and -1311).

Three control samples with expected methylation levels of 10%, 50% and 90% were first analysed (Figure 4.7). Given that this region has been shown to have varying levels of methylation, the controls were selected to confirm that preferential amplification of either unmethylated strands or methylated strands did not occur, which could bias the results. Apart from one outlier result (CpG 12 in the forward direction), the estimated methylation of all CpG sites for the 10% methylation standard was within 13% of the expected level in both directions, and for the 90% methylation standard it was within 10% (Figure 4.8). There was more variability between the strands for the 50% methylation standard. The forward strand overestimated the level by 9-35% and the reverse strand underestimated the level by 2-27% at individual CpG sites. The general trend was an increase in estimated methylation level in the direction of

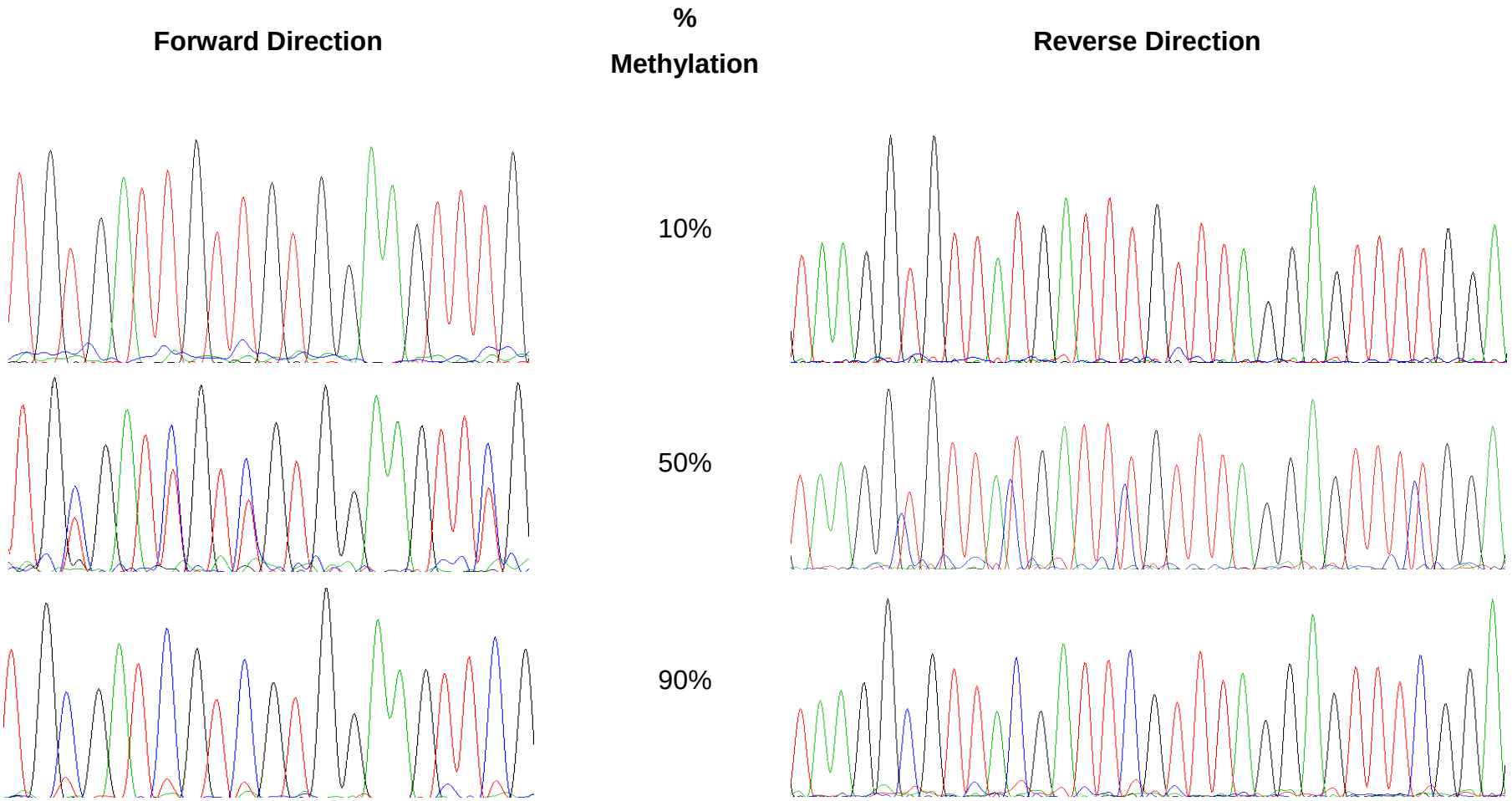


Figure 4.7 Bisulfite sequence of part of the *CEBPA* distal promoter region in control samples with known methylation levels.

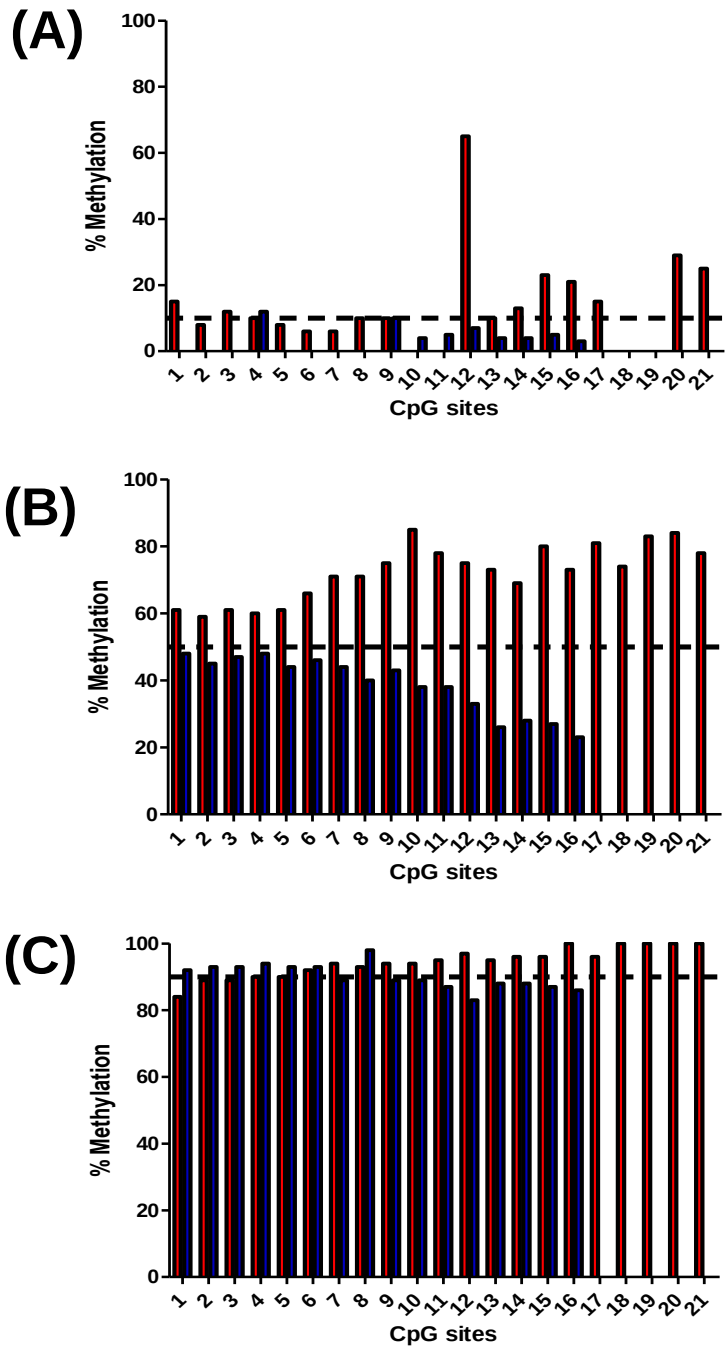


Figure 4.8 Estimated methylation levels at individual CpG sites examined by bisulfite sequencing of control samples. (A) 10%, (B) 50%, (C) 90% methylation. The 21 sites are those that were measured by the forward (CpG sites 1 to 21) and reverse PCRs (CpG sites 16 to 6). CpG sites 1 to 9 in the forward direction and 16 to 6 in the reverse direction were used for further analysis in patient samples based on peak height on the chromatogram.

■= Forward PCR; ■= Reverse PCR.

sequencing, which reflected the progressive reduction of the overall peak height of the T nucleotide along the sequence. This probably resulted from reagent depletion due to the repetitive regions and long T homopolymer runs after bisulfite conversion. Therefore, to maintain accuracy of methylation estimation, analysis of patient samples was restricted to CpG sites 1 to 9 in the forward direction and sites 16 to 6 in the reverse direction. The cut-offs were taken from the point in the sequences of the standards where the peak height of T nucleotides was less than 50% of the height of the T nucleotides at the beginning of the sequence.

Although one of the CpG sites included in the 450K array (Probe 3, 1275 bases upstream of the TSS) was within the forward PCR of the distal region, this CpG was not analysed as it was at the end of the sequencing and thus the peaks were too small to reliably interpret the methylation levels. Therefore, as the level of methylation tends to be similar in nearby CpG sites (Eckhardt *et al*, 2006), the probe 3 β values for the patient samples were compared with the estimated levels at CpG site 16 from the reverse PCR, which was 34 bases upstream of the probe 3 CpG and the closest of all CpG sites assessed. There was a significant degree of correlation between the two methods of quantification (r^2= 0.69, *P*<0.0001) (Figure 4.9). Of note, although levels at this CpG site were underestimated with the 50% standard, there was no evidence of such bias when comparing the results to the methylation array. The greatest variability in results was at the lower levels of methylation, where small changes in absolute nucleotide peak heights would have had a big effect on methylation estimation.

In order to determine whether there was any evidence for strand-specific methylation in the patient samples, the results for both strands at the four common CpG sites assessed were compared (CpG sites 6-9 in Figure 4.3). The median methylation level for the four sites for all samples analysed was calculated for each PCR and the two results compared. The correlation between the two estimations was highly significant (r^2= 0.56, *P*<0.0001), again with most variability occurring at the lower levels of methylation (Figure 4.10). Although the forward PCR consistently estimated a higher methylation value than the reverse PCR in the standards, there was no consistent bias in the patient samples. Of the 135 samples analysed, 11 (8%) showed a difference of

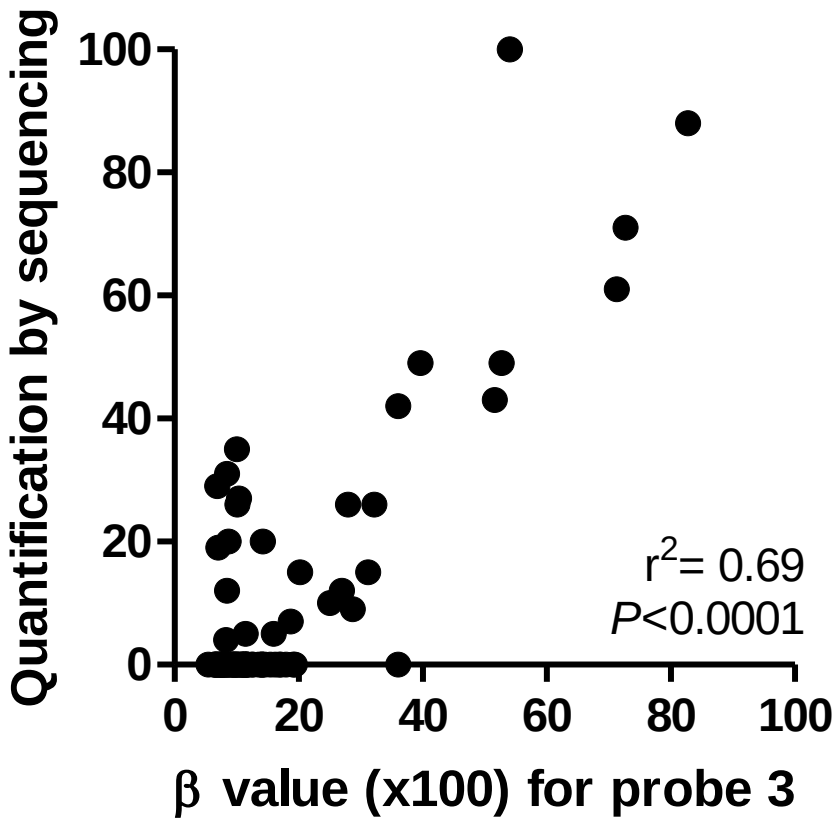
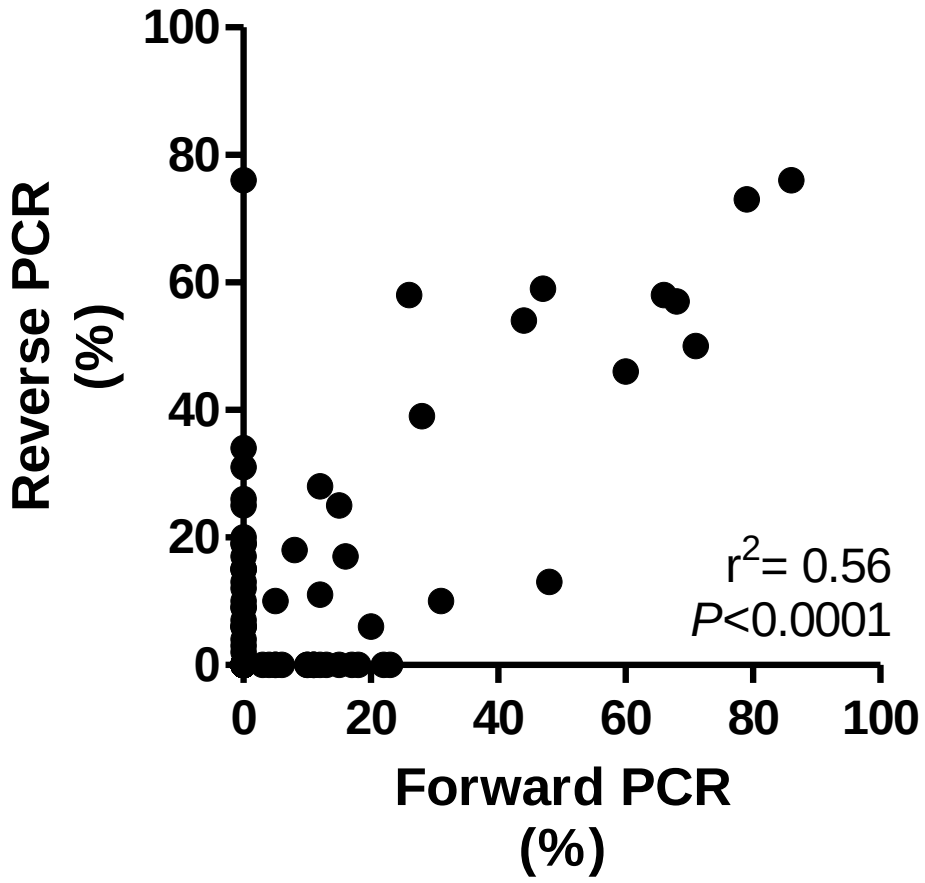


Figure 4.9 Comparison of methylation levels from the 450K array beta values with estimated levels from bisulfite sequencing in the distal *CEBPA* promoter region.

The array results are from probe 3, 1275 bases upstream from the transcription start site. The bisulfite sequencing results are the methylation level at a single CpG site 34 bases upstream from the array CpG, as estimated from the distal reverse PCR.



.

Figure 4.10 Comparison of the median methylation level in the forward and reverse strands at four CpG sites in the *CEBPA* distal promoter region as estimated by bisulfite sequencing in all 135 samples.

>20% in methylation level between the forward and reverse strands, 5 were more methylated in the forward direction and 6 in the reverse direction. This did not correlate with *CEBPA* genotype as five were *CEBPA*^{DM}, four *CEBPA*^{SM}, and three *CEBPA*^{WT}. One of the *CEBPA*^{SM} samples showed a difference of 76% between the two strands; despite repeating the PCRs this discrepancy remained.

Analysing the median methylation levels obtained according to the *CEBPA* genotype revealed that the majority of samples with *CEBPA*^{DM} did not show any evidence of methylation in the *CEBPA* distal promoter region, and those that did only showed low levels of methylation, median level 0% in both the 9 sites analysed in the forward PCR and the 11 sites in the reverse PCR (range, 0-46% and 0-32% respectively) (Figure 4.11).

The majority of the 38 *CEBPA*^{SM} samples also showed little methylation in this region, medians 0% (range, 0-45%) and 0% (range, 0-94%) in the forward and reverse directions (Figure 4.11). When individual CpG sites were examined, five *CEBPA*^{SM} samples had >25% methylation at at least one CpG site (Figure 4.12A and B). To assess whether methylation of the *CEBPA* promoter could lead to silencing of the wild-type allele and thus cause some *CEBPA*^{SM} to behave like *CEBPA*^{DM} samples, *CEBPA*^{SM} samples were split into *CEBPA*^{METH} (n=5) and *CEBPA*^{HYPOMETH} (n=33), with *CEBPA*^{METH} defined as >25% median methylation in at least one of the two distal PCRs. This level was chosen as it would be consistent with one allele being silenced in at least half of the cells, although not necessarily the wild-type allele. The methylation profile of the samples that were considered *CEBPA*^{METH} was then examined. Four of these samples did not have methylation signatures similar to *CEBPA*^{DM} (Figure 4.13A). The remaining sample was one of the two *CEBPA*^{SM} samples that satisfied the classic *CEBPA*^{DM} criteria. It had the highest methylation level in the *CEBPA* promoter region, with a marked strand-specific variation in methylation level, median 100% in the reverse direction and 0% in the forward direction. Furthermore, this sample was the only one of the five *CEBPA*^{METH} to have a classic C-LOF mutation, the other samples had a classic N mutation, a frameshift in the middle of the gene, a missense mutation in the C terminus and a missense mutation in the middle of the gene. Thus methylation of the distal

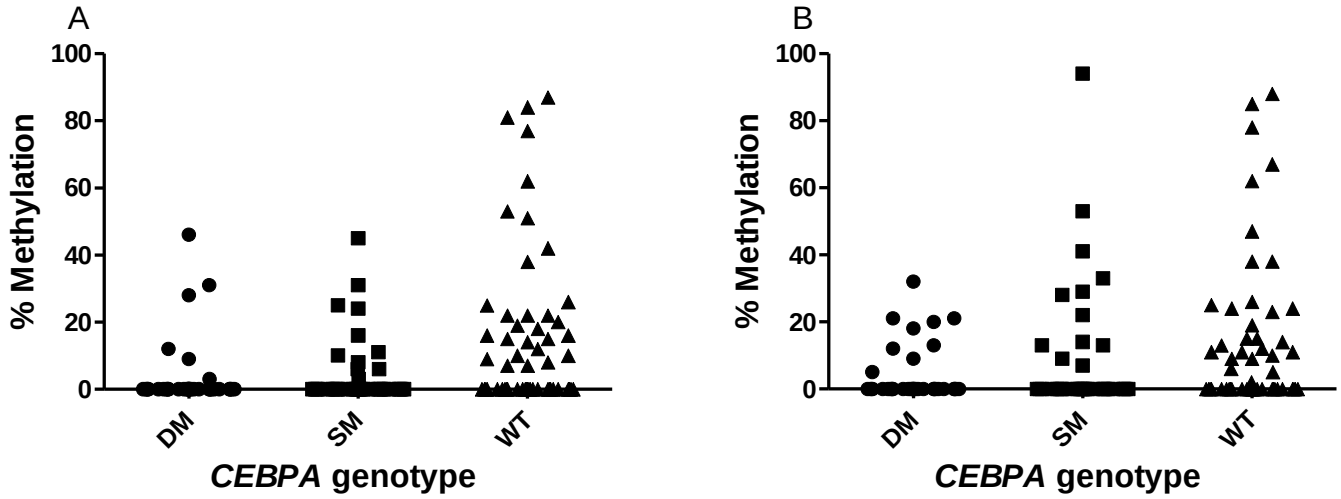


Figure 4.11 Median methylation levels in the distal *CEBPA* promoter region for AML samples according to their *CEBPA* genotype.

Results from (A) the forward and (B) the reverse PCR products.

.

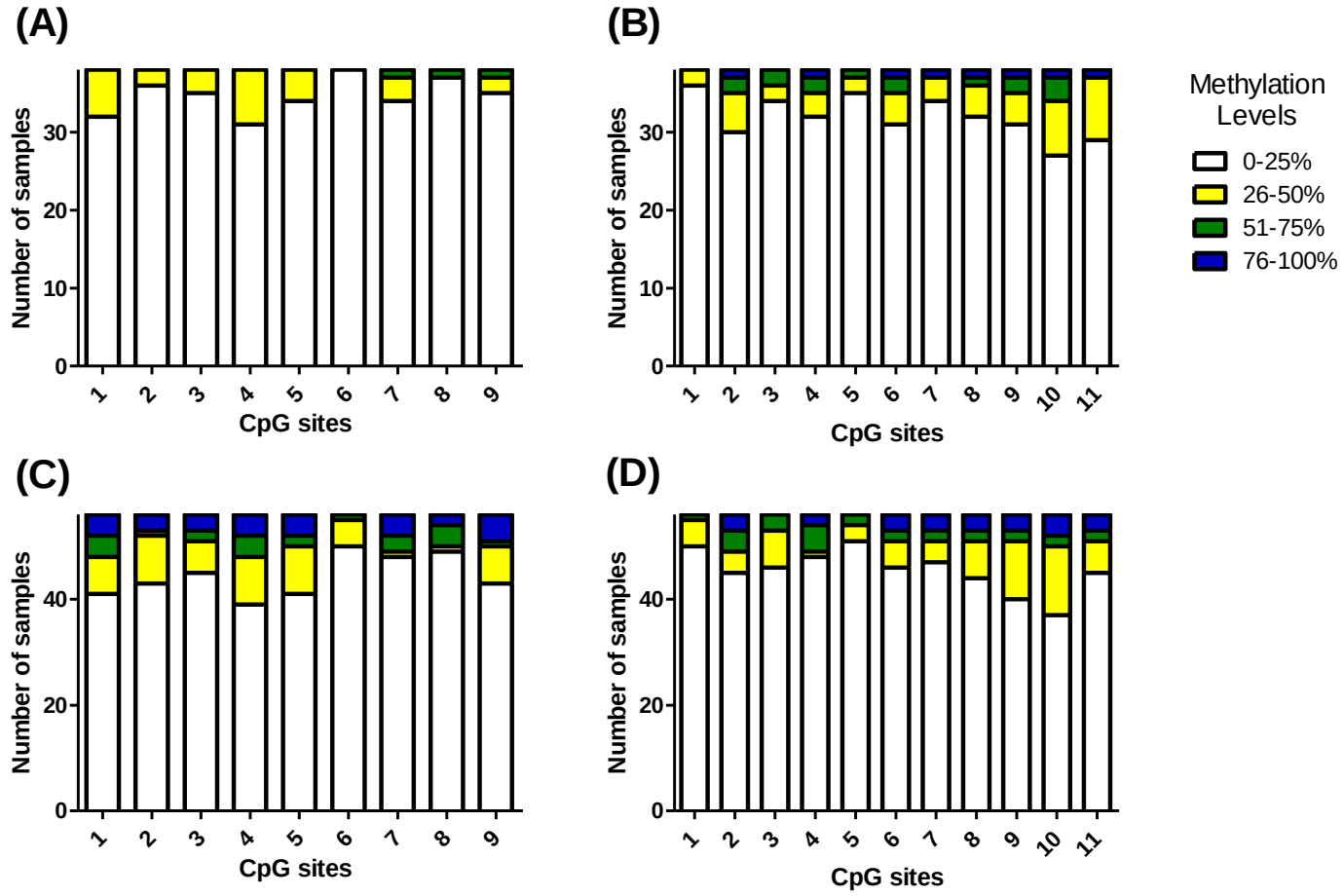


Figure 4.12 Number of samples with evidence of methylation at individual CpG sites in the *CEBPA* distal promoter region. *CEBPA*^{SM} samples (A) and (B), and *CEBPA*^{WT} samples (C) and (D) in forward and reverse PCRs respectively.

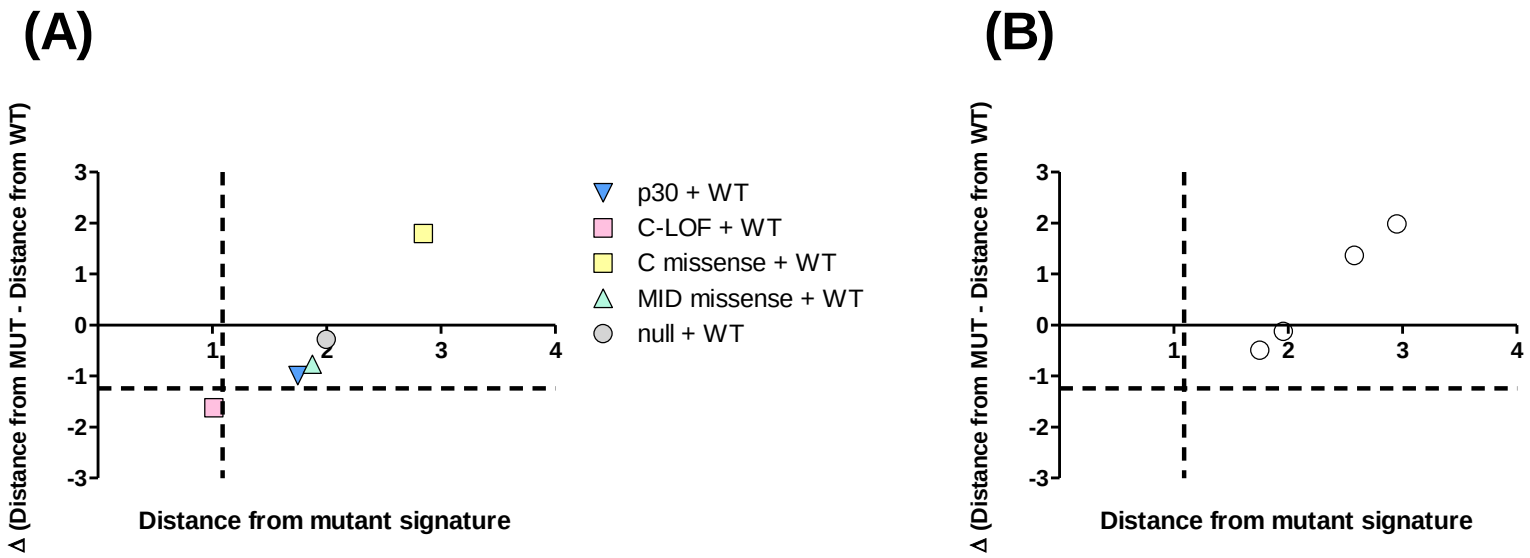


Figure 4.13 Distance scores for cases with hypermethylation in the distal *CEBPA* promoter region.

(A) *CEBPA*^{SM}, (B) *CEBPA*^{WT} cases. Hypermethylation is defined as >25% median methylation in at least one of the two PCRs for *CEBPA*^{SM} and >75% median methylation for *CEBPA*^{WT} cases.

CEBPA promoter was seen in one *CEBPA*^{SM} sample and if the wild-type allele was silenced this might explain why it fell in the *CEBPA*^{DM} quadrant. However, no RNA sample was available to confirm this.

The median results for the 56 *CEBPA*^{WT} samples ranged from 0-90% in the forward direction and 0-88% in the reverse direction (Figure 4.11). The majority of samples showed no evidence of methylation at any of the CpG sites in this region (Figure 4.12). As before, *CEBPA*^{WT} samples were split into *CEBPA*^{METH} (n=4) and *CEBPA*^{HYPOMETH} (n=52), but here *CEBPA*^{METH} was defined as >75% median methylation in at least one of the two distal PCRs. This level of methylation was chosen as Wouters *et al* (2007) had previously shown that four of their six *CEBPA*-silenced (*CEBPA*^{WT}) cases with gene expression profiles similar to *CEBPA*^{DM} had >75% methylation in the distal region. Four *CEBPA*^{WT} samples fulfilled this criterion, but as none of them were located within the *CEBPA*^{DM} quadrant, there was no evidence that methylation in this region was associated with a methylation profile similar to *CEBPA*^{DM} (Figure 4.13B).

4.4 Discussion

AML samples with classic *CEBPA*^{DM} have a distinct methylation signature, as has been shown in the previous chapter and published literature (Figueroa *et al*, 2010b). However, approximately 25% of patients with *CEBPA*^{DM} have a non-classic combination of mutations. The functional impact of these mutations and the prognostic consequences are unknown. As these mutations are very varied, it is difficult to assess their impact on prognosis. The aim of the studies presented in this chapter was to assess the methylation signature of the non-classic *CEBPA*^{DM} and *CEBPA*^{SM} samples to explore whether this could serve as a surrogate for providing information on the more broad-spectrum functional significance of different mutations.

The first part of this chapter explored some of the potential reasons why certain *CEBPA*^{DM} did not satisfy the classic *CEBPA*^{DM} criteria and were located outside this quadrant. Cloning of full-length *CEBPA* amplicons was performed for five *CEBPA*^{DM} samples and this showed that they were all biallelic. This is not surprising, as most studies have confirmed that the vast majority of mutations in *CEBPA*^{DM} are biallelic (Barjesteh van Waalwijk van Doorn-Khosrovani *et al*,

2003; Dufour *et al*, 2010; Frohling *et al*, 2004; Green *et al*, 2010b; Lin *et al*, 2005; Preudhomme *et al*, 2002). There is a suggestion in the data presented here that the mutant level of samples may have an impact on the methylation signature as the one classic DM sample that was not located within the classic *CEBPA*^{DM} quadrant had a lower level of both mutations than all other samples. If methylation signatures are to be used to investigate the functional impact of different mutations, further work must be done to explore the impact of mutant levels on the methylation levels in samples.

The type of *CEBPA* mutation also had an impact on the methylation signature of the *CEBPA*^{DM} samples. The mutations identified in *CEBPA* are diverse and are predicted to have different functional consequences. The p30 isoform that results from classic N-terminal mutations lacks TAD1, which is important for transcriptional activation and regulation of cell cycle progression, whereas classic C mutants have been shown to disrupt either DNA binding or dimerisation of C/EBPα (Friedman, 2015; Ohlsson *et al*, 2016). However, the impact of most of the non-classic mutations has not been investigated previously and thus the functional impact of these mutations is still unclear. Five of the six samples with mutations that would lead to functional p30 only (classic N mutation and frameshift after the second ATG start site) were not located in the classic *CEBPA*^{DM} quadrant. Two of the three samples with predicted C loss of function satisfied the classic *CEBPA*^{DM} criteria. Thus it seems a key feature of the *CEBPA*^{DM} signature is that no wild-type p42 is present, and that a C-LOF mutation may be necessary. In their mouse models Kato *et al* (2011) demonstrated that mice transplanted with transduced cells containing a C-terminal mutation alone developed AML, however if they were transduced with an N-terminal mutation alone this did not occur. The seven other non-classic *CEBPA*^{DM} samples all had at least one missense mutation, of these two were located in the classic *CEBPA*^{DM} quadrant. This demonstrates that missense mutations are quite varied and it is difficult to predict their impact on function. Of the 38 *CEBPA*^{SM}, only one was located within the *CEBPA*^{DM} quadrant, and one was borderline. This is consistent with published data on gene expression profiles and the prognostic impact of *CEBPA*^{SM} showing that they are more similar to *CEBPA*^{WT} than *CEBPA*^{DM}.

The *CEBPA* promoter was analysed to assess whether methylation of the promoter was associated with a methylation signature similar to classic *CEBPA*^{DM} as previous studies had shown that samples with silenced *CEBPA* secondary to methylation had similar gene expression profiles to *CEBPA*^{DM} (Wouters *et al*, 2007). The core and distal regions of the promoter were selected as these regions had been previously been shown to be methylated in some AML samples. The degree of methylation of the *CEBPA* promoter observed in the current data set was generally consistent with published data. No *CEBPA*^{MUT} sample had evidence of methylation in the core region. The median methylation level in the distal region for *CEBPA*^{DM} and *CEBPA*^{SM} samples was 0%; however 12 of the 79 samples (15%) did have methylation levels above 20%, so *CEBPA*^{MUT} and *CEBPA*^{METH} were not mutually exclusive, as had been reported by other groups (Fasan *et al*, 2013a; Lu *et al*, 2010). No *CEBPA*^{WT} samples had evidence of methylation in the core promoter region, in keeping with published studies where the rate of methylation for *CEBPA*^{WT} samples in this region is low at approximately 4% (Chim *et al*, 2002; Fasan *et al*, 2013a; Hackanson *et al*, 2008; Jost *et al*, 2009; Lin *et al*, 2011; Lu *et al*, 2010; Musialik *et al*, 2014; Szankasi *et al*, 2011). In contrast, the rate of methylation in the distal promoter in *CEBPA*^{WT} samples was 16%, which is again consistent with the average rate of 34% reported by others, although individual studies have marked differences with a range of 15-51%. This variation is likely to reflect the different methods used to detect methylation and how it is defined (Fasan *et al*, 2013a; Hackanson *et al*, 2008; Lin *et al*, 2011; Musialik *et al*, 2014).

The data from the 450K array showed that CpG sites of all samples, regardless of underlying *CEBPA* genotype, were methylated at 4618 bases upstream from the TSS for *CEBPA* (probe 1) and unmethylated from 720 bases upstream to 1323 bases downstream of the TSS (probes 6-15). Thus somewhere in the boundary between the shore and the CGI the CpG sites go from being methylated to unmethylated. Of note, one site at 1919 bases upstream of the TSS (probe 2), which is further upstream than the defined distal promoter region, was significantly differentially methylated between *CEBPA*^{DM} samples (median 7% methylated) compared to both *CEBPA*^{SM} and *CEBPA*^{WT} (medians 61% and 79% methylated respectively. More extensive analysis of this region

would therefore be of interest to explore this finding. All published studies of the distal region of the *CEBPA* promoter have examined the methylation levels at CpG sites in a relatively short stretch of DNA, and have classified a sample as either “methylated” or “unmethylated”. However, perhaps a more robust way of discriminating samples would be to analyse the distance from the TSS at which CpG sites change from being methylated to unmethylated. For example, Fasan *et al* (2013a) observed that many of the 238 samples that had >15% methylation in the distal region at approximately -1400 bases upstream of the TSS gradually became unmethylated closer to the TSS, with half of them unmethylated at about -1300 bases and only 8% methylated at -1100 bases. Although the data presented here also shows variability in the position that CpG sites go from being methylated to unmethylated, the functional significance of this is currently unknown. Some studies have shown an association between methylation of the distal promoter and decreased *CEBPA* expression (Fasan *et al*, 2013a; Lin *et al*, 2011; Musialik *et al*, 2014), but this association was not seen in other studies (Hackanson *et al*, 2008).

Of note, a novel long noncoding RNA arising from 800 bases upstream of the TSS for *CEBPA* was recently identified, coined extra coding *CEBPA*, ec*CEBPA*, which inhibits DNA methylation of the *CEBPA* promoter in cell lines and leads to a concomitant increase in gene expression (Di Ruscio *et al*, 2013). The TSS was predicted to be 800 bases upstream of the canonical *CEBPA* TSS, however ChIP analysis led to enrichment of RNA polymerases from 1800 bases upstream. Hence the differential methylation in this distal region at 1919 bases upstream seen in this study may be linked to ec*CEBPA*. Expression of ec*CEBPA* was shown to be concordant with *CEBPA* expression in normal human tissue such as liver and lung, primary haematopoietic samples were not examined. *CEBPA* mutated samples have been shown to have a distinctive long noncoding RNA signature, when compared to *CEBPA*^{WT} samples, but this signature did not include ec*CEBPA* (Garzon *et al*, 2014).

Although studies have shown that *CEBPA*^{WT} samples with methylation of the *CEBPA* promoter have similar gene expression to samples with *CEBPA*^{DM} (Figueroa *et al*, 2009b; Taskesen *et al*, 2011; van Vliet *et al*, 2013), there was no evidence from the data presented here that the methylome for these two groups was similar. All four *CEBPA*^{WT} samples and four of the five *CEBPA*^{SM}

samples exhibiting hypermethylation of the *CEBPA* promoter did not have methylation profiles that were located within the classic *CEBPA*^{DM} quadrant. Examining different areas of the *CEBPA* promoter may lead to different results, but the areas interrogated were very similar to those published.

There were too few cases to determine whether outcome differed for double-mutated cases with non-classic mutations that had methylation profiles that either did (n=5) or did not fall (n=11) in the classic *CEBPA*^{DM} quadrant. Given that all *CEBPA*^{DM} patients lacking a *FLT3*/ITD are classified as good-risk and are therefore not usually considered for allogeneic transplantation in first remission (Cornelissen *et al*, 2012), these studies suggest that further work is required to determine which non-classic *CEBPA*^{DM} are equivalent to the classic *CEBPA*^{DM}, and thus caution should be employed about considering them as part of the favourable prognostic group.

CHAPTER 5:**DETECTION OF ASXL1 MUTATIONS IN AML AND THEIR PROGNOSTIC IMPACT**

5.1 Introduction

In addition to DNA methylation, the structure of chromatin has been implicated in the regulation of gene expression. Two families of proteins encoded for by evolutionarily conserved genes play a role in post translational modification of histones. The Polycomb and trithorax groups lead to repressive and active conformations of chromatin respectively. Mutations in these genes were first discovered in *Drosophila* and led to abnormalities in embryonic segmental development (Jurgens, 1985; Lewis, 1978). Mutations in Polycomb genes resulted in transformation of anterior segments of the *Drosophila* to posterior segments through derepression of homeotic genes, whilst mutations in the trithorax group proteins led to silencing of the homeotic genes and the opposite phenotype, with features of the anterior segments on the posterior of the fly.

One of the implicated genes is the *Asx* (additional sex combs) gene, which was first found in *Drosophila* in 1984 by screening hundreds of embryos with lethal mutations and examining their morphology. Mutations in this gene led to partial transformation of the head and thorax to additional abdomen and sex combs (short bristles found on the males' front legs that help with mating) (Nüsslein-Volhard *et al*, 1984). It was initially classified as a Polycomb gene, as its mutation gave rise to ectopic posterior segments (Breen & Duncan, 1986; Jurgens, 1985). However, certain mutations in *Asx* were later noted to also cause anterior transformations typical of the trithorax group proteins (Milne *et al*, 1999). Given this dual function, *Asx* was reclassified in a new group, the enhancers of trithorax and Polycomb genes (Gildea *et al*, 2000).

There are three human homologues of *Asx*, *ASXL1* (Additional sex combs like 1), *ASXL2*, and *ASXL3*. The *ASXL1* gene was discovered by searching for human sequence similarity to the *Drosophila Asx* gene and mapped to chromosome 20q11, a region that is frequently amplified or deleted in various tumours including leukaemias (Fisher *et al*, 2003).

5.1.1 Structure and function of ASXL1

The *ASXL1* gene consists of 12 or 13 exons, depending on whether a 3 base pair region is considered as a separate exon, 3 of 13, as originally proposed by Fisher *et al* (2003) or the 5' end of the following exon. The dominant transcript is 7031 base pairs and is expressed in most adult tissues at low to moderate levels, apart from liver and kidneys where it is not expressed (Fisher *et al*, 2003). The ASXL1 protein consists of 1541 amino acids with conserved sequence within the ASXH (Asx homology) region, which consists of ASXN (Asx N terminus) and ASXM (Asx Middle) regions (Figure 5.1). Within these regions there are at least 3 nuclear receptor boxes, two within the ASXM region and one further downstream. There is also a conserved plant homeo domain (PHD) located in the C terminus. The PHD zinc finger binds specific residues on histones and is found in many proteins that modify chromatin structure (Sanchez & Zhou, 2011).

In vitro studies of ASXL1 structure have shown that it interacts with several proteins, some of which seem to have opposing effects. For example, in non-haematopoietic cells, it has been shown that ASXL1 binds HP1a and LSD1, a histone demethylase, to repress retinoic acid-receptor activity (Lee *et al*, 2010). Conversely, ASXL1 can interact with SRC-1 to activate retinoic acid receptors (Cho *et al*, 2006). Similarly, transcriptional silencing is initiated or maintained by the recruitment of the PRC2 complex by ASXL1 to specific loci, allowing EZH2 to catalyse the di and tri-methylation of K27 in histone H3. The PRC1 complex then recognises the H3K27me3 marks and mediates ubiquitination of Lysine 119 of histone H2A. Both these histone marks are thought to silence transcription through several mechanisms including recruitment of DNA methyltransferases, prevention of RNA pol II activity and compaction of the chromatin (Abdel-Wahab *et al*, 2012) (Figure 5.2). On the other hand, in *Drosophila*, ASX has also been shown to interact with the polycomb-repressive deubiquitinase, PR-DUB, which removes monoubiquitin from histone H2A, causing activation of the homeotic genes (Scheuermann *et al*, 2010), although the PR-DUB complex has not been confirmed *in vivo* in mammals.

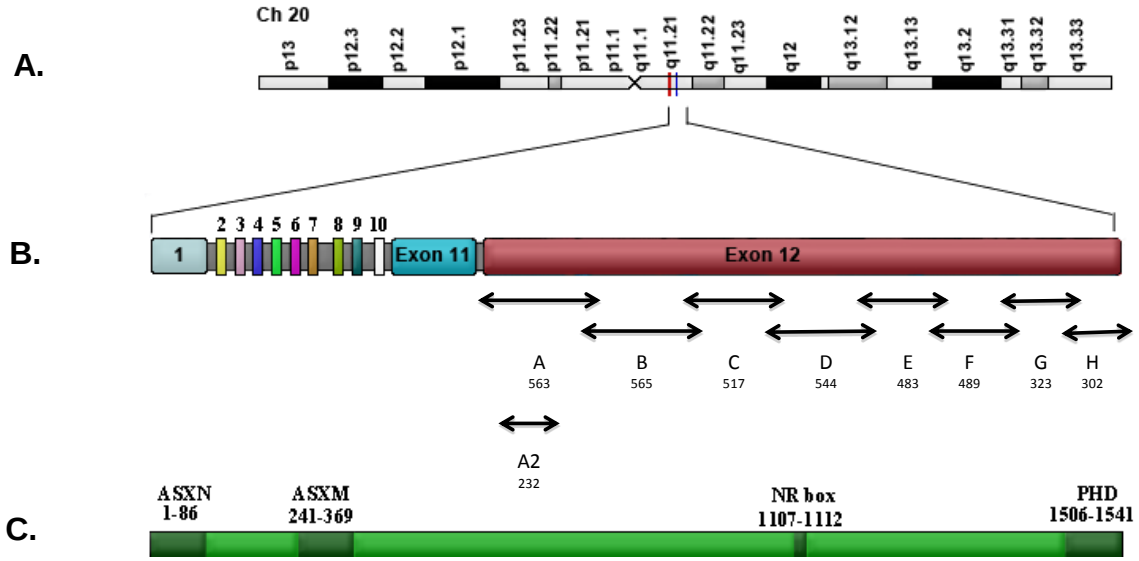


Figure 5.1 Representation of *ASXL1* locus and PCR fragments used for detection of *ASXL1* mutations by WAVE® analysis. (A) Chromosome 20 with site of *ASXL1.* (B) *ASXL1* gene with location of PCR fragments and length of each amplicon (base pairs). (C) ASXL1 protein with proposed domains. ASXN, conserved domain in the N terminus; ASXM conserved domain in the middle; NR box (nuclear receptor) box; PHD, plant homeodomain. Adapted from http://atlasgeneticsoncology.org/Genes/ASXL1ID44553ch20q11.html (Huret *et al*, 2013)

(COPYRIGHT PROTECTED IMAGE REMOVED)

Figure 5.2 Proposed function of ASXL1 and the polycomb complexes. Transcriptional silencing is initiated or maintained by the recruitment of the PRC2 complex by ASXL1 to specific loci, this allows EZH2 to catalyse the di and tri-methylation of K27 in histone 3. The PRC1 complex recognises the H3K27me3 and mediates ubiquitination of Lysine 119 of histone H2A. Removal of H2AK119Ub has been shown to occur by the PR-DUB complex (containing BAP1 and ASX) in *Drosophila* and mammalian cells *in vitro*. Adapted from Shih *et al* (2012)

5.1.2 Mutations in the *ASXL1* gene

ASXL1 mutations were first discovered by array comparative genomic hybridisation of 40 samples from patients with myelodysplasia (MDS) and secondary AML (Gelsi-Boyer *et al*, 2009). A small deletion in the 20q region that included *ASXL1* was identified in one of the samples, and therefore the gene was sequenced in 35 MDS and 39 chronic myelomonocytic leukaemia (CMML) samples and found to be mutated in 5 (11%) and 17 (43%) cases respectively. All mutations were localised to exon 12. Since this initial report, several groups have published their findings of recurrent *ASXL1* mutations in the spectrum of myeloid malignancies (Table 5.1). Of the studies that have examined the whole gene, only 8 nonsense or frameshift mutations outside exon 12 have been identified in 836 samples (1%). Given that this gene is 7031 base pairs and initial reports confirmed the majority of mutations were restricted to exon 12, which is itself 2907 bases, many studies have limited analysis to this exon only.

Considering all reported studies of mutations in *ASXL1* in myeloid malignancies, they are most commonly seen in CMML, with an overall incidence of 40%. They are found in MDS and myeloproliferative neoplasms (MPN) at similar rates of 20% and 23% respectively. Within the MPNs, they are most common in primary myelofibrosis (26%). The incidence in AML has varied in studies from 5%-25%, however overall it is 11% (Table 5.1). Mutations have also been seen infrequently in chronic lymphocytic leukaemia, and non-haematological tumours such as prostate and breast cancer (Katoh, 2013). Constitutional mutations in this gene have been discovered in the rare Bohring-Opitz syndrome (Hoischen *et al*, 2011; Magini *et al*, 2012), which is not associated with haematological abnormalities, although most patients die in early childhood.

The mutations are predominantly nonsense or frameshift mutations that would lead to premature truncation of the protein and absence of the conserved Plant Homeo Domain finger that is necessary for histone binding. The most commonly found mutation, c.1934dupG (p.G646WfsX12), is a duplication of a guanine within an 8-base guanine homopolymer run. It constitutes 43% of all *ASXL1* mutations detected in myeloid diseases. One group has suggested that this mutation could be a PCR artefact (Abdel-Wahab *et al*, 2010a), as they noted that this alteration was present in paired normal DNA from patients with

Table 5.1 Mutation rate of *ASXL1* in myeloid malignancies

Reference	Method of analysis	Exons analysed	Number of samples analysed	Incidence of *ASXL1* mutation (%)						
				De novo AML	Secondary AML	MDS	CMML	MPN	CML	JMML
Gelsi-Boyer *et al* (2009)	Sequencing	All	74			4/35 (11)	17/39 (44)			
Carbuccia *et al* (2009)	Sequencing	All	64					5/64 (8)		
Carbuccia *et al* (2010)	Sequencing	All	63	3/46 (7)	9/17 (53)					
Abdel-Wahab *et al* (2010b)	Sequencing	All	63		12/63 (19)					
Boultwood *et al* (2010a)	Sequencing	12	300	8/27 (30)	9/40 (23)	28/182 (15)	17/51 (33)			
Szpurka *et al* (2010)	Sequencing	All	23			2/23[1] (10)				
Sugimoto *et al* (2010)	Sequencing	12	49							2/49 (4)
Boultwood *et al* (2010b)	Sequencing	12	41						6/41 (15)	
Rocquain *et al* (2010)	Sequencing	12	129	11/64 (17)		12/65 (19)				
Chou *et al* (2010b)	Sequencing	12	501	54/501 (11)						
Gelsi-Boyer *et al* (2010)	Sequencing	12	53				25/53 (47)			
Perez *et al* (2010)	Sequencing	12	68							3/68 (4)
Makishima *et al* (2011)	Sequencing	12	54						2/54 (4)	
Abdel-Wahab *et al* (2011)	Sequencing	All	110				16/39 (41)	9/71[2] (12)		

Table 5.1 continued

Reference	Method of analysis	Exons analysed	Number of samples analysed	Incidence of *ASXL1* mutation (%)						
				De novo AML	Secondary AML	MDS	CMML	MPN	CML	JMML
Thol *et al* (2011b)	Sequencing	12	193			40/193 (21)				
Roche-Lestienne *et al* (2011)	Sequencing	12	91						8/91 (9)	
Stein *et al* (2011)	Sequencing	12	166					31/166 (19)		
Bejar *et al* (2011)	NGS and mass spectrometry	All	439			63/439 (14)				
Grossman *et al* (2011a)	Sequencing	12	39						12/39[3] (31)	
Jankowska *et al* (2011)	Sequencing	12	72		7/20[4] (35)		24/52 (46)			
Metzeler *et al* (2011a)	Sequencing	12	423	44/423 (10)						
Paschka *et al* (2011)	Sequencing	12	1429	90/1429[5] (6)						
Martinez-Aviles *et al* (2012)	Sequencing	12	62					3/62 (5)		
Ricci *et al* (2012)	Sequencing	12	65					28/65[2] (43)		
Pratcorona *et al* (2012)	WAVE on cDNA	12[6]	836	35/775 (5)	7/61 (12)					
Brecqueville *et al* (2012)	Sequencing	12	149					17/149 (11)		
Devillier *et al* (2012)	Sequencing	12	48		17/48[7] (35)					
Traina *et al* (2012)	Sequencing	12	26					5/26[8] (19)		

Table 5.1 continued

Reference	Method of analysis	Exons analysed	Number of samples analysed	Incidence of *ASXL1* mutation (%)						
				De novo AML	Secondary AML	MDS	CMML	MPN	CML	JMML
Fernandez-Mercado *et al* (2012)	Sequencing	12	84	2/51 (4)	16/33 (48)					
Ibanez *et al* (2012)	HRM analysis	12	175	16/175 (6)						
Schnittger *et al* (2013)	Sequencing	12	740	127/740[5] (17)						
Wang *et al* (2013a)	Sequencing	12	153			33/153 (22)				
Itzykson *et al* (2013)	Sequencing	12	314				125/314 (40)			
Schwaab *et al* (2013)	Sequencing	12	39					8/39[8] (21)		
Hou *et al* (2014)	Sequencing	12	444	49/444 (11)						
Haferlach *et al* (2014)	Sequencing	12	944			221/944 (23)				
Krauth *et al* (2014)	Sequencing	12	139	16/139[9] (12)						
Guglielmelli *et al* (2014a)	Sequencing	12	166					54/166[2] (33)		
Damaj *et al* (2014)	Sequencing	12	43					6/43[8] (14)		
Tefferi *et al* (2014)	Sequencing	12	570					142/570[2] (25)		
Patnaik *et al* (2014)	Sequencing	12	420				164/420 (39)			
Guglielmelli *et al* (2014b)	Sequencing	12	797					203/797[2] (26)		

Table 5.1 continued

Reference	Method of analysis	Exons analysed	Number of samples analysed	Incidence of *ASXL1* mutation (%)						
				De novo AML	**Secondary AML**	**MDS**	**CMML**	**MPN**	**CML**	**JMML**
Renneville *et al* (2014)	Sequencing	12	226	21/226 (9)						
TOTAL			10884	259/2871[5] (9)	77/282[5] (27)	**MDS** 182/1090 (17)	**CMML** 388/968 (40)	**MPN** 511/2218 (23)	**CML** 26/225 (12)	**JMML** 5/117 (4)
				All AML 336/3153 (11)						

[1]Refractory anaemia with ringed sideroblasts in transformation. [2]Primary myelofibrosis. [3]CML in blast crisis. [4]AML with preceding CMML. [5]Breakdown of primary and secondary AML not given, excluded from subtotals. [6]Only first 325 bases of exon 12 examined. [7]AML with myelodysplastia-related changes. [8]Systemic mastocytosis. [9]t(8;21) AML.

Abbreviations: aCGH, array comparative genomic hybridisation; AML, acute myeloid leukaemia; BC-CML, chronic myeloid leukaemia in blast crisis; CML, chronic myeloid leukaemia; CMML, chronic myelomonocytic leukaemia; HRM, high resolution melting; MDS, myelodysplasia; MPN, myeloproliferative neoplasia; NGS, next generation sequencing; PMF, primary myelofibrosis; SM, systemic mastocytosis; SNP, single nucleotide polymorphism.

myeloid diseases as well as in 24 of 96 normal bone marrow samples. However, the findings from this study have not been replicated, for example, Grossmann *et al* (2012) analysed 491 normal control bone marrow samples and only one sample (0.2%) had this mutation, which they thought might be due to an undetected early myeloid disease. Other groups have also confirmed the validity of this mutation by using different proof-reading polymerases (Brecqueville *et al*, 2012), and analysing serial dilutions of the mutation with known wild-type samples to see if the mutant levels are as expected (Chou *et al*, 2010b). Furthermore, the consistency across groups in terms of incidence and correlation with other mutations has led most groups to consider the c.1934dupG mutation to be genuine (Gelsi-Boyer *et al*, 2012).

There is also conflicting evidence as to whether mutations in *ASXL1* have an effect on RNA expression levels. The gene expression profile from CD34-positive cells of 23 MDS and 9 CMML samples was investigated, and whilst there was variable *ASXL1* mRNA expression across the whole cohort, there was no difference in levels between the 6 mutated samples and the remaining wild-type samples (Gelsi-Boyer *et al*, 2009). Conversely, analysis of microarray results from 162 MDS samples of which 35 had nonsense or frameshift mutations, showed a significant reduction in gene expression of *ASXL1* in the mutant samples compared to the wild-type samples, with mean copy numbers of 5.8 and 7.6 respectively (p= 0.025) (Thol *et al*, 2011b). Furthermore, Western blot analysis of protein expression using both N- and C- terminal directed anitbodies in 2 primary AML samples with *ASXL1* mutations (one nonsense and one frameshift) showed reduced or absent expression (Abdel-Wahab *et al*, 2012).

5.1.3 Functional consequences of *ASXL1* mutations

To assess the effect of ASXL1 loss on the chromatin state, Abdel-Wahab *et al* (2012) used short hairpin RNAs, shRNAs, to silence *ASXL1* in the leukaemia cell line, UKE1. Chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq) data revealed a significant decrease in H3K27me3 transcriptional start site occupancy in ASXL1 knockdown cell lines compared to cell lines with an empty vector. Western blot analysis of purified histones from these two cell lines confirmed a significant reduction in H3K27me3 marks in the

knockdown cell line, but with preserved protein expression of the PRC2 complex members EZH2, SUZ12 and EED. Knockdown of *ASXL1* in primary CD34-positive cells, using small interfering RNAs, led to upregulation of homeotic genes including *HOXA5-9* as analysed by gene expression microarray (Abdel-Wahab *et al*, 2012). Furthermore, ChIP for EZH2 followed by qPCR of bound DNA at the HOXA locus in *ASXL1* knockdown cells revealed decreased HOXA product compared to empty vector cells. The authors concluded that these findings were consistent with the loss of ASXL1 causing a loss of PRC2 recruitment to the HOXA locus. This was confirmed by another group who investigated the effects of shRNA ASXL1 knockdown in CD34-positive cells (Davies *et al*, 2013). They noted impaired granulocytic differentiation in these cells, along with a significant increase in the number of multipotent mixed lineage colony-forming units. Gene expression profiling of these cell lines also revealed dysregulated gene expression of PRC2 targets.

Whilst initial investigations of *ASXL1* knock-out in mouse models showed some disruption in haematopoiesis, it did not trigger MDS or leukaemia (Fisher *et al*, 2010). However in this study, *ASXL1* was constitutively ablated by the introduction of a neomycin resistance cassette with several premature stop codons. This led to a high rate of perinatal lethality in the homozygous mutant embryos, which may have masked the haematological effects of knocking down *ASXL1*. Subsequent knock-out mouse models have displayed features of myelodysplasia (Abdel-Wahab *et al*, 2013; Inoue *et al*, 2013; Wang *et al*, 2014). Wang *et al* (2014) generated a knock-out model by the introduction of a cassette disrupting exon 1. Although they too found 80% embryonic lethality, in those mice that survived there were morphological features of myelodysplasia. Inoue *et al* (2013) used a BM transplant model, where cells were transduced with a vector containing either c.1934dupG or c.1900_1922del *ASXL1* constructs or mock transduced and these cells were then transplanted into sublethally irradiated mice. Morphological abnormalities in myeloid and erythroid cell lines were seen 12 months after transplant of the mutant vector, and these mice had a reduced median survival of 400 days compared to 2 years in the mock transduced mice. Abdel-Wahab *et al* (2013) used conditional allele targeting of the *ASXL1* gene to delete it specifically in the haematopoietic

compartment. Microscopic examination of blood revealed dysplastic myeloid and erythroid cells in the 6 month old *ASXL1* knock-out mice.

5.1.4 Incidence and characteristics of AML patients with *ASXL1* mutations

In the 14 studies investigating *ASXL1* mutations in AML, the overall incidence is 11% (Table 5.1). Mutations in this gene are more frequent in older patients. In a cohort of 501 patients with *de novo* AML, Chou *et al* (2010b) found *ASXL1* mutations in 54 patients (11%). The median age of those with the mutation was 66 years versus 49 years in the wild type patients (p <0.001). Similarly, Metzeler *et al (2011a)* found the mutation to be present in 16% of 234 patients ≥60 years but only in 3% of 189 patients <60 years (p <0.001). An association between *ASXL1* mutations and secondary AML has also been noted by several groups (Devillier *et al*, 2012; Fernandez-Mercado *et al*, 2012; Pratcorona *et al*, 2012; Rocquain *et al*, 2010; Schnittger *et al*, 2013). The total mutation rate in these papers for primary AML was 154 in 1606 patients (10%) compared to 53 in 185 patients with secondary disease (28%) (including AML with myelodysplastic changes, therapy-related, post-MDS and post-CMML AML) (p <0.001).

Initially, *ASXL1* mutations were thought to be mutually exclusive with *NPM1* mutations in AML (Carbuccia *et al*, 2010), however subsequent studies have shown that although there is a strong inverse correlation between the two mutations, they do rarely co-occur (Chou *et al*, 2010b; Fernandez-Mercado *et al*, 2012; Metzeler *et al*, 2011a; Pratcorona *et al*, 2012; Schnittger *et al*, 2013). The two largest studies found 10 out of a total of 181 *ASXL1* mutant samples to also have an *NPM1* mutation (Chou *et al*, 2010b; Schnittger *et al*, 2013). Given this strong inverse correlation, it is perhaps unsurprising that an inverse correlation has also been seen with *FLT3*/ITDs, and *DNMT3A* mutations, both of which frequently co-occur with *NPM1*, with these mutations present in only 3-10% and 4-15% of *ASXL1* mutant samples respectively (Chou *et al*, 2010b; Devillier *et al*, 2012; Fernandez-Mercado *et al*, 2012; Metzeler *et al*, 2011a; Pratcorona *et al*, 2012; Schnittger *et al*, 2013). Mutations in *ASXL1* have been associated with *RUNX1* mutations, as this mutation was present in 30% of *ASXL1* mutant samples compared to 10% in the *ASXL1* wild type samples (Chou *et al*, 2010b).

5.1.5 Impact of *ASXL1* mutations on outcome in AML patients

Some studies have suggested that a mutation in *ASXL1* is associated with a worse overall prognosis (Metzeler *et al*, 2011a; Paschka *et al*, 2011; Pratcorona *et al*, 2012; Schnittger *et al*, 2013), although this has not been corroborated by all groups (Ibanez *et al*, 2012; Shen *et al*, 2011). Others have found that, whilst presence of an *ASXL1* mutation was associated with a shorter overall survival in univariate analysis, it was not an independent risk factor in multivariate analysis when age, karyotype, *NPM1* mutation without *FLT3*/ITD, biallelic *CEBPA* mutations (Chou *et al*, 2010b) or age, karyotype, transplantation status were taken into account respectively (Patel *et al*, 2012).

Given the association of *ASXL1* mutations with older age and secondary disease, it is not clear whether knowing its status in a patient with newly diagnosed AML would provide sufficient additional information to significantly alter the expected outcome or mandate consideration of treatment modification.

This chapter outlines the screening of mutations in *ASXL1* exon 12 in 371 samples from four predefined groups of younger (15-59 years) and older (≥60 years）patients with primary or secondary AML, and relates genotype to outcome.

5.2 Patients, Materials and Methods

5.2.1 Patients

Genomic DNA screened was from diagnostic peripheral blood or bone marrow samples of 371 adult patients with AML entered onto the UK Medical Research Council (MRC) AML10 (n =178), AML11 (n=145) and AML12 (n=48) trials between 1988 and 2002. Patients were selected according to age and type of disease but not karyotype except those with acute promyelocytic leukaemia, who were excluded. All had known genotype for *FLT3*, *NPM1*, *IDH1* and *IDH2*. They were analysed in four pre-defined groups: 153 younger patients (15 – 59 years) with primary AML, 69 younger patients with secondary disease (40 post-MDS, 29 due to other causes including therapy-related AML), 116 older patients (60 or more years) with primary AML and 33 older patients with secondary AML (20 post-MDS, 13 other). All younger patients (n=222) were enrolled on the

AML10 and AML12 trials. Four of the older patients with secondary disease were enrolled on the AML12 trial; the remaining 145 older patients were enrolled on the AML11 trial. Ethical approval was obtained from the Multi-Centre Research Committee of Wales. Informed consent was obtained in accordance with the Declaration of Helsinki.

5.2.2 Screening for *ASXL1* exon 12 mutations by dHPLC

Mutation screening of *ASXL1* exon 12 was carried out by denaturing high performance liquid chromatography using the WAVE® DNA Fragment Analysis System (Transgenomic, Glasgow, UK) (see section 2.1.4). The entire coding sequence of *ASXL1* exon 12 was amplified by PCR using the proof-reading DNA polymerase, Optimase® (Transgenomic, Glasgow, UK) in 8 overlapping fragments, A-H (Figure 5.1). A further PCR product (A2) was used to confirm the presence or absence of the most commonly documented mutation (c.1934dupG). PCRs were performed as outlined in section 2.1.2. Details of all PCR primers and annealing temperatures are shown in Appendix 1. The presence of PCR products was confirmed on an agarose gel. To ensure that mutations associated with loss of heterozygosity due to chromosome (20q) deletion or uniparental disomy were not missed, the amplicons were mixed with known wild-type amplicons, denatured and cooled slowly to allow the formation of heteroduplexes on re-annealing. Resultant products were then analysed on the WAVE® at optimal melting temperatures (given in Appendix 1) calculated using Transgenomic Navigator software. A temperature was chosen if at least part of the amplicon was predicted to be approximately 75% helical. For most fragments, more than one temperature was used in order to examine the full amplicon, and *ASXL1* exon12 fragments A, A2 and E are given as examples in Figure 5.3.

5.2.3 Investigation of abnormal WAVE chromatograms

All samples with abnormal WAVE chromatograms, except fragment E, were sequenced using fresh PCR products. A common synonymous single nucleotide polymorphism, SNP rs 4911231, (c.3759T>C, p.S1253S) was detected in fragment E; the 1000 Genomes Project (2012) reports the T allele frequency to be 58% and C allele 42%. As both alleles are common, PCR

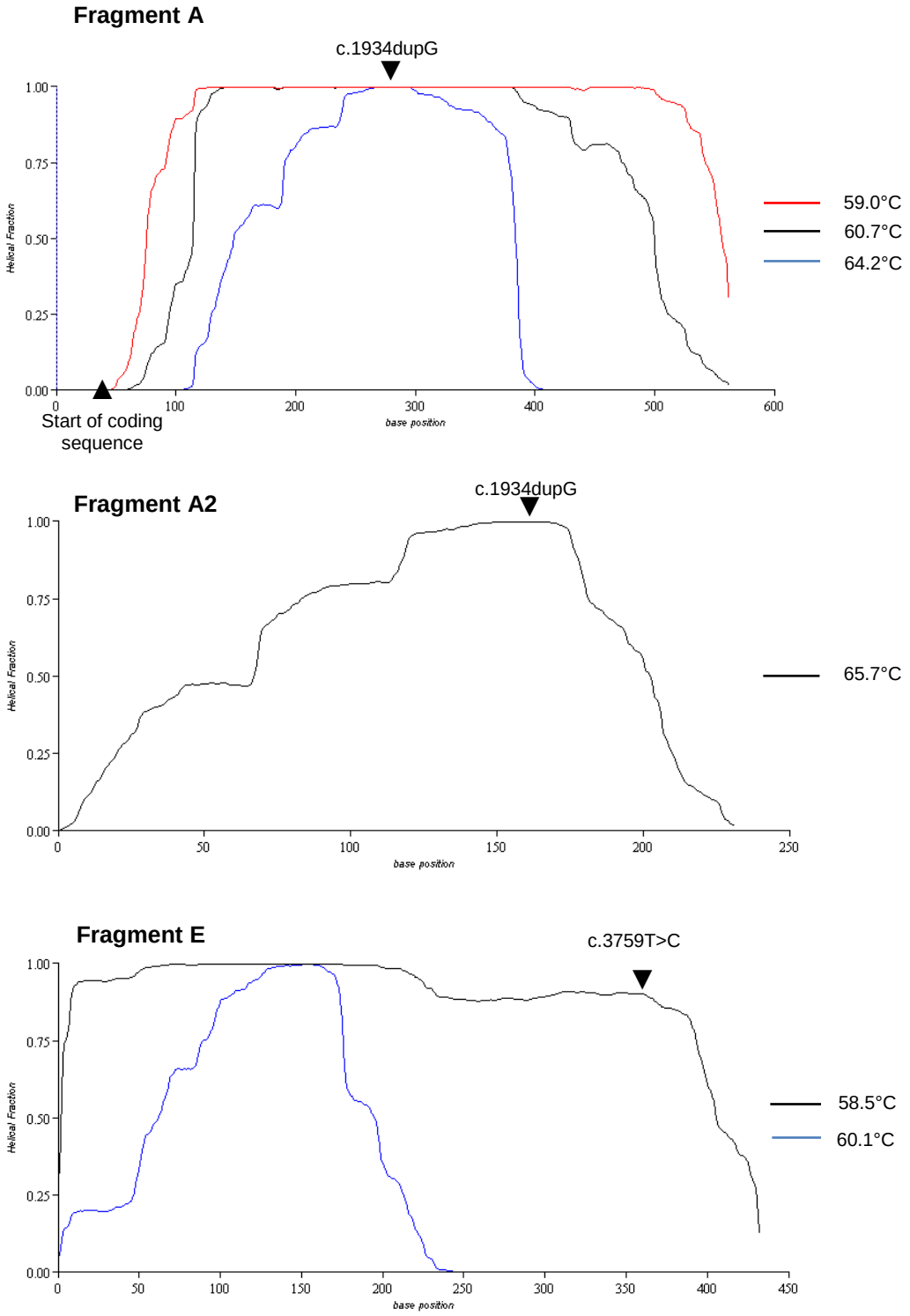


Figure 5.3 Representative amplicon melting curve profiles for WAVE® analysis. Examples of the melting curve profiles for 3 of the fragments analysed on the WAVE® at the selected temperatures. Predicted helical proportion is plotted against position in the amplicon. Downward arrows show site of the most common mutation, c.1934dupG in fragment A and A2, and the frequent synonymous SNP, c.3759T>C in fragment E. Upward arrow shows the start of the coding sequence in fragment A amplicon.

products for this fragment were initially analysed on the WAVE® unmixed, i.e. no wild-type amplicons added. Characteristic chromatograms were produced if samples were heterozygous for this SNP, thus if there was another mutation it should not be missed. If the fragment amplicon showed a single peak, it was rerun as a 3:1 mix with HL60 DNA (T/T at SNP site) to differentiate homozygous T/T from C/C and identify potential homozygous/ hemizygous mutations (Figure 5.4). To confirm that the abnormal chromatogram was due to the SNP, a PCR with a mismatch primer that introduced a restriction enzyme digestion site was created that would allow discrimination between the alleles (Figure 5.4C). PCR products were obtained using BioTaq DNA polymerase (Bioline, UK) (see section 2.1.2) and digested with *Bsr*1 (New England Biolabs (UK) Ltd., Hitchin, UK) for 2hrs at 65^{o}C, to give bands of 99+27 bp for T-alleles and a single uncut band of 126bp for C-alleles (Figure 5.4D).

Where sequencing of amplicons indicated more than one mutation within the same fragment, a repeat PCR was performed using BioTaq DNA polymerase (Bioline, UK) and the products were cloned using the TOPO TA cloning kit (Invitrogen, Paisley, UK) (section 2.1.9). Approximately 20 clones were harvested and grown up overnight and then PCR-amplified. At least four clones were then sequenced directly in order to assess whether the mutations were on the same allele.

Where available, follow-up samples for patients with a mutation in their presentation sample were analysed to assess whether the mutation persisted, using WAVE® analysis, sequencing or enzyme digestion as appropriate. Microsatellite analysis of 2 polymorphic markers, D11S554 and FES (primers in Appendix 1), was used to verify that the paired samples were from the same patient. Regions of the genome known to contain variable number tandem repeats were amplified to create PCR products of different fragment lengths which could be differentiated using fragment size separation on the CEQ 8000 Genetic Analysis System (Beckman Coulter, USA) (section 2.1.5). DNA samples derived from the same patient would produce the same length products.

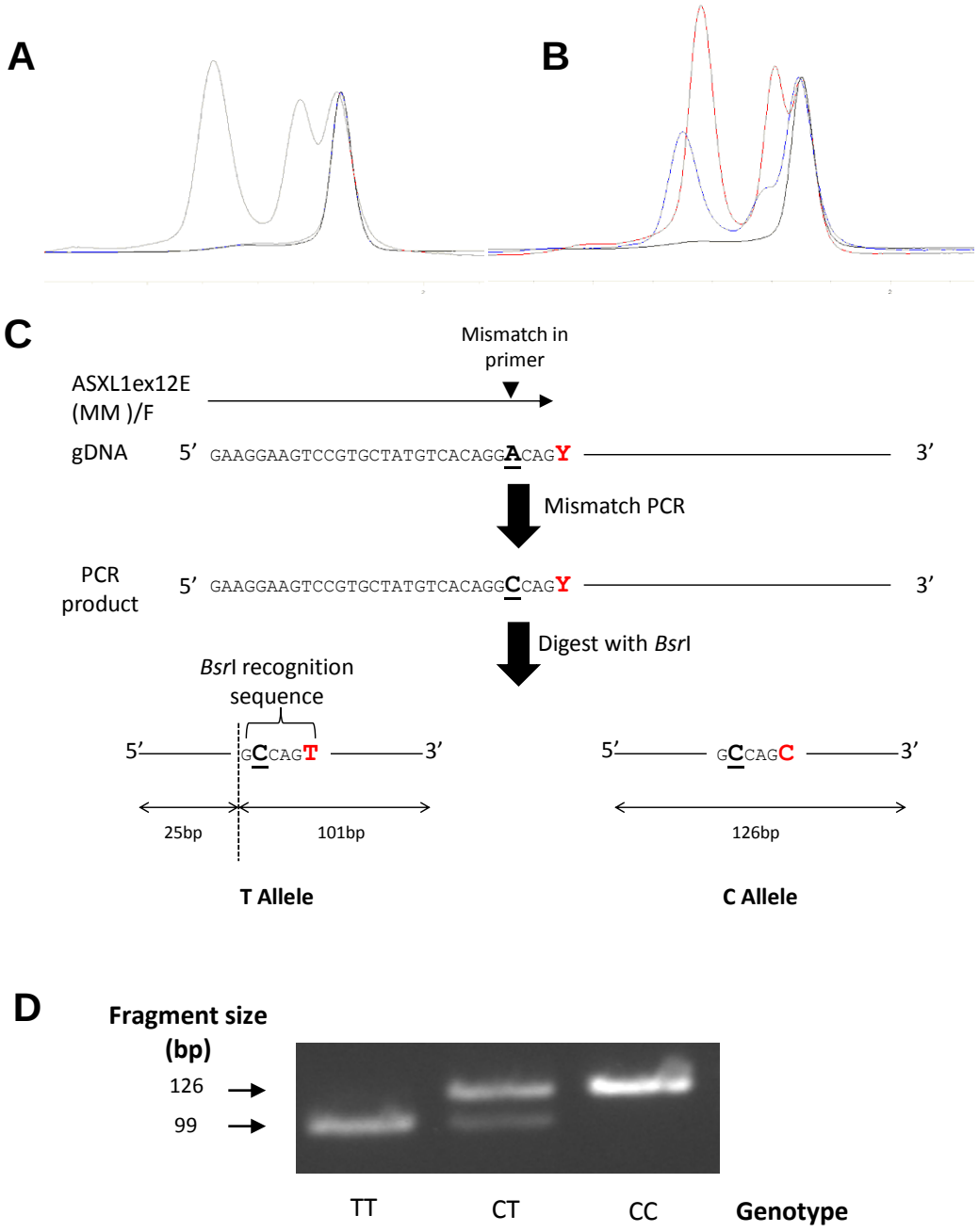


Figure 5.4 Identification and confirmation of the common synonymous SNP, c.3759T>C. Characteristic chromatograms at 58.5°C for samples that are homozygous T (black), heterozygous (red) and homozygous C (blue) for the SNP in (A) unmixed samples and (B) mixed with known TT product (HL60), to differentiate homozygous CC from TT. (C) The principle of the mismatch primer PCR and restriction digest to identify the base at the SNP site. The underlined base is the location of the mismatch (A>C) introduced in the primer and the base in red is the SNP under investigation. (D) An example of the confirmatory BsrI digest visualised on a 2% agarose gel stained with ethidium bromide.

5.2.4 Therapy

The MRC AML10 trial (Figure 5.5) recruited patients suitable for intensive chemotherapy, the majority of whom were under the age of 56, between 1988and 1995 (Hann *et al*, 1997). At diagnosis, patients were randomised to receive two cycles of either DAT or ADE induction chemotherapy. Both regimens included daunorubicin on days 1, 3 and 5 and intravenous cytarabine every 12 hours for 10 days in the first cycle, and 8 days in the second. The difference between the two arms was that DAT entailed 6-thioguanine given orally twice daily for 10 days in the first cycle and 8 days in the second cycle, whereas ADE involved intravenous etoposide for 5 days per cycle. After induction, if patients had achieved complete remission (CR) they were given two further cycles of consolidation chemotherapy consisting of MACE then MIDAC. For those patients with no HLA-matched sibling, there was a second randomisation between further high dose therapy with autologous stem cell transplant (SCT) versus no further therapy.

The MRC AML12 trial (Burnett *et al*, 2010) recruited patients from 1994-2002 who were suitable for intensive chemotherapy (Figure 5.5). Randomisation at induction was between ADE (given as in AML10) or MAE chemotherapy (as for ADE but with mitoxantrone given on days 1,3 and 5 rather than daunorubicin). Halfway through the trial there was a protocol amendment, all patients were given DAT chemotherapy for induction with either high dose (200mg/m^2) or standard dose (100mg/m^2) cytarabine. Patients were also randomly assigned to receive all trans retinoic acid (ATRA) or not. If CR was reached after induction chemotherapy, patients had a cycle of MACE chemotherapy followed by a second randomisation to either ICE, MiDAC, SCT (allogeneic if available, otherwise autologous) or ICE followed by SCT as a fifth cycle of treatment.

AML11 was designed for older patients with AML or those not suitable for intensive chemotherapy and was open from 1990 to 1998 (Goldstone *et al*, 2001) (Figure 5.5). The first randomisation was to either have two cycles of DAT (first cycle as in AML10 trial and second consisting of daunorubicin on days 1 and 3, and cytarabine and 6-thioguanine for 5 days rather than 8), ADE (first cycle as in AML10, second cycle also had two doses of daunorubicin and 5 days of cytarabine) or MAC. Patients in remission after two cycles had a

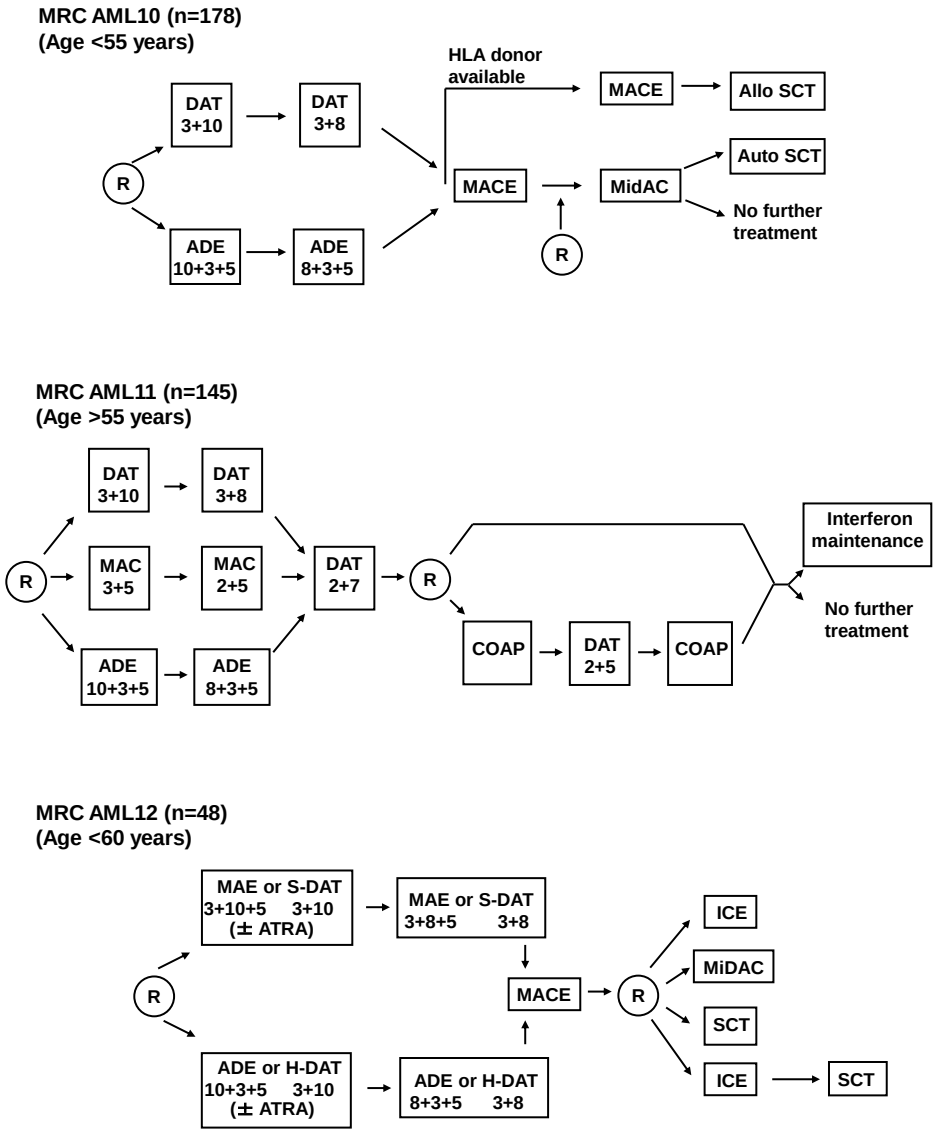


Figure 5.5 Outline of the relevant treatment protocols for patients in Medical Research Council trials AML10, AML11 and AML12.

Abbreviations: DAT, daunorubicin + cytarabine + 6-thioguanine (in AML12, patients received either a standard (S) or high (H) dose of cytarabine); ADE, cytarabine + daunorubicin + etoposide; MACE, amsacrine + cytarabine + etoposide; MidAC, mitoxantrone + cytarabine; MAC, mitoxantrone + cytarabine; COAP, cyclophosphamide + vincristine + cytarabine + prednisolone; MAE, mitoxantrone + cytarabine + etoposide; ATRA, all trans retinoic acid; ICE, idarubicin + cytarabine + etoposide; R, randomisation; SCT, stem cell transplantation

consolidation course of DAT 2+7. There were two further randomisations at this stage, firstly either stopping treatment after the third course versus continuing to have 6 courses of chemotherapy (COAP, DAT and COAP being the last 3 cycles). The second randomisation was to have IFN-α maintenance for a year or not.

5.2.5 End points

CR was defined as a normocellular bone marrow with less than 5% blasts, showing evidence of normal maturation, although persistent myelodysplastic features did not preclude CR being achieved. Patients who failed to achieve remission were either clinically classified as induction death (ID), related to treatment and/ or hypoplasia, or resistant disease (RD) due to persistence of the leukaemia despite treatment. Overall survival (OS) was measured from the point of randomisation until death. Cumulative incidence of relapse (CIR) was defined as the incidence of relapse following CR, with death as a competing risk.

5.2.6 Statistical methods

Patient information and outcome data were analysed by Dr Robert Hills, Department of Haematology, University of Cardiff. Samples with missense mutations of unknown significance were excluded from outcome analysis. Mantel-Haenszel and χ^2 tests were used to test for differences in demographic and clinical data by genotype. Kaplan – Meier curves were constructed for survival data and compared by means of the log-rank test, with standard tests for heterogeneity between subgroups. Surviving patients were censored at 9 August 2010, with follow-up complete to this date for 85% of patients. Median follow-up for survival was 18.4 years (range 5.2 – 22.3 years). Multivariate logistic regression analysis was used to find the factors most closely associated with complete remission (CR) rate, and multivariate Cox models were used to analyse CIR and OS. Models were fitted using forward selection, with variables added to the model if they had a *p* value, derived using the deviance statistic, of less than 0.05. Odds ratios (ORs) or hazard ratios (HRs) and 95% confidence intervals (CIs) are quoted for endpoints. In all cases a ratio of <1 indicates

benefit for a mutation. All *p* values are two-tailed and all end points are given at 5 years.

5.3 Results

5.3.1 Detection of sequence alterations by WAVE analysis

Of the 371 patient samples analysed, abnormal dHPLC chromatograms were detected in one or more fragments from 249 (67%) patients. All abnormal chromatograms were confirmed and the alteration identified by sequencing or restriction digestion of a fresh PCR product. Examples of abnormal chromatograms are given in Figure 5.6. There were 230 samples with an abnormal chromatogram in fragment E consistent with the genotype CC or CT at the common synonymous SNP site, c.3759T>C (p.S1253S). The genotype at this site was investigated for all samples by restriction enzyme digest (see section 5.2.3), an example of which is given in Figure 5.4. There was 100% concordance between the results from the WAVE® and the restriction digest. The overall frequency was 38% patients with TT, 16% with CC and 46% with CT, consistent with the expected percentages based on the 1000 Genome Project of 33%, 18% and 49% respectively.

Excluding abnormal chromatograms explained by c.3759T>C, there were 59 samples with at least one abnormal chromatogram and the appropriate fragment for each sample was sequenced to identify the abnormality. A single synonymous mutation was seen in 6 samples, these SNPs were considered to be wild-type (WT). Missense alterations were seen in 22 patient samples, of which 1 had two separate alterations, 3 had at least one synonymous substitution as well, and 1 also had a frameshift mutation. These changes were compared to published data and the SNP database (dbSNP, NCBI, Bethesda, USA) and, where available, remission samples or DNA from CD3+ cells were analysed for presence of the substitution that had been detected in the diagnostic sample. The remission sample for a patient who had the missense mutation p.G643V in their diagnostic sample was analysed by sequencing and the mutation was shown to persist. Two patients with remission samples available had a p.E1102D alteration which had been seen in paired fibroblasts

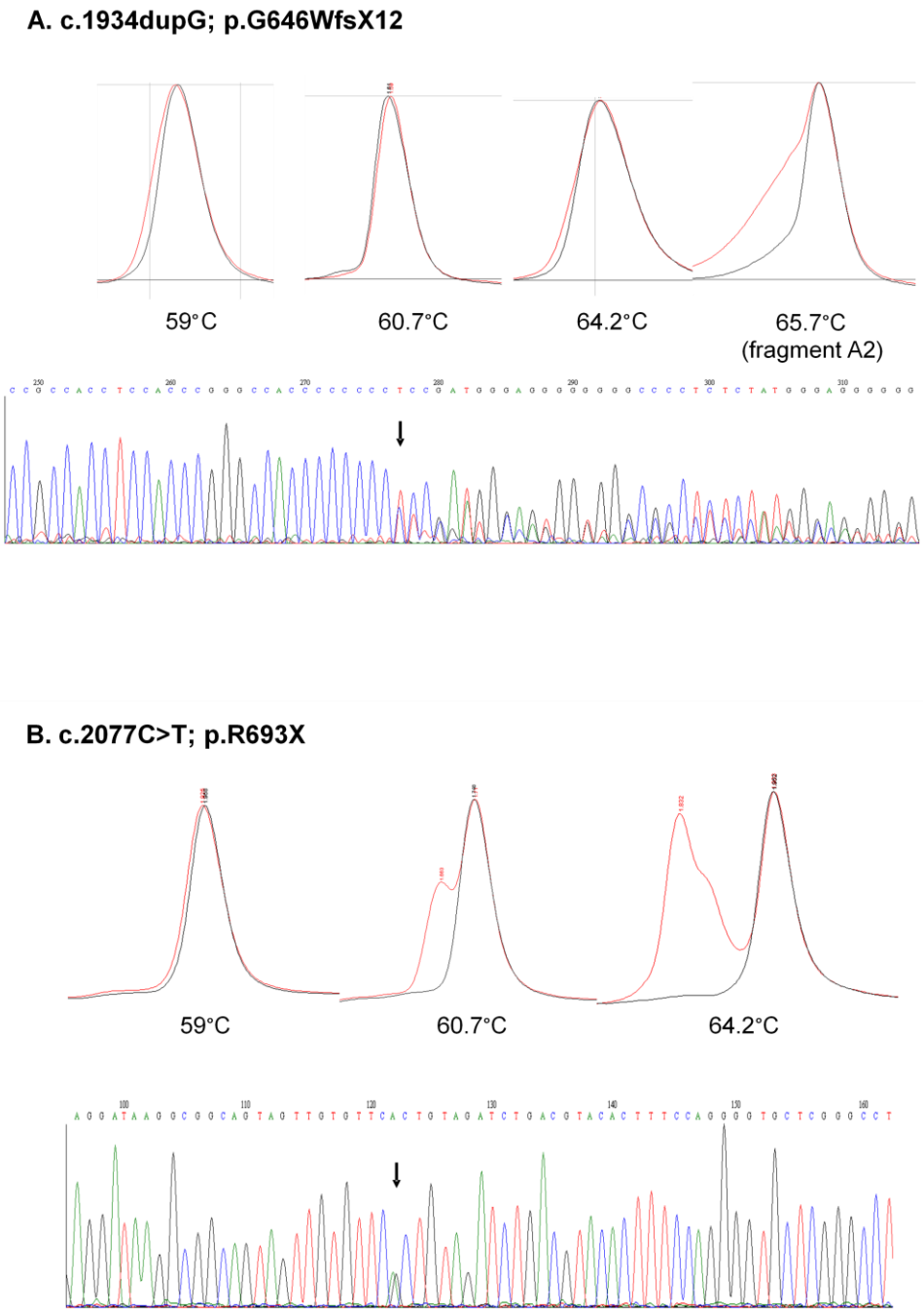


Figure 5.6 Representative analysis of two *ASXL1* mutations. In each case, the upper panel shows the WAVE® chromatogram for fragment A at the specified temperatures. The patient sample (red) is normalised and compared to the wild-type control (black). Fragment A2 (designed to pick up c.1934dupG) is also shown for example A. The lower panel for each example is the sequencing results for the patient, with the abnormality highlighted with an arrow.

(A) The most common mutation seen, c.1934dupG predicted to give rise to p.G646WfsX12. (B) The nonsense mutation, c.2077C>T predicted to produce p.R693X.

in a previous study (Perez *et al*, 2010). PCR products for fragment D were digested with *Hae*III, which cuts the “E” allele twice to create 3 fragments of 60, 213 and 271 bases, but only cuts the “D” allele once to create fragments of 213 and 331 bases; the fragments were then separated on agarose gel. In both cases, fragments of sizes 213, 271 and 331 bases were seen with the remission samples consistent with persistence of both alleles. The alteration p.S1231F was also analysed by restriction digest of the PCR product for fragment E using *Hinf*I enzyme. If the “S” allele was present, the digest created 3 fragments of 57, 170 and 206 bases and the “F” allele, 2 fragments of 206 and 227 bases. The follow-up sample showed persistence of the S1231F heterozygosity. This suggested that all three missense alterations (p.G643V, p.E1102D, p.S1231F) were germline in origin. Consequently, 17 of the missense mutations were either previously reported to be SNPs and/ or shown to be present in paired normal tissue and were scored as WT (p.G643V, p.G652S, p.G704R, p.P779L, p.E1102D, p.S1231F; M1249V, p.A1312V, p.G1397S), (Table 5.2). Four patients without available paired tissue had missense mutations of unknown significance (p.A636V, p.C687R, p.E1015G plus p.S1099C, p.P1134L). They were scored as WT for mutation incidence, and were excluded from the outcome analysis. The PCR product for Fragment D for the sample with both p.E1015G and p.S1099C missense alterations was cloned and used to transform One Shot DH5α cells (see section 2.1.9). Four colonies were picked and sequenced and showed that the alterations were on the same allele.

Of the remaining 32 patient samples with at least one abnormal chromatogram, 23 had frameshift mutations and 9 had nonsense mutations (Table 5.3) (Figure 5.7). Thirteen of the 32 patients (41%) had the most commonly reported frameshift mutation p.G646WfsX12. One patient had two separate frameshift mutations (p.G645VfsX58 and p.G645WfsX12), another had both a frameshift (p.G646WfsX12) and a missense mutation (p.E865K). Six of these patients had follow-up samples available. Microsatellite analysis (see section 2.1.5) using two polymorphic markers was performed and confirmed that each pair of diagnostic and follow-up samples originated from the same patient. Four of the six had the most frequent mutation, p.G646WfsX12. Three of these four follow-up samples were WT by both WAVE® analysis and direct sequencing. In the

Table 5.2 *ASXL1* missense mutations considered to be SNPs

DNA Change	Predicted AA change	PCR fragment	Number of Cases	Reason for considering as SNP
c.1928G>T	p.G643V	A	1	Point mutation persisted in follow-up remission sample
c.1954G>A	p.G652S	A	1	SNP rs3746609
c.2110G>A	p.G704R	A	1	Point mutation persisted in follow-up remission sample
c.2336C>T	p.P779L	B	1	SNP rs41289850 Point mutation persisted in follow-up remission sample
c.3306G>T	p.E1102D	D	9	Point mutation persisted in follow-up remission sample Mutation present in paired fibroblasts (Perez *et al*, 2010)
c.3692C>T	p.S1231F	E	1	SNP rs74638057 Mutation present in paired CD3 sample (Sugimoto *et al*, 2010)
c.3745A>G	p.M1249V	E	1	Point mutation in paired normal tissue (Abdel-Wahab *et al*, 2011)
c.3935C>T	p.A1312V	F	1	Point mutation in paired normal tissue (Bejar *et al*, 2011)
c.4189G>A	p.G1397S	F	1	Point mutation in paired normal tissue (Bejar *et al*, 2011)

Table 5.3 Frameshift and nonsense mutations detected in *ASXL1* exon 12

DNA Change	Predicted AA change	Number of cases in patients <60 years		Number of cases in patients ≥60 years		Total number of cases
		De novo (n=153)	Secondary (n= 69)	De novo (n=116)	Secondary (n=33)	(n=371)
WILD TYPE		147	61	103	28	339
c.1773C>A	p.Y591X	0	0	1	1	2
c.1873C>T	p.R625X	0	0	1	0	1
c.1900_1922del	p.E635RfsX15	1	1	2	0	4
c.1926delA/ c.1926_7delAG	p.G645VfsX58/ p.G645WfsX12	1	0	0	0	1
c.1934dupG	p.G646WfsX12	1	2	7*	2	12
c.2077C>T	p.R693X	0	2	0	1	3
c.2081delC	p.T694NfsX9	0	0	0	1	1
c.2122C>T	p.Q708X	0	0	1	0	1
c.2179delG	p.E727RfsX17	0	0	1	0	1
c.2205delT	p.T736QfsX8	0	1	0	0	1
c.2291_2300del	p.L764PfsX5	0	1	0	0	1
c.2295_2299del	p.S766PfsX6	0	1	0	0	1
c.2356A>T	p.R786X	1	0	0	0	1
c.2388G>A	p.W796X	1	0	0	0	1
c.2751_2761del	p.P920CfsX2	1	0	0	0	1
TOTAL NUMBER OF MUTANT CASES		**6 (4%)**	**8 (12%)**	**13 (11%)**	**5 (15%)**	**32 (9%)**

* One patient with c.1934dupG also had c.2593G>A (p.E865K) with a remission follow-up sample that was wild-type for both mutations.

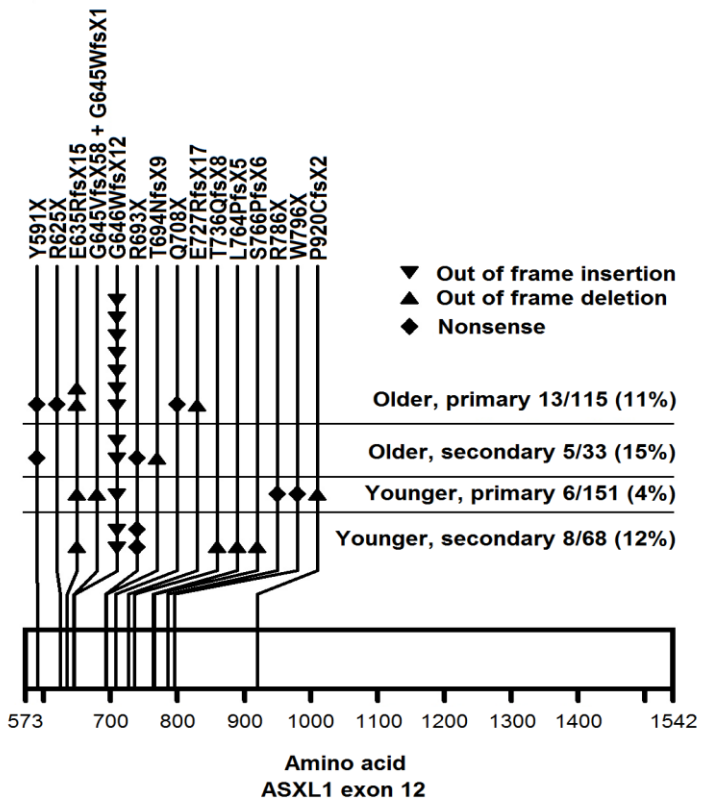


Figure 5.7 Distribution, incidence and type of truncating ASXL1 exon 12 mutations detected in the 4 patient groups

remaining sample, although the mutation could be seen on direct sequencing it was at a much lower level than the diagnostic sample. One of these patients also had a missense mutation p.E865K, which was not present on sequencing of the remission sample. One patient with the mutation p.Y591X and another with the two frameshift mutations p.G645VfsX58 and p.G645WfsX1 also had remission samples, which were WT on sequencing. Hence overall, 32 patients (9%) had mutations of likely pathological significance that would lead to loss of the PHD finger.

5.3.2 Patient characteristics according to *ASXL1* genotype

A summary of patient characteristics according to *ASXL1* genotype is given in Table 5.4. *ASXL1*-mutated patients (*ASXL1*^{MUT}) had a trend for a higher presenting white cell count than *ASXL1* WT patients (*ASXL1*^{WT}) (median 42.8 versus 25.9 x10^{9}/l respectively, *p*= 0.08) but were not associated with cytogenetic risk groups as defined by the MRC classification (Grimwade *et al*, 1998), or FAB subtype. There was a marked difference in the incidence of mutations according to age, 6% in the younger and 12% in the older age groups (*p*=0.05), and *ASXL1*^{MUT} patients were significantly older than *ASXL1*^{WT} patients (median 61.5 versus 51 years, *p*=0.01). They were also more likely to have secondary than primary disease (13% versus 7%), although this did not reach statistical significance (*p*=0.08). In younger patients with primary disease, the incidence was 4% compared to 12% in those with secondary disease (*p*=0.03), but in older patients the difference between primary and secondary disease was not significant (11% versus 15%, *p*=0.5).

Presence of an *ASXL1* mutation was inversely correlated with an *NPM1* mutation (*p*=0.0008), although they were not mutually exclusive (9% of *ASXL1*^{MUT} cases were *NPM1*^{MUT}), and showed a trend for association with *IDH2* mutations (*p*=0.05), of which 57% (4 of 7) were *IDH2*$^{MUT\text{-}R140}$ and 43% (3 of 7) *IDH2*$^{MUT\text{-}R172}$. There was no association with *FLT3* or *IDH1* mutations (Table 5.5).

5.3.3 Clinical outcome according to *ASXL1* genotype

Correlating the outcome data of the whole cohort with *ASXL1* status revealed no difference in response to chemotherapy between the two groups (Table 5.6).

Table 5.4 Patient demographics according to *ASXL1* mutant status

	Total No.	ASXL1-WT No. (%)	ASXL1 mutant No. (%)	*p*
Patients	367	335 (91%)	32 (9%)	
Age, years				0.01*
15-29	43	42 (13%)	1 (3%)	
30-39	36	34 (10%)	2 (6%)	
40-49	82	81 (24%)	1 (3%)	
50-59	58	48 (14%)	10 (31%)	
60-69	101	87 (26%)	14 (43%)	
70+	47	43 (13%)	4 (13%)	
Median (range)	52 (16-80)	51 (16-80)	61.5 (19-74)	0.01**
Sex				0.11†
Female	187	175 (52%)	12 (38%)	
Male	180	160 (48%)	20 (63%)	
Diagnosis				0.08†
de Novo	266	247 (73%)	19 (59%)	
Secondary	101	88 (26%)	13 (41%)	
WBC, X 10^9/L				
Median (range)	27.4 (0.4-528.0)	25.9 (0.4-349.0)	42.75 (3.0-528.0)	0.07**
Cytogenetics				0.7†
Favourable	11	10 (4%)	1 (3%)	
Normal Karyotype	153	139 (54%)	14 (61%)	
Other Intermediate	66	60 (23%)	6 (26%)	
Adverse	49	47 (18%)	2 (9%)	
Unknown	88	79	9	
FAB Type				0.2†
M0	6	6 (2%)	0	
M1	75	70 (22%)	5 (18%)	
M2	103	97 (31%)	6 (21%)	
M4	101	91 (29%)	10 (36%)	
M5	37	30 (30%)	7 (25%)	
M6	9	9 (3%)	0	
M7	4	4 (1%)	0	
RAEB-t	7	7 (2%)	0	
Other/Unknown	25	21	4	

*Test for trend; † test for heterogeneity, ** Wilcoxon Rank Sum Test

Table 5.5 Correlation of *ASXL1* mutation status with other mutations

	Total No.	ASXL1-WT No. (%)	ASXL1 mutant No. (%)	*p*
FLT3/ITD				
Wild-type	284	258 (77%)	26 (81%)	0.6†
Mutant	82	76 (23%)	6 (19%)	
FLT3/TKD				
Wild-type	331	300 (90%)	31 (97%)	0.18†
Mutant	36	35 (10%)	1 (3%)	
NPM1				
Wild-type	232	203 (61%)	29 (91%)	0.0008†
Mutant	135	132 (39%)	3 (9%)	
ITD/NPM1				
Wild-type/Wild-type	197	174 (52%)	23 (72%)	0.005†
Wild-type/Mutant	87	84 (25%)	3 (9%)	
Mutant/Wild-type	35	29 (9%)	6 (19%)	
Mutant/Mutant	47	47 (14%)	0	
IDH1 WT	323	295 (88%)	28 (88%)	0.9†
IDH1 Mutant	44	40 (12%)	4 (13%)	
IDH2 WT	325	300 (90%)	25 (78%)	0.05†
IDH2 Mutant	42	35 (10%)	7 (22%)	

† test for heterogeneity

Table 5.6 Outcome of patients with and without a mutation in *ASXL1 exon 12.*

Response to therapy	*ASXL1*^{WT}	*ASXL1*^{MUT}	Univariate OR or HR, CI	*p*	Multivariate OR or HR, CI*	*p*
CR (with recovery)	59%	56%	1.11 (0.53-2.32)	0.8	0.47 (0.16-1.34)	0.15
CR/Cri	66%	56%	1.53 (0.72-3.28)	0.3	1.06 (0.35-3.20)	0.9
RD	20%	25%	1.39 (0.56-3.42)	0.5	0.85 (0.28-2.60)	0.8
ID	14%	19%	1.42 (0.51-3.95)	0.5	0.89 (0.16-5.09)	0.9
Outcome at 5 years						
CIR	56%	83%	2.41 (1.19-4.89)	0.01	1.61 (0.85-3.06)	0.14
OS	22%	6%	1.66 (1.07-2.59)	0.02	1.27 (0.80-2.02)	0.3

CR, complete remission; CRi, complete remission with incomplete haematological recovery; RD, resistant disease; ID, induction death; CIR, cumulative incidence of relapse; OS, overall survival; OR, odds ratio; HR, hazards ratio
*Adjusted for age, sex, white blood cell count, performance status, cytogenetic risk group, *FLT3*/ITD, *FLT3*/TKD, *NPM1*, *IDH1* and *IDH2* mutant status

CR rate, with or without complete haematological recovery, was 66% versus 56% in the wild-type and mutant groups respectively (OR 1.53, 95% CI 0.72-3.28, *p*= 0.3). However, the *ASXL1*^{MUT} group had a significantly worse CIR at 5 years, 89% versus 56% for the *ASXL1*^{WT} patients (HR 2.41, 95% CI 1.19- 4.89, *p*= 0.01) (Figure 5.8A) and OS at 5 years, 6% versus 22% respectively (HR 1.66, 95% CI 1.07- 2.59, *p*= 0.02) (Figure 5.8B).

Analysing outcome data in subgroups of the cohort was limited by the small numbers of patients. However, when considered in the four pre-defined groups, in the younger patients, CIR was significantly higher in the *ASXL1*^{MUT} group with secondary disease (100% versus 50%, *p*= 0.04) but not primary disease (75% versus 43%, *p*= 0.6), although there were only four *ASXL1*^{MUT} patients in each group that achieved remission. OS was lower for patients with *ASXL1*^{MUT} in both younger cohorts but the difference was not significant (secondary disease, 0% versus 18%, *p*= 0.2; primary disease, 17% versus 32%, *p*= 0.7) (Figure 5.9 A and B). In the older patients, there was no difference according to mutant status in the groups for either CIR (secondary disease, 50% versus 92%, *p*= 0.4; primary disease, 88% versus 74%, *p*= 0.1) or OS (secondary disease, 20% versus 4%, *p*= 0.7; primary disease, 0% versus 14%, *p*= 0.2), although again there were only 8 and 6 *ASXL1*^{MUT} patients respectively in the two groups (Figure 5.9 C and D).

In multivariate analysis adjusted for age, sex, white cell count, performance status at diagnosis, cytogenetic risk group, *FLT3*/ITD, *FLT3*/TKD, *NPM1*, *IDH1* and *IDH2* genotype, *ASXL1* mutations lost any prognostic significance because age was a major confounding risk factor (Table 5.6).

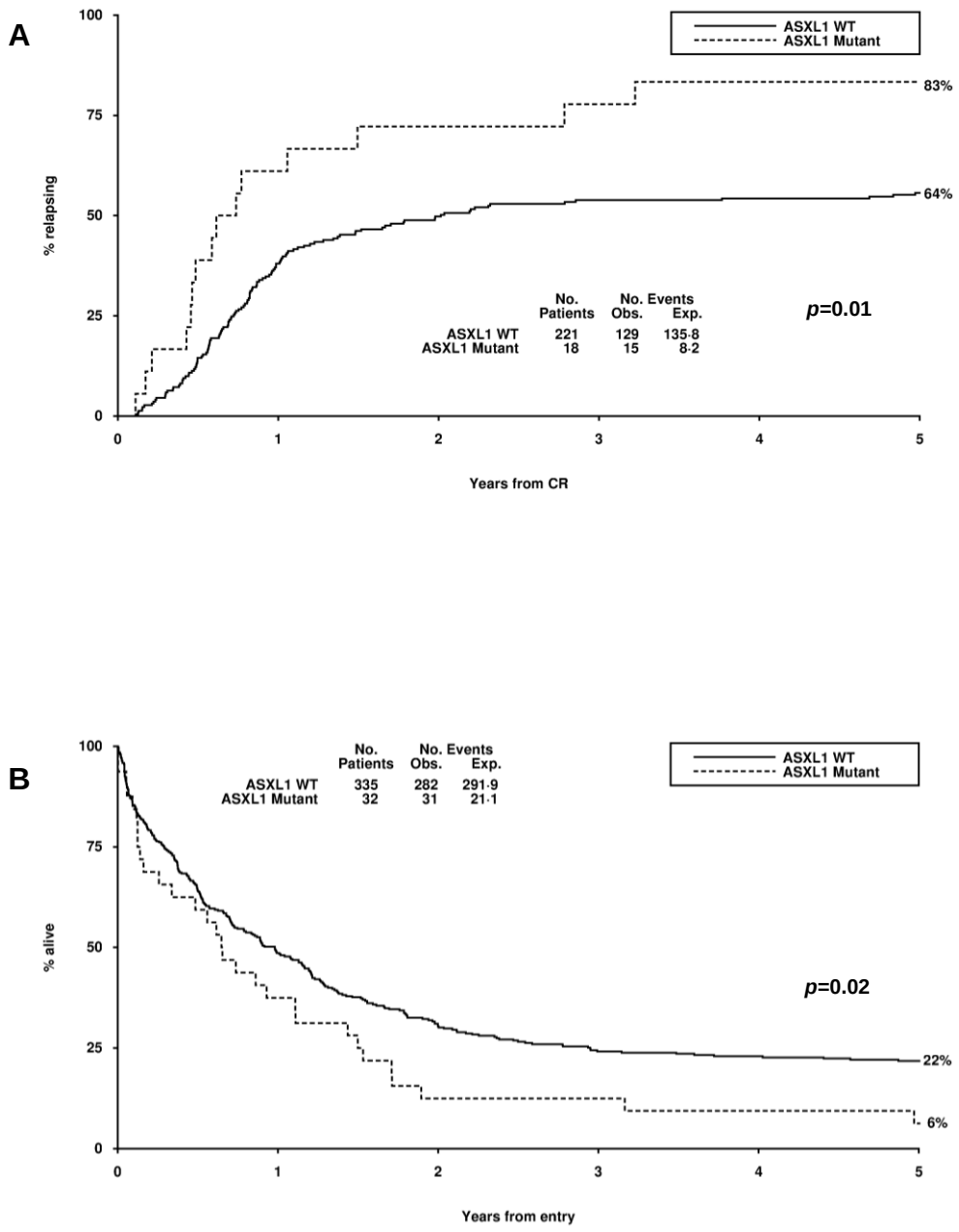


Figure 5.8 Outcome according to ASXL1 mutant status in the total cohort.
(A) Cumulative Incidence of Relapse, (B) Overall survival.

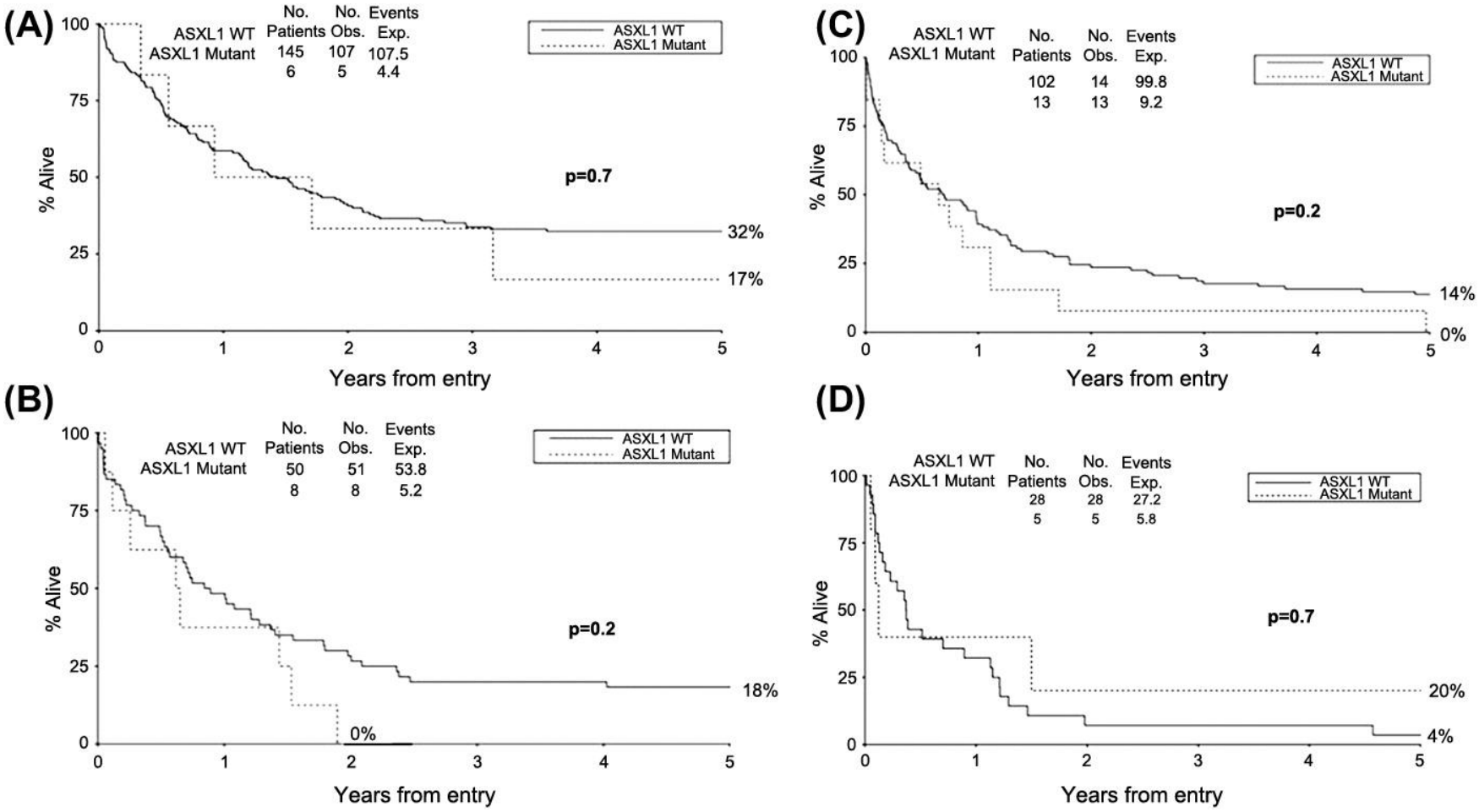


Figure 5.9 Overall survival in the different patient groups. Younger patients with (A) primary and (B) secondary disease. Older patients with (C) primary and (D) secondary disease.

5.4 Discussion

Abnormalities in genes affecting the “epigenome”, including *ASXL1*, are now recognised to be mutated in the spectrum of myeloid disorders. Although a number of papers have suggested that presence of an *ASXL1* mutation is associated with a worse overall survival, not all agree and therefore controversy exists as to whether knowledge of the *ASXL1* genotype at diagnosis would influence management. In order to assess this, the mutation status of four defined groups of patients, categorised according to age and primary or secondary disease status, was analysed and correlated with outcome data.

Most studies have used sequencing to detect mutations but the WAVE® was used in this study, as it provided a faster method of screening for point mutations as well as size changes, with minimal processing of the PCR product required. Furthermore, this medium throughput method has a higher sensitivity than sequencing for detecting mutations. To ensure that mutations were not missed, as deletions in 20q are found in 1% of patients with primary AML and 8% secondary AML (Grimwade *et al*, 2010), the amplicons under investigation were mixed with known wild-type amplicons to create heteroduplexes. The investigation was restricted to exon 12 because those studies that looked at the whole gene found mutations in other exons in only 1% of samples. Some studies have limited the region analysed further to the first 385 bases of this exon (Pratcorona *et al*, 2012) or just screened for size changes using a more rapid technique (Paschka *et al*, 2011), but in our study this would have missed 38% and 28%, respectively, of our mutant-positive patients. Despite limiting analysis to a single exon, due to its size of 2907 bases, it still required 8 overlapping PCR fragments to cover the coding region, because the optimal amplicon size for WAVE® analysis is 200-450 bases, although amplicons up to 1000 bases can be assessed.

The characteristic WAVE® chromatogram in fragment A at 64.2°C for the most common mutation c.1934dupG was a very subtle widening of the peak plus a slight shoulder, this is because fragment A was a relatively large product of 563 bases and the duplication occurred in a part of the amplicon predicted to have a high helical fraction within an 8 base homopolymer run (Figure 5.6A). Hence to avoid missing the abnormality, a smaller amplicon A2 of 232 bases was

analysed that displayed a more exaggerated shoulder on the chromatogram in samples with the c.1934dupG mutation (Figure 5.6A). Secondly, given the concern that this mutation may be a chance PCR artefact as it is a duplicated G within an 8 base guanine homopolymer run, a second amplicon confirming the duplication gave further support for it being a true mutation. Further confirmation that this is a true mutation was sought by analysing follow-up samples. Three of the 4 patients with the duplication present in the diagnostic sample no longer had it in their remission sample, and in the remaining sample it was at a much lower level, consistent with this being a true mutation and not a random error during the amplification process. This is in agreement with most other groups (Gelsi-Boyer *et al*, 2012).

Overall, 249 of the 371 patient samples analysed had abnormal chromatograms in at least one fragment. Most (n=217, 87%) of the abnormalities consisted of either synonymous or missense alterations. Synonymous changes were considered wild-type as there was no predicted change in amino acid. Of the 23 missense mutations seen in 22 patient samples, 18 were considered somatic as they were seen in normal controls or paired normal tissue by us or other groups. Remission samples were not available for 3 samples with missense mutations that had not been reported previously, and although considered to be wild-type for mutation incidence, we excluded these from outcome analysis. Only one missense mutation was detected that was not present in a follow-up remission sample, however, this sample also had a frameshift mutation and thus was considered mutant. Hence, the overall mutation rate of *ASXL1* exon 12 was 9%, and this is in keeping with other studies in AML (Table 5.1).

The mutation rate was significantly higher in patients older than 60 years compared to those less than 60 years, 12% versus 6% respectively (p= 0.05). In those patients under 60 years, the mutation rate was significantly higher in secondary disease compared to primary, 12% versus 4% respectively (p= 0.03). Hence in patients younger than 60 years with primary disease, the overall mutation rate was 4%. The association of *ASXL1* mutations with older age and secondary disease seen in our cohort has also been seen by other groups (reviewed in Gelsi-Boyer *et al*, 2012).

There was no association between karyotype and *ASXL1* genotype, although it has been reported that *ASXL1* mutations are more frequent in intermediate-risk

patients with an aberrant karyotype than those with a normal karyotype (Schnittger *et al*, 2013). *ASXL1* mutations were inversely correlated with *NPM1* mutations, only 9% of those with an *ASXL1* mutation also had an *NPM1* mutation compared to 39% of the *ASXL1* WT patients (p <0.001). This strong inverse correlation has previously been reported (Carbuccia *et al*, 2010; Paschka *et al*, 2011; Schnittger *et al*, 2013).

In the total cohort, *ASXL1*^{MUT} patients had a poorer outcome with significantly worse OS, which is compatible with the adverse impact seen in most studies (Chou *et al*, 2010b; Metzeler *et al*, 2011a; Paschka *et al*, 2011; Patel *et al*, 2012; Pratcorona *et al*, 2012; Schnittger *et al*, 2013). We could not attribute this to a higher primary resistance of *ASXL1*^{MUT} patients to chemotherapy, as has been suggested (Chou *et al*, 2010b; Pratcorona *et al*, 2012), as we observed no difference in the rates of remission and resistant disease according to *ASXL1* genotype in our patients, but the relapse rate was significantly increased in *ASXL1*^{MUT} patients. However, this analysis included older and secondary disease patients, characteristics that are likely to impact on outcome. When stratified according to age and type of disease, we found no impact on outcome in the younger patients with primary disease, although with small numbers of mutant-positive patients some caution is required in the interpretation of these data, and no differences in outcome were found in the older patients according to type of disease. Furthermore, *ASXL1* mutations were not a significant factor in multivariate analysis because of the overriding impact of age. Only one other study by Schnittger *et al*. (2013) has reported outcome data specifically in younger patients with *de novo* AML in a similar size cohort of 223 patients ≤ 60 years of age, although all their patients had intermediate-risk cytogenetics. They found that the mutations were a significant adverse factor for OS but not event-free survival in these patients. The reason(s) for the difference is (are) not clear, although it should be noted that the survival rates for their patients were much higher for all patient groups than those found in not only our study but also other studies.

In conclusion, the low incidence of *ASXL1* mutations in younger patients with primary disease and the lack of independent prognostic significance in multivariate analysis in our cohort suggests that there is a limited role for diagnostic screening of *ASXL1* for the purpose of prognostic stratification.

CHAPTER 6:**CONCLUSIONS AND FUTURE DIRECTIONS**

Acute myeloid leukaemia is a biologically heterogenous disease and this is reflected in the differing outcomes seen clinically. Currently, when it is diagnosed, patients are stratified into one of three prognostic groups depending on their cytogenetic status and mutation status in certain genes. Patients with a normal karyotype (40% of all patients with AML) fall into the intermediate prognostic group, of which only 35% will survive for greater than 10 years (Grimwade *et al*, 2010). Despite this biological heterogeneity, whole genome sequencing of samples from AML patients has identified the fewest number of mutations per patient of any malignancy to date, with a median of 13 mutations per sample compared to other malignancies, which can have hundreds (2013; Kandoth *et al*, 2013). Recently identified recurrent mutations have been in genes known to have epigenetic functions. The aim of the studies presented in this thesis was therefore to investigate whether genome-wide methylation patterns could provide additional information in predicting response to treatment, and in addition to ascertain whether knowledge of the mutation status of the epigenetic modifier gene *ASXL1* would add prognostic value to warrant testing in all patients at diagnosis.

The methylation profile of samples from 40 well-characterised patients in whom the outcome was known was first analysed using the Illumina Methylation 27K array (Chapter 3). All patients selected had an NK, and were WT for *NPM1* and *FLT3*; half the patients had chemosensitive disease and the others were chemoresistant. At the time the methylation profile was analysed, more was known about the molecular profile of the samples, in particular their *CEBPA*, *DNMT3A*, *WT1*, *IDH1*, *IDH2*, *GATA2* and *TET2* genotype. The results of the methylation array were validated using pyrosequencing assays of four of the genes that were found to be variably methylated. This demonstrated good correlation between the two methods of methylation analysis with r^2 ranging from 0.76 to 0.93 indicating the reliability of the Illumina Methylation array results.

Although the initial aim was to identify a methylation signature associated with prognosis, unsupervised clustering of the samples based on their methylation profiles did not correlate with overall outcome (Figure 3.9). However, it was

apparent from this analysis that the samples did group according to their *CEBPA* genotype and that *CEBPA*^{DM} samples had a distinct methylation profile compared to the other *CEBPA*^{WT} samples. Ten of the 40 samples in this cohort (25%) were *CEBPA*^{DM}, which is higher than the published overall incidence of 10%, but this was not surprising due to the criteria by which this cohort was selected, as this genotype is known to be more frequent in NK, *NPM1*^{WT} samples. This methylation profile was then explored further.

A methylation signature consisting of the top ranked 25 CpG sites that differentiated the eight classic *CEBPA*^{DM} samples with an N terminal out-of-frame mutation coupled with an in-frame C terminal mutation from the 30 *CEBPA*^{WT} samples was created. Two scores were derived for each sample dependent on how close they were to the classic *CEBPA*^{DM} and *CEBPA*^{WT} signatures, and the difference between these two scores was plotted against the distance from the classic *CEBPA*^{DM} score. The *CEBPA*^{DM} samples and *CEBPA*^{WT} samples formed two distinct groups on this plot. This was validated with a further 17 classic *CEBPA*^{DM} and 26 *CEBPA*^{WT} samples. One classic *CEBPA*^{DM} case did not group with the other cases. Although biallelic, it had low mutant levels of 24% and 31% compared to the mean mutant level of 44% and, as methylation levels are quantitative, it was hypothesised that this was the reason that the methylation signature was more similar to *CEBPA*^{WT} in this case. Using scores from all classic *CEBPA*^{DM} cases except this one, a classic *CEBPA*^{DM} quadrant for the plot was subsequently defined.

To assess whether the methylation profiles identified were specific to *CEBPA*^{DM} samples or whether they could be seen in other cohorts of samples that are associated with good prognosis or reduced C/EBPα, three CpG sites that formed part of the signature were analysed in 42 samples with *NPM1* mutations and 40 CBF leukaemia samples. The methylation levels at these sites were not the same as classic *CEBPA*^{DM} samples, indicating that the methylation profile was indeed specific to this mutation.

The methylation signature was then used to explore the significance of a variety of non-classic *CEBPA*^{DM} and *CEBPA*^{SM} samples (Chapter 4). Overall, 16 non-classic *CEBPA*^{DM} samples were assessed. Only one of the six samples that were predicted to produce just p30 protein was located in the classic *CEBPA*^{DM} quadrant. Two of the three samples that led to C loss of function only and two of

the seven *CEBPA*^{DM} samples that had a missense mutation were within the classic *CEBPA*^{DM} quadrant. This suggests that C loss of function is necessary, but not sufficient, to produce the *CEBPA*^{DM} methylation pattern seen in the classic samples. Only one of the 38 *CEBPA*^{SM} samples was located in the classic *CEBPA*^{DM} quadrant. This is further evidence that *CEBPA*^{SM} AML is biologically distinct from *CEBPA*^{DM}. The results from these studies also suggest that non-classic *CEBPA*^{DM} AML should not automatically be considered as part of the favourable prognostic group which is of clinical significance as stem cell transplants are not recommended in patients with favourable prognostic molecular aberrations.

Methylation of the *CEBPA* promoter was also investigated in the total cohort of 135 samples in order to ascertain whether this had an impact on the overall methylome. Firstly, CpGs that were interrogated as part of the Illumina Methylation arrays were analysed. One site 1919 bases upstream of the TSS for *CEBPA* showed differential methylation between *CEBPA*^{DM} and other samples, otherwise there was little variability in the methylation levels at the remaining 16 CpG sites related to *CEBPA*. An estimate of methylation levels in the distal and core promoter regions was determined using direct sequencing. No methylation was seen in the core promoter region in any of the samples. Only a few *CEBPA*^{DM} samples had evidence of methylation in the distal promoter region, and nearly all those that did had only low levels of methylation. Similarly, most *CEBPA*^{WT} and *CEBPA*^{SM} samples were also unmethylated in this region, however five *CEBPA*^{SM} and four *CEBPA*^{WT} were considered to be “hypermethylated”, defined as >25% and >50% methylation respectively. Only one of the hypermethylated *CEBPA*^{SM} samples was located in the classic *CEBPA*^{DM} quadrant, theoretically the wild-type allele in this case could have been silenced and this would be akin to C-LOF mutant only.

The aim of chapter 5 was to compare the frequencies of *ASXL1* mutations in younger and older patients with primary and secondary AML, and to ascertain whether knowledge of *ASXL1* genotype would significantly alter expected outcome. Denaturing HPLC was used to screen for mutations in exon 12 of the gene. Analysis was restricted to a single exon, as in those studies that analysed the whole gene, only 1% of mutations were located outside exon 12. Overall 32 out of 371 patients had *ASXL1* mutations (9%), which is comparable to

previously published data. The rate of mutations in younger patients with primary disease was 4% and secondary disease was 12%. In older patients, the rate of mutation was 11% in primary disease and 15% in secondary disease. Although mutations in *ASXL1* were associated with a worse overall survival, this was not significant in multivariate analysis due to the strong association with older age. Knowledge of *ASXL1* genotype at diagnosis thus did not provide further information regarding outcome of the patient to necessitate screening at diagnosis.

6.1 Future directions

The data presented in these studies suggest that different mutations identified in the *CEBPA* gene in patients with AML may affect the normal protein function in different ways. The methylation profile of the different types of *CEBPA* mutant samples is just one way of assessing the functional variability associated with the different types of mutations, however it does not explain why the methylation patterns are different. Attempting to correlate the methylation profile of the different types of *CEBPA* mutations with functional studies may provide a potential explanation for the variability. For example, the impact of the different *CEBPA* mutations on the function of the protein could be explored using *in vitro* transactivation assays whereby cells are transfected with *CEBPA* mutant constructs together with luciferase reporter constructs for known downstream targets of C/EBPα.

As further recurrent mutations are identified, it would be interesting to assess the status of the samples used in this study to ascertain whether any further mutations have an impact on the methylation profile seen, for example activating *NOTCH* mutations which have been seen in a minority of AML samples and have been shown to lead to gene expression profiles that are similar to *CEBPA*^{DM} samples (Wouters *et al*, 2007).

From a clinical perspective, any potential prognostic role of using DNA methylation analysis to assess the significance of a non-classic *CEBPA*^{DM} would be at diagnosis rather than detecting minimal residual disease, as there would be too much variability in methylation level in normal bone marrow to pick up small clones of disease. However, as previously discussed, although the methylation profile of different *CEBPA*^{DM} samples indicates that there are

differences between some of the non-classic *CEBPA*^{DM} cases compared to the classic *CEBPA*^{DM} cases, this does not necessarily mean that the outcome will be different in these two groups, and due to the small numbers of patients it would be difficult to confirm or refute this.

6.2 Conclusions

In conclusion, the results presented in this thesis show that the aberrant methylome in AML can be linked with the underlying molecular and cytogenetic status of the samples. In particular, in the cohort of AML samples selected in this study that were cytogenetically normal with no *NPM1* mutation or *FLT3*/ITD, samples with classic *CEBPA*^{DM} mutations had a characteristic methylome distinct from all *CEBPA*^{WT} and the majority of *CEBPA*^{SM} samples. A classic *CEBPA*^{DM} signature could be derived that was used to investigate the impact of non-classic *CEBPA*^{DM} mutations on the methylation profile of samples. Not all samples with non-classic *CEBPA*^{DM} mutations had the classic signature, particularly those lacking a C-LOF mutation. As molecular genotyping is increasingly being used as the basis for risk-adapted therapy, it is important to understand whether all mutations within a particular gene are associated with the same prognosis. Moreover, it is important to assess whether mutations detected are associated with other confounding factors that confer a poor prognosis. In these studies, the impact of *ASXL1* mutations in AML was analysed. Although mutations were associated with a worse outcome, they were not significant in multivariate analysis due to the low incidence and the confounding role of age. Thus, as technological advances continue and allow whole genome and exome sequencing to become more frequent in AML, the interpretation of the results becomes increasingly complex, particularly as clinical trials will not be able to power increasingly small sub-group analysis to prove which combination of mutations are good or poor risk. It is in this context, that the methylation profile may play a role in demonstrating the global effects of multiple genetic abnormalities.

REFERENCES

(2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**(7422)**:** 56-65

(2013) Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine* **368**(22)**:** 2059-74

Abbas S, Lugthart S, Kavelaars FG, Schelen A, Koenders JE, Zeilemaker A, van Putten WJ, Rijneveld AW, Lowenberg B, Valk PJ (2010) Acquired mutations in the genes encoding IDH1 and IDH2 both are recurrent aberrations in acute myeloid leukemia: prevalence and prognostic value. *Blood* **116**(12)**:** 2122-6

Abdel-Wahab O, Adli M, Lafave LM, Gao J, Hricik T, Shih AH, Pandey S, Patel JP, Chung YR, Koche R, Perna F, Zhao X, Taylor JE, Park CY, Carroll M, Melnick A, Nimer SD, Jaffe JD, Aifantis I, Bernstein BE, Levine RL (2012) ASXL1 Mutations Promote Myeloid Transformation through Loss of PRC2-Mediated Gene Repression. *Cancer Cell* **22**(2)**:** 180-93

Abdel-Wahab O, Gao J, Adli M, Dey A, Trimarchi T, Chung YR, Kuscu C, Hricik T, Ndiaye-Lobry D, LaFave LM, Koche R, Shih AH, Guryanova OA, Kim E, Li S, Pandey S, Shin JY, Telis L, Liu J, Bhatt PK, Monette S, Zhao X, Mason CE, Park CY, Bernstein BE, Aifantis I, Levine RL (2013) Deletion of Asxl1 results in myelodysplasia and severe developmental defects in vivo. *J Exp Med* **210**(12)**:** 2641-2659

Abdel-Wahab O, Kilpivaara O, Patel J, Busque L, Levine RL (2010a) The most commonly reported variant in ASXL1 (c.1934dupG;p.Gly646TrpfsX12) is not a somatic alteration. *Leukemia* **24**(9)**:** 1656-7

Abdel-Wahab O, Manshouri T, Patel J, Harris K, Yao J, Hedvat C, Heguy A, Bueso-Ramos C, Kantarjian H, Levine RL, Verstovsek S (2010b) Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias. *Cancer Res* **70**(2)**:** 447-52

Abdel-Wahab O, Pardanani A, Patel J, Wadleigh M, Lasho T, Heguy A, Beran M, Gilliland DG, Levine RL, Tefferi A (2011) Concomitant analysis of EZH2 and ASXL1 mutations in myelofibrosis, chronic myelomonocytic leukemia and blast-phase myeloproliferative neoplasms. *Leukemia* **25**(7)**:** 1200-2

Agrawal S, Hofmann WK, Tidow N, Ehrich M, van den Boom D, Koschmieder S, Berdel WE, Serve H, Muller-Tidow C (2007) The C/EBPdelta tumor suppressor is silenced by hypermethylation in acute myeloid leukemia. *Blood* **109**(9)**:** 3895-905

Allen CG, Hills RK, Evans CM, Lamb K, Sellar R, Moorman AV, Sale M, Liu-Yin JA, Burnett AK, Linch DC, Gale RE (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**(21)**:** 419-

Arber DA, Stein AS, Carter NH, Ikle D, Forman SJ, Slovak ML (2003) Prognostic impact of acute myeloid leukemia classification. Importance of detection of recurring cytogenetic abnormalities and multilineage dysplasia on survival. *Am J Clin Pathol* **119**(5)**:** 672-80

Aryee MJ, Wu Z, Ladd-Acosta C, Herb B, Feinberg AP, Yegnasubramanian S, Irizarry RA (2011) Accurate genome-scale percentage DNA methylation estimates from microarray data. *Biostatistics* **12**(2)**:** 197-210

Bacher U, Haferlach C, Schnittger S, Kohlmann A, Kern W, Haferlach T (2010) Mutations of the TET2 and CBL genes: novel molecular markers in myeloid malignancies. *Ann Hematol* **89**(7)**:** 643-52

Bacher U, Haferlach T, Schoch C, Kern W, Schnittger S (2006) Implications of NRAS mutations in AML: a study of 2502 patients. *Blood* **107**(10)**:** 3847-53

Barjesteh van Waalwijk van Doorn-Khosrovani S, Erpelinck C, Meijer J, van Oosterhoud S, van Putten WL, Valk PJ, Berna Beverloo H, Tenen DG, 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**(1)**:** 31-40

Becker H, Yoshida K, Blagitko-Dorfs N, Claus R, Pantic M, Abdelkarim M, Niemöller C, Greil C, Hackanson B, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Döhner K, Schnittger S, Henneke P, Niemeyer CM, Flotho C, Pfeifer D, Ogawa S, Lübbert M (2014) Tracing the development of acute myeloid leukemia in CBL syndrome. *Blood* **123**(12)**:** 1883-1886

Behdad A, Weigelin HC, Elenitoba-Johnson KS, Betz BL (2015) A clinical grade sequencing-based assay for CEBPA mutation testing: report of a large series of myeloid neoplasms. *J Mol Diagn* **17**(1)**:** 76-84

Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G, Kantarjian H, Raza A, Levine RL, Neuberg D, Ebert BL (2011) Clinical effect of point mutations in myelodysplastic syndromes. *The New England journal of medicine* **364**(26)**:** 2496-506

Bennett KL, Hackanson B, Smith LT, Morrison CD, Lang JC, Schuller DE, Weber F, Eng C, Plass C (2007) Tumor suppressor activity of CCAAT/enhancer

binding protein alpha is epigenetically down-regulated in head and neck squamous cell carcinoma. *Cancer Res* **67**(10)**:** 4657-64

Benthaus T, Schneider F, Mellert G, Zellmeier E, Schneider S, Kakadia PM, Hiddemann W, Bohlander SK, Feuring-Buske M, Braess J, Spiekermann K, Dufour A (2008) Rapid and sensitive screening for CEBPA mutations in acute myeloid leukaemia. *British journal of haematology* **143**(2)**:** 230-9

Bereshchenko O, Mancini E, Moore S, Bilbao D, Mansson R, Luc S, Grover A, Jacobsen SE, 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**(5)**:** 390-400

Bestor TH (1988) Cloning of a mammalian DNA methyltransferase. *Gene* **74**(1)**:** 9-12

Bestor TH, Ingram VM (1983) Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proceedings of the National Academy of Sciences of the United States of America* **80**(18)**:** 5559-63

Bhattacharyya S, Yu Y, Suzuki M, Campbell N, Mazdo J, Vasanthakumar A, Bhagat TD, Nischal S, Christopeit M, Parekh S, Steidl U, Godley L, Maitra A, Greally JM, Verma A (2013) Genome-wide hydroxymethylation tested using the HELP-GT assay shows redistribution in cancer. *Nucleic Acids Res* **41**(16)**:** e157

Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan JB, Shen R (2011) High density DNA methylation array with single CpG site resolution. *Genomics* **98**(4)**:** 288-95

Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL (2009) Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics* **1**(1)**:** 177-200

Bienz M, Ludwig M, Leibundgut EO, Mueller BU, Ratschiller D, Solenthaler M, Fey MF, Pabst T (2005) Risk assessment in patients with acute myeloid leukemia and a normal karyotype. *Clin Cancer Res* **11**(4)**:** 1416-24

Bird A (2007) Perceptions of epigenetics. *Nature* **447**(7143)**:** 396-398

Bird AP (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* **8**(7)**:** 1499-1504

Boissel N, Leroy H, Brethon B, Philippe N, de Botton 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**(6)**:** 965-70

Boissel N, Nibourel O, Renneville A, Gardin C, Reman O, Contentin N, Bordessoule D, Pautas C, de Revel 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. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**(23)**:** 3717-23

Borssen M, Palmqvist L, Karrman K, Abrahamsson J, Behrendtz M, Heldrup J, Forestier E, Roos G, Degerman S (2013) Promoter DNA methylation pattern identifies prognostic subgroups in childhood T-cell acute lymphoblastic leukemia. *PLoS One* **8**(6)**:** e65373

Boultwood J, Perry J, Pellagatti A, Fernandez-Mercado M, Fernandez-Santamaria C, Calasanz MJ, Larrayoz MJ, Garcia-Delgado M, Giagounidis A, Malcovati L, Della Porta MG, Jadersten M, Killick S, Hellstrom-Lindberg E, Cazzola M, Wainscoat JS (2010a) Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. *Leukemia* **24**(5)**:** 1062-5

Boultwood J, Perry J, Zaman R, Fernandez-Santamaria C, Littlewood T, Kusec R, Pellagatti A, Wang L, Clark RE, Wainscoat JS (2010b) High-density single nucleotide polymorphism array analysis and ASXL1 gene mutation screening in chronic myeloid leukemia during disease progression. *Leukemia* **24**(6)**:** 1139-45

Bowen D, Groves MJ, Burnett AK, Patel Y, Allen C, Green C, Gale RE, Hills R, Linch DC (2009) TP53 gene mutation is frequent in patients with acute myeloid leukemia and complex karyotype, and is associated with very poor prognosis. *Leukemia* **23**(1)**:** 203-6

Bowen DT, Frew ME, Hills R, Gale RE, Wheatley K, Groves MJ, Langabeer SE, Kottaridis PD, Moorman AV, Burnett AK, Linch DC (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**(6)**:** 2113-9

Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, Bell GW, Otte AP, Vidal M, Gifford DK, Young RA, Jaenisch R (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**(7091)**:** 349-353

Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Nemes A, Temper V, Razin A, Cedar H (1994) Sp1 elements protect a CpG island from de novo methylation. *Nature* **371**(6496)**:** 435-8

Brecqueville M, Rey J, Bertucci F, Coppin E, Finetti P, Carbuccia N, Cervera N, Gelsi-Boyer V, Arnoulet C, Gisserot O, Verrot D, Slama B, Vey N, Mozziconacci MJ, Birnbaum D, Murati A (2012) Mutation analysis of ASXL1, CBL, DNMT3A, IDH1, IDH2, JAK2, MPL, NF1, SF3B1, SUZ12, and TET2 in myeloproliferative neoplasms. *Genes, chromosomes & cancer* **51**(8)**:** 743-55

Breen TR, Duncan IM (1986) Maternal expression of genes that regulate the bithorax complex of Drosophila melanogaster. *Developmental biology* **118**(2)**:** 442-56

Bullinger L, Ehrich M, Dohner K, Schlenk RF, Dohner H, Nelson MR, van den Boom D (2010) Quantitative DNA methylation predicts survival in adult acute myeloid leukemia. *Blood* **115**(3)**:** 636-42

Burnett A, Wetzler M, Löwenberg B (2011) Therapeutic Advances in Acute Myeloid Leukemia. *Journal of Clinical Oncology* **29**(5)**:** 487-494

Burnett AK (2013) Treatment of acute myeloid leukaemia. *Clin Med (Lond)* **13 Suppl 6:** s58-61

Burnett AK, Hills RK (2011) Who should be transplanted in first remission of acute myeloid leukaemia? *Curr Treat Options Oncol* **12**(4)**:** 329-40

Burnett AK, Hills RK, Milligan DW, Goldstone AH, Prentice AG, McMullin MF, Duncombe A, Gibson B, Wheatley K (2010) Attempts to optimize induction and consolidation treatment in acute myeloid leukemia: results of the MRC AML12 trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**(4)**:** 586-95

Cairoli R, Beghini A, Grillo G, Nadali G, Elice F, Ripamonti CB, 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**(9)**:** 3463-8

Calkhoven CF, Müller C, Leutz A (2000) Translational control of C/EBPα and C/EBPβ isoform expression. *Genes & Development* **14**(15)**:** 1920-1932

Calvanese V, Fernandez AF, Urdinguio RG, Suarez-Alvarez B, Mangas C, Perez-Garcia V, Bueno C, Montes R, Ramos-Mejia V, Martinez-Camblor P, Ferrero C, Assenov Y, Bock C, Menendez P, Carrera AC, Lopez-Larrea C,

Fraga MF (2012) A promoter DNA demethylation landscape of human hematopoietic differentiation. *Nucleic Acids Res* **40**(1)**:** 116-31

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y (2002) Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* **298**(5595)**:** 1039-1043

Carbuccia N, Murati A, Trouplin V, Brecqueville M, Adelaide J, Rey J, Vainchenker W, Bernard OA, Chaffanet M, Vey N, Birnbaum D, Mozziconacci MJ (2009) Mutations of ASXL1 gene in myeloproliferative neoplasms. *Leukemia* **23**(11)**:** 2183-6

Carbuccia N, Trouplin V, Gelsi-Boyer V, Murati A, Rocquain J, Adelaide J, Olschwang S, Xerri L, Vey N, Chaffanet M, Birnbaum D, Mozziconacci MJ (2010) Mutual exclusion of ASXL1 and NPM1 mutations in a series of acute myeloid leukemias. *Leukemia* **24**(2)**:** 469-73

Care RS, Valk PJ, Goodeve AC, Abu-Duhier FM, Geertsma-Kleinekoort WM, Wilson GA, Gari MA, Peake IR, Lowenberg B, Reilly JT (2003) Incidence and prognosis of c-KIT and FLT3 mutations in core binding factor (CBF) acute myeloid leukaemias. *British journal of haematology* **121**(5)**:** 775-7

Chen T, Ueda Y, Dodge JE, Wang Z, Li E (2003) Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol* **23**(16)**:** 5594-605

Chen W, Xie H, Wang H, Chen L, Sun Y, Chen Z, Li Q (2016) Prognostic Significance of KIT Mutations in Core-Binding Factor Acute Myeloid Leukemia: A Systematic Review and Meta-Analysis. *PLoS One* **11**(1)**:** e0146614

Chim CS, Wong AS, Kwong YL (2002) Infrequent hypermethylation of CEBPA promotor in acute myeloid leukaemia. *British journal of haematology* **119**(4)**:** 988-90

Cho YS, Kim EJ, Park UH, Sin HS, Um SJ (2006) Additional sex comb-like 1 (ASXL1), in cooperation with SRC-1, acts as a ligand-dependent coactivator for retinoic acid receptor. *J Biol Chem* **281**(26)**:** 17588-98

Chou WC, Chou SC, Liu CY, Chen CY, Hou HA, Kuo YY, Lee MC, Ko BS, Tang JL, Yao M, Tsay W, Wu SJ, Huang SY, Hsu SC, Chen YC, Chang YC, Kuo KT, Lee FY, Liu MC, Liu CW, Tseng MH, Huang CF, Tien HF (2011a) TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. *Blood* **118**(14)**:** 3803-10

Chou WC, Hou HA, Chen CY, Tang JL, Yao M, Tsay W, Ko BS, Wu SJ, Huang SY, Hsu SC, Chen YC, Huang YN, Chang YC, Lee FY, Liu MC, Liu CW, Tseng

MH, Huang CF, Tien HF (2010a) Distinct clinical and biologic characteristics in adult acute myeloid leukemia bearing the isocitrate dehydrogenase 1 mutation. *Blood* **115**(14)**:** 2749-54

Chou WC, Huang HH, Hou HA, Chen CY, Tang JL, Yao M, Tsay W, Ko BS, Wu SJ, Huang SY, Hsu SC, Chen YC, Huang YN, Chang YC, Lee FY, Liu MC, Liu CW, Tseng MH, Huang CF, Tien HF (2010b) Distinct clinical and biological features of de novo acute myeloid leukemia with additional sex comb-like 1 (ASXL1) mutations. *Blood* **116**(20)**:** 4086-94

Chou WC, Lei WC, Ko BS, Hou HA, Chen CY, Tang JL, Yao M, Tsay W, Wu SJ, Huang SY, Hsu SC, Chen YC, Chang YC, Kuo KT, Lee FY, Liu MC, Liu CW, Tseng MH, Huang CF, Tien HF (2011b) The prognostic impact and stability of Isocitrate dehydrogenase 2 mutation in adult patients with acute myeloid leukemia. *Leukemia* **25**(2)**:** 246-53

Cornelissen JJ, Gratwohl A, Schlenk RF, Sierra J, Bornhauser M, Juliusson G, Racil Z, Rowe JM, Russell N, Mohty M, Lowenberg B, Socie G, Niederwieser D, Ossenkoppele GJ (2012) The European LeukemiaNet AML Working Party consensus statement on allogeneic HSCT for patients with AML in remission: an integrated-risk adapted approach. *Nat Rev Clin Oncol* **9**(10)**:** 579-90

Cornelissen JJ, van Putten WL, Verdonck LF, Theobald M, Jacky E, Daenen SM, van Marwijk Kooy M, Wijermans P, Schouten H, Huijgens PC, 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**(9)**:** 3658-66

Costello JF, Fruhwald MC, Smiraglia DJ, Rush LJ, Robertson GP, Gao X, Wright FA, Feramisco JD, Peltomaki P, Lang JC, Schuller DE, Yu L, Bloomfield CD, Caligiuri MA, Yates A, Nishikawa R, Su Huang H, Petrelli NJ, Zhang X, O'Dorisio MS, Held WA, Cavenee WK, Plass C (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nature genetics* **24**(2)**:** 132-8

Cross SH, Charlton JA, Nan X, Bird AP (1994) Purification of CpG islands using a methylated DNA binding column. *Nature genetics* **6**(3)**:** 236-44

Damaj G, Joris M, Chandesris O, Hanssens K, Soucie E, Canioni D, Kolb B, Durieu I, Gyan E, Livideanu C, Cheze S, Diouf M, Garidi R, Georgin-Lavialle S, Asnafi V, Lhermitte L, Lavigne C, Launay D, Arock M, Lortholary O, Dubreuil P, Hermine O (2014) ASXL1 but Not TET2 Mutations Adversely Impact Overall Survival of Patients Suffering Systemic Mastocytosis with Associated Clonal Hematologic Non-Mast-Cell Diseases. *PLoS One* **9**(1)**:** e85362

Davies C, Yip BH, Fernandez-Mercado M, Woll PS, Agirre X, Prosper F, Jacobsen SE, Wainscoat JS, Pellagatti A, Boultwood J (2013) Silencing of ASXL1 impairs the granulomonocytic lineage potential of human CD34+ progenitor cells. *British Journal of Haematology* **160**(6)**:** 842-850

De Smet C, Lurquin C, Lethe B, Martelange V, Boon T (1999) DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol Cell Biol* **19**(11)**:** 7327-35

Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F (2011) Evaluation of the Infinium Methylation 450K technology. *Epigenomics* **3**(6)**:** 771-84

Deneberg S, Grovdal M, Karimi M, Jansson M, Nahi H, Corbacioglu A, Gaidzik V, Dohner K, Paul C, Ekstrom TJ, Hellstrom-Lindberg E, Lehmann S (2010) Gene-specific and global methylation patterns predict outcome in patients with acute myeloid leukemia. *Leukemia* **24**(5)**:** 932-941

Deneberg S, Guardiola P, Lennartsson A, Qu Y, Gaidzik V, Blanchet O, Karimi M, Bengtzen S, Nahi H, Uggla B, Tidefelt U, Hoglund M, Paul C, Ekwall K, Dohner K, Lehmann S (2011) Prognostic DNA methylation patterns in cytogenetically normal acute myeloid leukemia are predefined by stem cell chromatin marks. *Blood* **118**(20)**:** 5573-82

Devillier R, Gelsi-Boyer V, Brecqueville M, Carbuccia N, Murati A, Vey N, Birnbaum D, Mozziconacci MJ (2012) Acute myeloid leukemia with myelodysplasia-related changes are characterized by a specific molecular pattern with high frequency of ASXL1 mutations. *American journal of hematology* **87**(7)**:** 659-62

Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, Terragni J, Figueroa ME, De Figueiredo Pontes LL, Alberich-Jorda M, Zhang P, Wu M, D'Alo F, Melnick A, Leone G, Ebralidze KK, Pradhan S, Rinn JL, Tenen DG (2013) DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**(7476)**:** 371-6

Döhner H, Estey EH, Amadori S, Appelbaum FR, Büchner T, Burnett AK, Dombret H, Fenaux P, Grimwade D, Larson RA, Lo-Coco F, Naoe T, Niederwieser D, Ossenkoppele GJ, Sanz MA, Sierra J, Tallman MS, Löwenberg B, Bloomfield CD (2010) Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* **115**(3)**:** 453-474

Dohner K, Schlenk RF, Habdank M, Scholl C, Rucker FG, 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**(12)**:** 3740-6

Dohner K, Tobis K, Ulrich R, Frohling S, Benner A, Schlenk RF, Dohner H (2002) Prognostic significance of partial tandem duplications of the MLL gene in adult patients 16 to 60 years old with acute myeloid leukemia and normal cytogenetics: a study of the Acute Myeloid Leukemia Study Group Ulm. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **20**(15)**:** 3254-61

Dombret H, Seymour JF, Butrym A, Wierzbowska A, Selleslag D, Jang JH, Kumar R, Cavenagh J, Schuh AC, Candoni A, Recher C, Sandhu I, Bernal del Castillo T, Al-Ali HK, Martinelli G, Falantes J, Noppeney R, Stone RM, Minden MD, McIntyre H, Songer S, Lucy LM, Beach CL, Dohner H (2015) International phase 3 study of azacitidine vs conventional care regimens in older patients with newly diagnosed AML with >30% blasts. *Blood* **126**(3)**:** 291-9

Dong KB, Maksakova IA, Mohn F, Leung D, Appanah R, Lee S, Yang HW, Lam LL, Mager DL, Schubeler D, Tachibana M, Shinkai Y, Lorincz MC (2008) DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *Embo J* **27**(20)**:** 2691-701

Dufour A, Schneider F, Hoster E, Benthaus T, Ksienzyk B, Schneider S, Kakadia PM, Sauerland MC, Berdel WE, Buchner T, Wormann B, Braess J, Subklewe M, Hiddemann W, Bohlander SK, Spiekermann K, group ACs (2012) Monoallelic CEBPA mutations in normal karyotype acute myeloid leukemia: independent favorable prognostic factor within NPM1 mutated patients. *Ann Hematol* **91**(7)**:** 1051-63

Dufour A, Schneider F, Metzeler KH, Hoster E, Schneider S, Zellmeier E, Benthaus T, Sauerland MC, Berdel WE, Buchner T, Wormann B, Braess J, Hiddemann W, Bohlander SK, 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**(4)**:** 570-7

Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Liddle J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Berlin K, Beck S (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nature genetics* **38**(12)**:** 1378-85

Eden A, Gaudet F, Waghmare A, Jaenisch R (2003) Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **300**(5618)**:** 455

Eden S, Hashimshony T, Keshet I, Cedar H, Thorne AW (1998) DNA methylation models histone acetylation. *Nature* **394**(6696)**:** 842

Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, Gehrke C (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res* **10**(8)**:** 2709-21

El-Sharkawi D, Ali A, Evans CM, Hills RK, Burnett AK, Linch DC, Gale RE (2013) ASXL1 mutations are infrequent in young patients with primary acute myeloid leukemia and their detection has a limited role in therapeutic risk stratification. *Leukemia & Lymphoma*

Esteller M (2002) CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* **21**(35)**:** 5427-40

Esteller M, Corn PG, Baylin SB, Herman JG (2001) A gene hypermethylation profile of human cancer. *Cancer research* **61**(8)**:** 3225-9

Esteve PO, Chin HG, Smallwood A, Feehery GR, Gangisetty O, Karpf AR, Carey MF, Pradhan S (2006) Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes & Development* **20**(22)**:** 3089-103

Fasan A, Alpermann T, Haferlach C, Grossmann V, Roller A, Kohlmann A, Eder C, Kern W, Haferlach T, Schnittger S (2013a) Frequency and Prognostic Impact of CEBPA Proximal, Distal and Core Promoter Methylation in Normal Karyotype AML: A Study on 623 Cases. *PLoS One* **8**(2)**:** e54365

Fasan A, Eder C, Haferlach C, Grossmann V, Kohlmann A, Dicker F, Kern W, Haferlach T, Schnittger S (2013b) GATA2 mutations are frequent in intermediate-risk karyotype AML with biallelic CEBPA mutations and are associated with favorable prognosis. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* **27**(2)**:** 482-5

Fasan A, Haferlach C, Alpermann T, Jeromin S, Grossmann V, Eder C, Weissmann S, Dicker F, Kohlmann A, Schindela S, Kern W, Haferlach T, Schnittger S (2014) The role of different genetic subtypes of CEBPA mutated AML. *Leukemia* **28**(4)**:** 794-803

Feinberg AP, Vogelstein B (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**(5895)**:** 89-92

Feldman N, Gerson A, Fang J, Li E, Zhang Y, Shinkai Y, Cedar H, Bergman Y (2006) G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* **8**(2)**:** 188-94

Fernandez-Mercado M, Yip BH, Pellagatti A, Davies C, Larrayoz MJ, Kondo T, Perez C, Killick S, McDonald EJ, Odero MD, Agirre X, Prosper F, Calasanz MJ, Wainscoat JS, Boultwood J (2012) Mutation Patterns of 16 Genes in Primary and Secondary Acute Myeloid Leukemia (AML) with Normal Cytogenetics. *PLoS One* **7**(8)**:** e42334

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**(7347)**:** 398-402

Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF, Tallman MS, Sun Z, Wolniak K, Peeters JK, Liu W, Choe SE, Fantin VR, Paietta E, Lowenberg B, Licht JD, Godley LA, Delwel R, Valk PJ, Thompson CB, Levine RL, Melnick A (2010a) Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**(6)**:** 553-67

Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, Schifano E, Booth J, van Putten W, Skrabanek L, Campagne F, Mazumdar M, Greally JM, Valk PJ, Lowenberg B, Delwel R, Melnick A (2010b) DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17**(1)**:** 13-27

Figueroa ME, Skrabanek L, Li Y, Jiemjit A, Fandy TE, Paietta E, Fernandez H, Tallman MS, Greally JM, Carraway H, Licht JD, Gore SD, Melnick A (2009a) MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation. *Blood* **114**(16)**:** 3448-58

Figueroa ME, Wouters BJ, Skrabanek L, Glass J, Li Y, Erpelinck-Verschueren CA, Langerak AW, Lowenberg B, Fazzari M, Greally JM, Valk PJ, Melnick A, Delwel R (2009b) Genome-wide epigenetic analysis delineates a biologically distinct immature acute leukemia with myeloid/T-lymphoid features. *Blood* **113**(12)**:** 2795-804

Fisher CL, Berger J, Randazzo F, Brock HW (2003) A human homolog of Additional sex combs, ADDITIONAL SEX COMBS-LIKE 1, maps to chromosome 20q11. *Gene* **306:** 115-26

Fisher CL, Pineault N, Brookes C, Helgason CD, Ohta H, Bodner C, Hess JL, Humphries RK, Brock HW (2010) Loss-of-function Additional sex combs like 1 mutations disrupt hematopoiesis but do not cause severe myelodysplasia or leukemia. *Blood* **115**(1)**:** 38-46

Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, De T, Teague JW, Stratton MR,

McDermott U, Campbell PJ (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* **43**(Database issue)**:** D805-11

Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, Iyer NG, Perez-Rosado A, Calvo E, Lopez JA, Cano A, Calasanz MJ, Colomer D, Piris MA, Ahn N, Imhof A, Caldas C, Jenuwein T, Esteller M (2005) Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nature genetics* **37**(4)**:** 391-400

Fraga MF, Herranz M, Espada J, Ballestar E, Paz MF, Ropero S, Erkek E, Bozdogan O, Peinado H, Niveleau A, Mao JH, Balmain A, Cano A, Esteller M (2004) A mouse skin multistage carcinogenesis model reflects the aberrant DNA methylation patterns of human tumors. *Cancer Research* **64**(16)**:** 5527-34

Frank D, Keshet I, Shani M, Levine A, Razin A, Cedar H (1991) Demethylation of CpG islands in embryonic cells. *Nature* **351**(6323)**:** 239-41

Friedman AD (2015) C/EBPalpha in normal and malignant myelopoiesis. *Int J Hematol* **101**(4)**:** 330-41

Frischmeyer PA, Dietz HC (1999) Nonsense-Mediated mRNA Decay in Health and Disease. *Hum Mol Genet* **8**(10)**:** 1893-1900

Fröhling S, Lipka DB, Kayser S, Scholl C, Schlenk RF, Döhner H, Gilliland DG, Levine RL, Döhner K (2006) Rare occurrence of the JAK2 V617F mutation in AML subtypes M5, M6, and M7. *Blood* **107**(3)**:** 1242-1243

Frohling S, Schlenk RF, Krauter J, Thiede C, Ehninger G, Haase D, Harder L, Kreitmeier S, Scholl C, Caligiuri MA, Bloomfield CD, 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**(4)**:** 427-32

Frohling S, Schlenk RF, 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**(4)**:** 624-33

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America* **89**(5)**:** 1827-31

Fuchs O, Provaznikova D, Kocova M, Kostecka A, Cvekova P, Neuwirtova R, Kobylka P, Cermak J, Brezinova J, Schwarz J, Markova J, Salaj P, Klamova H, Maaloufova J, Lemez P, Novakova L, Benesova K (2008) CEBPA polymorphisms and mutations in patients with acute myeloid leukemia, myelodysplastic syndrome, multiple myeloma and non-Hodgkin's lymphoma. *Blood cells, molecules & diseases* **40**(3)**:** 401-5

Gaidzik VI, Bullinger L, Schlenk RF, Zimmermann AS, Rock J, Paschka P, Corbacioglu A, Krauter J, Schlegelberger B, Ganser A, Spath D, Kundgen A, Schmidt-Wolf IG, Gotze K, Nachbaur D, Pfreundschuh M, Horst HA, 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**(10)**:** 1364-72

Gaidzik VI, Paschka P, Spath D, Habdank M, Kohne CH, Germing U, von Lilienfeld-Toal M, Held G, Horst HA, Haase D, Bentz M, Gotze K, Dohner H, Schlenk RF, 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. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **30**(12)**:** 1350-7

Gaidzik VI, Schlenk RF, 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**(19)**:** 4505-11

Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, Lin JC, Liang G, Jones PA, Tanay A (2008) Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proceedings of the National Academy of Sciences of the United States of America* **105**(35)**:** 12979-84

Gale RE, Green C, Allen C, Mead AJ, Burnett AK, Hills RK, Linch DC (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**(5)**:** 2776-84

Gale RE, Lamb K, Allen C, El-Sharkawi D, Stowe C, Jenkinson S, Tinsley S, Dickson G, Burnett AK, Hills RK, Linch DC (2015) Simpson's Paradox and the Impact of Different DNMT3A Mutations on Outcome in Younger Adults With Acute Myeloid Leukemia. *J Clin Oncol* **33**(18)**:** 2072-83

Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. *J Mol Biol* **196**(2)**:** 261-82

Garzon R, Volinia S, Papaioannou D, Nicolet D, Kohlschmidt J, Yan PS, Mrozek K, Bucci D, Carroll AJ, Baer MR, Wetzler M, Carter TH, Powell BL, Kolitz JE, Moore JO, Eisfeld AK, Blachly JS, Blum W, Caligiuri MA, Stone RM, Marcucci G, Croce CM, Byrd JC, Bloomfield CD (2014) Expression and prognostic impact of lncRNAs in acute myeloid leukemia. *Proc Natl Acad Sci U S A* **111**(52)**:** 18679-84

Gebhard C, Benner C, Ehrich M, Schwarzfischer L, Schilling E, Klug M, Dietmaier W, Thiede C, Holler E, Andreesen R, Rehli M (2010) General Transcription Factor Binding at CpG Islands in Normal Cells Correlates with Resistance to De novo DNA Methylation in Cancer Cells. *Cancer Research* **70**(4)**:** 1398-1407

Gelsi-Boyer V, Brecqueville M, Devillier R, Murati A, Mozziconacci MJ, Birnbaum D (2012) Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases. *J Hematol Oncol* **5:** 12

Gelsi-Boyer V, Trouplin V, Adelaide J, Bonansea J, Cervera N, Carbuccia N, Lagarde A, Prebet T, Nezri M, Sainty D, Olschwang S, Xerri L, Chaffanet M, Mozziconacci MJ, Vey N, Birnbaum D (2009) Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br J Haematol* **145**(6)**:** 788-800

Gelsi-Boyer V, Trouplin V, Roquain J, Adelaide J, Carbuccia N, Esterni B, Finetti P, Murati A, Arnoulet C, Zerazhi H, Fezoui H, Tadrist Z, Nezri M, Chaffanet M, Mozziconacci MJ, Vey N, Birnbaum D (2010) ASXL1 mutation is associated with poor prognosis and acute transformation in chronic myelomonocytic leukaemia. *Br J Haematol* **151**(4)**:** 365-75

Gildea JJ, Lopez R, Shearn A (2000) A Screen for New Trithorax Group Genes Identified little imaginal discs, the Drosophila melanogaster Homologue of Human Retinoblastoma Binding Protein 2. *Genetics* **156**(2)**:** 645-663

Giles FJ, Borthakur G, Ravandi F, Faderl S, Verstovsek S, Thomas D, Wierda W, Ferrajoli A, Kornblau S, Pierce S, Albitar M, Cortes J, Kantarjian H (2007) The haematopoietic cell transplantation comorbidity index score is predictive of early death and survival in patients over 60 years of age receiving induction therapy for acute myeloid leukaemia. *British Journal of Haematology* **136**(4)**:** 624-627

Goldstone AH, Burnett AK, Wheatley K, Smith AG, Hutchinson RM, Clark RE (2001) Attempts to improve treatment outcomes in acute myeloid leukemia (AML) in older patients: the results of the United Kingdom Medical Research Council AML11 trial. *Blood* **98**(5)**:** 1302-11

Gombart AF, Hofmann WK, Kawano S, Takeuchi S, Krug U, Kwok SH, Larsen RJ, Asou H, Miller CW, Hoelzer D, Koeffler HP (2002) Mutations in the gene encoding the transcription factor CCAAT/enhancer binding protein alpha in myelodysplastic syndromes and acute myeloid leukemias. *Blood* **99**(4)**:** 1332-40

Gonzalez-Zulueta M, Bender CM, Yang AS, Nguyen T, Beart RW, Van Tornout JM, Jones PA (1995) Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. *Cancer research* **55**(20)**:** 4531-5

Gonzalo S, Jaco I, Fraga MF, Chen T, Li E, Esteller M, Blasco MA (2006) DNA methyltransferases control telomere length and telomere recombination in mammalian cells. *Nat Cell Biol* **8**(4)**:** 416-424

Green CL, Evans CM, Hills RK, Burnett AK, Linch DC, Gale RE (2010a) The prognostic significance of IDH1 mutations in younger adult patients with acute myeloid leukemia is dependent on FLT3/ITD status. *Blood* **116**(15)**:** 2779-82

Green CL, Evans CM, Zhao L, Hills RK, Burnett AK, Linch DC, Gale RE (2011) The prognostic significance of IDH2 mutations in AML depends on the location of the mutation. *Blood* **118**(2)**:** 409-12

Green CL, Koo KK, Hills RK, Burnett AK, Linch DC, Gale RE (2010b) 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. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**(16)**:** 2739-47

Green CL, Tawana K, Hills RK, Bodor C, Fitzgibbon J, Inglott S, Ancliff P, Burnett AK, Linch DC, Gale RE (2013) GATA2 mutations in sporadic and familial acute myeloid leukaemia patients with CEBPA mutations. *British Journal of Haematology* **161**(5)**:** 701-5

Greger V, Passarge E, Hopping W, Messmer E, Horsthemke B (1989) Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum Genet* **83**(2)**:** 155-8

Griffiths EA, Gore SD, Hooker CM, Mohammad HP, McDevitt MA, Smith BD, Karp JE, Herman JG, Carraway HE (2010) Epigenetic differences in cytogenetically normal versus abnormal acute myeloid leukemia. *Epigenetics : official journal of the DNA Methylation Society* **5**(7)**:** 590-600

Grimwade D, Hills RK, Moorman AV, Walker H, Chatters S, Goldstone AH, Wheatley K, Harrison CJ, Burnett AK, Group obotNCRIALW (2010) Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among

5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood* **116**(3)**:** 354-365

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**(7)**:** 2322-33

Grossmann V, Haferlach C, Nadarajah N, Fasan A, Weissmann S, Roller A, Eder C, Stopp E, Kern W, Haferlach T, Kohlmann A, Schnittger S (2013) CEBPA double-mutated acute myeloid leukaemia harbours concomitant molecular mutations in 76.8% of cases with TET2 and GATA2 alterations impacting prognosis. *British Journal of Haematology* **161**(5)**:** 649-58

Grossmann V, Kohlmann A, Zenger M, Schindela S, Eder C, Weissmann S, Schnittger S, Kern W, Muller MC, Hochhaus A, Haferlach T, Haferlach C (2011a) A deep-sequencing study of chronic myeloid leukemia patients in blast crisis (BC-CML) detects mutations in 76.9% of cases. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* **25**(3)**:** 557-60

Grossmann V, Schnittger S, Kohlmann A, Eder C, Roller A, Dicker F, Schmid C, Wendtner CM, Staib P, Serve H, Kreuzer KA, Kern W, Haferlach T, Haferlach C (2012) A novel hierarchical prognostic model of AML solely based on molecular mutations. *Blood* **120**(15)**:** 2963-72

Grossmann V, Tiacci E, Holmes AB, Kohlmann A, Martelli MP, Kern W, Spanhol-Rosseto A, Klein H-U, Dugas M, Schindela S, Trifonov V, Schnittger S, Haferlach C, Bassan R, Wells VA, Spinelli O, Chan J, Rossi R, Baldoni S, De Carolis L, Goetze K, Serve H, Peceny R, Kreuzer K-A, Oruzio D, Specchia G, Di Raimondo F, Fabbiano F, Sborgia M, Liso A, Farinelli L, Rambaldi A, Pasqualucci L, Rabadan R, Haferlach T, Falini B (2011b) Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype. *Blood* **118**(23)**:** 6153-6163

Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG, Iqbal K, Shi YG, Deng Z, Szabo PE, Pfeifer GP, Li J, Xu GL (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**(7366)**:** 606-10

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**(1)**:** 77-88

Guglielmelli P, Biamonte F, Rotunno G, Artusi V, Artuso L, Bernardis I, Tenedini E, Pieri L, Paoli C, Mannarelli C, Fjerza R, Rumi E, Stalbovskaya V, Squires M, Cazzola M, Manfredini R, Harrison C, Tagliafico E, Vannucchi AM (2014a) Impact of mutational status on outcomes in myelofibrosis patients treated with ruxolitinib in the COMFORT-II study. *Blood* **123**(14)**:** 2157-60

Guglielmelli P, Lasho TL, Rotunno G, Score J, Mannarelli C, Pancrazzi A, Biamonte F, Pardanani A, Zoi K, Reiter A, Duncombe A, Fanelli T, Pietra D, Rumi E, Finke C, Gangat N, Ketterling RP, Knudson RA, Hanson CA, Bosi A, Pereira A, Manfredini R, Cervantes F, Barosi G, Cazzola M, Cross NC, Vannucchi AM, Tefferi A (2014b) The number of prognostically detrimental mutations and prognosis in primary myelofibrosis: an international study of 797 patients. *Leukemia*

Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, Zhong C, Hu S, Le T, Fan G, Zhu H, Chang Q, Gao Y, Ming G-l, Song H (2014) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci* **17**(2)**:** 215-222

Hackanson B, Bennett KL, Brena RM, Jiang J, Claus R, Chen SS, Blagitko-Dorfs N, Maharry K, Whitman SP, Schmittgen TD, Lubbert M, Marcucci G, Bloomfield CD, Plass C (2008) Epigenetic modification of CCAAT/enhancer binding protein alpha expression in acute myeloid leukemia. *Cancer research* **68**(9)**:** 3142-51

Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, Schnittger S, Sanada M, Kon A, Alpermann T, Yoshida K, Roller A, Nadarajah N, Shiraishi Y, Shiozawa Y, Chiba K, Tanaka H, Koeffler HP, Klein HU, Dugas M, Aburatani H, Kohlmann A, Miyano S, Haferlach C, Kern W, Ogawa S (2014) Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* **28**(2)**:** 241-7

Hajkova H, Markova J, Haskovec C, Sarova I, Fuchs O, Kostecka A, Cetkovsky P, Michalova K, Schwarz J (2012) Decreased DNA methylation in acute myeloid leukemia patients with DNMT3A mutations and prognostic implications of DNA methylation. *Leuk Res* **36**(9)**:** 1128-33

Hann IM, Stevens RF, Goldstone AH, Rees JK, Wheatley K, Gray RG, Burnett AK (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**(7)**:** 2311-8

Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D, Briem E, Zhang K, Irizarry RA, Feinberg AP

(2011) Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* **43**(8)**:** 768-775

Hashimshony T, Zhang J, Keshet I, Bustin M, Cedar H (2003) The role of DNA methylation in setting up chromatin structure during development. *Nature genetics* **34**(2)**:** 187-92

Hayatsu H, Wataya Y, Kai K, Iida S (1970) Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry* **9**(14)**:** 2858-65

Helbling D, Mueller BU, Timchenko NA, Schardt J, Eyer M, Betts DR, Jotterand M, Meyer-Monard S, Fey MF, 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**(4)**:** 1369-75

Herman JG, Baylin SB (2003) Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* **349**(21)**:** 2042-54

Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S, Samid D, Duan DS, Gnarra JR, Linehan WM, et al. (1994) Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* **91**(21)**:** 9700-4

Hoischen A, van Bon BW, Rodriguez-Santiago B, Gilissen C, Vissers LE, de Vries P, Janssen I, van Lier B, Hastings R, Smithson SF, Newbury-Ecob R, Kjaergaard S, Goodship J, McGowan R, Bartholdi D, Rauch A, Peippo M, Cobben JM, Wieczorek D, Gillessen-Kaesbach G, Veltman JA, Brunner HG, de Vries BB (2011) De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat Genet* **43**(8)**:** 729-31

Holliday R (1987) The inheritance of epigenetic defects. *Science* **238**(4824)**:** 163-70

Holliday R, Grigg GW (1993) DNA methylation and mutation. *Mutat Res* **285**(1)**:** 61-7

Holliday R, Pugh JE (1975) DNA modification mechanisms and gene activity during development. *Science* **187**(4173)**:** 226-32

Hollink IH, van den Heuvel-Eibrink MM, Arentsen-Peters ST, Zimmermann M, Peeters JK, Valk PJ, Balgobind BV, Sonneveld E, Kaspers GJ, de Bont ES, Trka J, Baruchel A, Creutzig U, Pieters R, Reinhardt D, Zwaan CM (2011) Characterization of CEBPA mutations and promoter hypermethylation in pediatric acute myeloid leukemia. *Haematologica* **96**(3)**:** 384-92

Hou HA, Lin CC, Chou WC, Liu CY, Chen CY, Tang JL, Lai YJ, Tseng MH, Huang CF, Chiang YC, Lee FY, Kuo YY, Lee MC, Liu MC, Liu CW, Lin LI, Yao M, Huang SY, Ko BS, Hsu SC, Wu SJ, Tsay W, Chen YC, Tien HF (2014) Integration of cytogenetic and molecular alterations in risk stratification of 318 patients with de novo non-M3 acute myeloid leukemia. *Leukemia* **28**(1)**:** 50-8

Hou HA, Lin LI, Chen CY, Tien HF (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* **101**(4)**:** 738-40

Hugues L, Cave H, Philippe N, Pereira S, Fenaux P, Preudhomme C (2005) Mutations of PTPN11 are rare in adult myeloid malignancies. *Haematologica* **90**(6)**:** 853-854

Huret JL, Ahmad M, Arsaban M, Bernheim A, Cigna J, Desangles F, Guignard JC, Jacquemot-Perbal MC, Labarussias M, Leberre V, Malo A, Morel-Pair C, Mossafa H, Potier JC, Texier G, Viguie F, Yau Chun Wan-Senon S, Zasadzinski A, Dessen P (2013) Atlas of genetics and cytogenetics in oncology and haematology in 2013. *Nucleic Acids Res* **41**(Database issue)**:** D920-4

Ibanez M, Such E, Cervera J, Luna I, Gomez-Segui I, Lopez-Pavia M, Dolz S, Barragan E, Fuster O, Llop M, Rodriguez-Veiga R, Avaria A, Oltra S, Senent ML, Moscardo F, Montesinos P, Martinez-Cuadron D, Martin G, Sanz MA (2012) Rapid screening of ASXL1, IDH1, IDH2, and c-CBL mutations in de novo acute myeloid leukemia by high-resolution melting. *J Mol Diagn* **14**(6)**:** 594-601

Illingworth R, Kerr A, DeSousa D, Jørgensen H, Ellis P, Stalker J, Jackson D, Clee C, Plumb R, Rogers J, Humphray S, Cox T, Langford C, Bird A (2008) A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci. *PLoS Biol* **6**(1)**:** e22

Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr ARW, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP (2010) Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. *PLoS Genet* **6**(9)**:** e1001134

Illmer T, Schaich M, Ehninger G, Thiede C (2007) Tyrosine kinase mutations of JAK2 are rare events in AML but influence prognosis of patients with CBF-leukemias. *Haematologica* **92**(1)**:** 137-8

Inoue D, Kitaura J, Togami K, Nishimura K, Enomoto Y, Uchida T, Kagiyama Y, Kawabata KC, Nakahara F, Izawa K, Oki T, Maehara A, Isobe M, Tsuchiya A, Harada Y, Harada H, Ochiya T, Aburatani H, Kimura H, Thol F, Heuser M, Levine RL, Abdel-Wahab O, Kitamura T (2013) Myelodysplastic syndromes are induced by histone methylation-altering ASXL1 mutations. *J Clin Invest* **123**(11)**:** 4627-40

Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash JB, Sabunciyan S, Feinberg AP (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature genetics* **41**(2)**:** 178-86

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**(6047)**:** 1300-3

Itzykson R, Kosmider O, Renneville A, Gelsi-Boyer V, Meggendorfer M, Morabito M, Berthon C, Ades L, Fenaux P, Beyne-Rauzy O, Vey N, Braun T, Haferlach T, Dreyfus F, Cross NC, Preudhomme C, Bernard OA, Fontenay M, Vainchenker W, Schnittger S, Birnbaum D, Droin N, Solary E (2013) Prognostic score including gene mutations in chronic myelomonocytic leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **31**(19)**:** 2428-36

Ivey A, Hills RK, Simpson MA, Jovanovic JV, Gilkes A, Grech A, Patel Y, Bhudia N, Farah H, Mason J, Wall K, Akiki S, Griffiths M, Solomon E, McCaughan F, Linch DC, Gale RE, Vyas P, Freeman SD, Russell N, Burnett AK, Grimwade D, Group UKNCRIAW (2016) Assessment of Minimal Residual Disease in Standard-Risk AML. *N Engl J Med*

Jackson-Grusby L, Beard C, Possemato R, Tudor M, Fambrough D, Csankovszki G, Dausman J, Lee P, Wilson C, Lander E, Jaenisch R (2001) Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nature genetics* **27**(1)**:** 31-9

Jankowska AM, Makishima H, Tiu RV, Szpurka H, Huang Y, Traina F, Visconte V, Sugimoto Y, Prince C, O'Keefe C, Hsi ED, List A, Sekeres MA, Rao A, McDevitt MA, Maciejewski JP (2011) Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A. *Blood* **118**(14)**:** 3932-41

Jiang M, Zhang Y, Fei J, Chang X, Fan W, Qian X, Zhang T, Lu D (2009) Rapid quantification of DNA methylation by measuring relative peak heights in direct bisulfite-PCR sequencing traces. *Lab Invest* **90**(2)**:** 282-290

Johansen LM, Iwama A, Lodie TA, Sasaki K, Felsher DW, Golub TR, Tenen DG (2001) c-Myc is a critical target for c/EBPalpha in granulopoiesis. *Mol Cell Biol* **21**(11)**:** 3789-806

Jones PA, Liang G (2009) Rethinking how DNA methylation patterns are maintained. *Nature reviews Genetics* **10**(11)**:** 805-11

Jost E, do ON, Wilop S, Herman JG, Osieka R, Galm O (2009) Aberrant DNA methylation of the transcription factor C/EBPalpha in acute myelogenous leukemia. *Leukemia research* **33**(3)**:** 443-9

Jost E, Lin Q, Weidner CI, Wilop S, Hoffmann M, Walenda T, Schemionek M, Herrmann O, Zenke M, Brummendorf TH, Koschmieder S, Wagner W (2014) Epimutations mimic genomic mutations of DNMT3A in acute myeloid leukemia. *Leukemia* **28**(6)**:** 1227-1234

Juhl-Christensen C, Bomberg M, Melsvik D, Hokland P, Nyvold CG (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**(4)**:** 273-80

Jurgens G (1985) A group of genes controlling the spatial expression of the bithorax complex in Drosophila. *Nature* **316**(6024)**:** 153-155

Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MDM, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ, Wilson RK, Raphael BJ, Ding L (2013) Mutational landscape and significance across 12 major cancer types. *Nature* **502**(7471)**:** 333-339

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/EBPα mutations play distinct but collaborative roles in leukemogenesis: lessons from clinical data and BMT models. *Blood* **117**(1)**:** 221-233

Katoh M (2013) Functional and cancer genomics of ASXL family members. *Br J Cancer* **109**(2)**:** 299-306

Kelly LM, Gilliland DG (2002) Genetics of myeloid leukemias. *Annu Rev Genomics Hum Genet* **3:** 179-98

Kelly TK, De Carvalho DD, Jones PA (2010) Epigenetic modifications as therapeutic targets. *Nat Biotechnol* **28**(10)**:** 1069-1078

Khulan B, Thompson RF, Ye K, Fazzari MJ, Suzuki M, Stasiek E, Figueroa ME, Glass JL, Chen Q, Montagna C, Hatchwell E, Selzer RR, Richmond TA, Green RD, Melnick A, Greally JM (2006) Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Research* **16**(8)**:** 1046-55

Koschmieder S, Halmos B, Levantini E, Tenen DG (2009) Dysregulation of the C/EBPalpha differentiation pathway in human cancer. *Journal of clinical*

oncology : official journal of the American Society of Clinical Oncology **27**(4)**:** 619-28

Kottaridis PD, Gale RE, Frew ME, Harrison G, Langabeer SE, Belton AA, Walker H, Wheatley K, Bowen DT, Burnett AK, Goldstone AH, Linch DC (2001) The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. *Blood* **98**(6)**:** 1752-9

Krauth MT, Eder C, Alpermann T, Bacher U, Nadarajah N, Kern W, Haferlach C, Haferlach T, Schnittger S (2014) High number of additional genetic lesions in acute myeloid leukemia with t(8;21)/RUNX1-RUNX1T1: frequency and impact on clinical outcome. *Leukemia*

Krebs DL, Hilton DJ (2001) SOCS proteins: negative regulators of cytokine signaling. *Stem Cells* **19**(5)**:** 378-87

Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein BE (2008) Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS genetics* **4**(10)**:** e1000242

Larsen F, Gundersen G, Lopez R, Prydz H (1992) CpG islands as gene markers in the human genome. *Genomics* **13**(4)**:** 1095-1107

Lee SW, Cho YS, Na JM, Park UH, Kang M, Kim EJ, Um SJ (2010) ASXL1 represses retinoic acid receptor-mediated transcription through associating with HP1 and LSD1. *The Journal of biological chemistry* **285**(1)**:** 18-29

Leone G, Teofili L, Voso MT, Lubbert M (2002) DNA methylation and demethylating drugs in myelodysplastic syndromes and secondary leukemias. *Haematologica* **87**(12)**:** 1324-41

Leonhardt H, Page AW, Weier HU, Bestor TH (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71**(5)**:** 865-73

Lewis EB (1978) A gene complex controlling segmentation in Drosophila. *Nature* **276**(5688)**:** 565-570

Lewis RE, Cruse JM, Sanders CM, Webb RN, Suggs JL (2007) Aberrant expression of T-cell markers in acute myeloid leukemia. *Exp Mol Pathol* **83**(3)**:** 462-3

Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koboldt DC, Schmidt H, Zhang Q, Osborne JR, Lin L, O'Laughlin M, McMichael JF, Delehaunty KD, McGrath SD, Fulton LA, Magrini VJ, Vickery TL, Hundal J, Cook LL, Conyers JJ, Swift GW, Reed JP, Alldredge PA, Wylie T, Walker J, Kalicki J, Watson MA, Heath S, Shannon WD, Varghese N, Nagarajan R, Westervelt P, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Wilson RK (2010) DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**(25)**:** 2424-33

Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S, Locke D, Hillier LW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi X, Osborne JR, Minx P, Gordon D, Chinwalla A, Zhao Y, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson M, Baty J, Ivanovich J, Heath S, Shannon WD, Nagarajan R, Walter MJ, Link DC, Graubert TA, DiPersio JF, Wilson RK (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**(7218)**:** 66-72

Li E, Beard C, Jaenisch R (1993) Role for DNA methylation in genomic imprinting. *Nature* **366**(6453)**:** 362-5

Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**(6)**:** 915-26

Li M, Collins R, Jiao Y, Ouillette P, Bixby D, Erba H, Vogelstein B, Kinzler KW, Papadopoulos N, Malek SN (2011) Somatic mutations in the transcriptional corepressor gene BCORL1 in adult acute myelogenous leukemia. *Blood* **118**(22)**:** 5914-7

Li Q, Kopecky KJ, Mohan A, Willman CL, Appelbaum FR, Weick JK, Issa J-PJ (1999) Estrogen Receptor Methylation Is Associated with Improved Survival in Adult Acute Myeloid Leukemia. *Clinical Cancer Research* **5**(5)**:** 1077-1084

Liersch R, Muller-Tidow C, Berdel WE, Krug U (2014) Prognostic factors for acute myeloid leukaemia in adults--biological significance and clinical use. *Br J Haematol* **165**(1)**:** 17-38

Lin LI, Chen CY, Lin DT, Tsay W, Tang JL, Yeh YC, Shen HL, Su FH, Yao M, Huang SY, Tien HF (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**(4)**:** 1372-9

Lin TC, Hou HA, Chou WC, Ou DL, Yu SL, Tien HF, Lin LI (2011) CEBPA methylation as a prognostic biomarker in patients with de novo acute myeloid leukemia. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK* **25**(1)**:** 32-40

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**(7271)**:** 315-22

Lock LF, Takagi N, Martin GR (1987) Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell* **48**(1)**:** 39-46

Loh ML, Reynolds MG, Vattikuti S, Gerbing RB, Alonzo TA, Carlson E, Cheng JW, Lee CM, Lange BJ, Meshinchi S (2004) PTPN11 mutations in pediatric patients with acute myeloid leukemia: results from the Children's Cancer Group. *Leukemia* **18**(11)**:** 1831-1834

Lowenberg B, Downing JR, Burnett A (1999) Acute Myeloid Leukemia. *New England Journal of Medicine* **341**(14)**:** 1051-1062

Lu Y, Chen W, Stein A, Weiss LM, Huang Q (2010) C/EBPA gene mutation and C/EBPA promoter hypermethylation in acute myeloid leukemia with normal cytogenetics. *American journal of hematology* **85**(6)**:** 426-30

Macleod D, Ali RR, Bird A (1998) An alternative promoter in the mouse major histocompatibility complex class II I-Abeta gene: implications for the origin of CpG islands. *Mol Cell Biol* **18**(8)**:** 4433-43

Magini P, Della Monica M, Uzielli ML, Mongelli P, Scarselli G, Gambineri E, Scarano G, Seri M (2012) Two novel patients with Bohring-Opitz syndrome caused by de novo ASXL1 mutations. *Am J Med Genet A* **158A**(4)**:** 917-21

Makishima H, Jankowska AM, McDevitt MA, O'Keefe C, Dujardin S, Cazzolli H, Przychodzen B, Prince C, Nicoll J, Siddaiah H, Shaik M, Szpurka H, Hsi E, Advani A, Paquette R, Maciejewski JP (2011) CBL, CBLB, TET2, ASXL1, and IDH1/2 mutations and additional chromosomal aberrations constitute molecular events in chronic myelogenous leukemia. *Blood* **117**(21)**:** e198-206

Marcucci G, Maharry K, Radmacher MD, Mrozek K, Vukosavljevic T, Paschka P, Whitman SP, Langer C, Baldus CD, Liu CG, Ruppert AS, Powell BL, Carroll AJ, Caligiuri MA, Kolitz JE, Larson RA, Bloomfield CD (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. *Journal of*

clinical oncology : official journal of the American Society of Clinical Oncology **26**(31)**:** 5078-87

Marcucci G, Maharry K, Wu YZ, Radmacher MD, Mrozek K, Margeson D, Holland KB, Whitman SP, Becker H, Schwind S, Metzeler KH, Powell BL, Carter TH, Kolitz JE, Wetzler M, Carroll AJ, Baer MR, Caligiuri MA, Larson RA, Bloomfield CD (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. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**(14)**:** 2348-55

Marcucci G, Metzeler KH, Schwind S, Becker H, Maharry K, Mrozek K, Radmacher MD, Kohlschmidt J, Nicolet D, Whitman SP, Wu YZ, Powell BL, Carter TH, Kolitz JE, Wetzler M, Carroll AJ, Baer MR, Moore JO, Caligiuri MA, Larson RA, Bloomfield CD (2012) Age-related prognostic impact of different types of DNMT3A mutations in adults with primary cytogenetically normal acute myeloid leukemia. *J Clin Oncol* **30**(7)**:** 742-50

Martinez-Aviles L, Besses C, Alvarez-Larran A, Torres E, Serrano S, Bellosillo B (2012) TET2, ASXL1, IDH1, IDH2, and c-CBL genes in JAK2- and MPL-negative myeloproliferative neoplasms. *Ann Hematol* **91**(4)**:** 533-41

Mead AJ, Linch DC, Hills RK, Wheatley K, Burnett AK, Gale RE (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**(4)**:** 1262-70

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**(7205)**:** 766-70

Melki JR, Vincent PC, Clark SJ (1999) Concurrent DNA Hypermethylation of Multiple Genes in Acute Myeloid Leukemia. *Cancer Research* **59**(15)**:** 3730-3740

Mendler JH, Maharry K, Radmacher MD, Mrozek K, Becker H, Metzeler KH, Schwind S, Whitman SP, Khalife J, Kohlschmidt J, Nicolet D, Powell BL, Carter TH, Wetzler M, Moore JO, Kolitz JE, Baer MR, Carroll AJ, Larson RA, Caligiuri MA, Marcucci G, Bloomfield CD (2012) RUNX1 mutations are associated with poor outcome in younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and MicroRNA expression signatures. *J Clin Oncol* **30**(25)**:** 3109-18

Metzeler KH, Becker H, Maharry K, Radmacher MD, Kohlschmidt J, Mrozek K, Nicolet D, Whitman SP, Wu YZ, Schwind S, Powell BL, Carter TH, Wetzler M,

Moore JO, Kolitz JE, Baer MR, Carroll AJ, Larson RA, Caligiuri MA, Marcucci G, Bloomfield CD (2011a) ASXL1 mutations identify a high-risk subgroup of older patients with primary cytogenetically normal AML within the ELN Favorable genetic category. *Blood* **118**(26)**:** 6920-9

Metzeler KH, Maharry K, Radmacher MD, Mrozek K, Margeson D, Becker H, Curfman J, Holland KB, Schwind S, Whitman SP, Wu YZ, Blum W, Powell BL, Carter TH, Wetzler M, Moore JO, Kolitz JE, Baer MR, Carroll AJ, Larson RA, Caligiuri MA, Marcucci G, Bloomfield CD (2011b) TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **29**(10)**:** 1373-81

Milne TA, Sinclair DA, Brock HW (1999) The Additional sex combs gene of Drosophila is required for activation and repression of homeotic loci, and interacts specifically with Polycomb and super sex combs. *Mol Gen Genet* **261**(4-5)**:** 753-61

Mizuki M, Schwäble J, Steur C, Choudhary C, Agrawal S, Sargin B, Steffen B, Matsumura I, Kanakura Y, Böhmer FD, Müller-Tidow C, Berdel WE, Serve H (2003) Suppression of myeloid transcription factors and induction of STAT response genes by AML-specific Flt3 mutations. *Blood* **101**(8)**:** 3164-3173

Mohandas T, Sparkes RS, Shapiro LJ (1981) Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* **211**(4480)**:** 393-6

Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schubeler D (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* **30**(6)**:** 755-66

Monk M, Boubelik M, Lehnert S (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**(3)**:** 371-82

Morris T, Lowe R (2012) Report on the Infinium 450k methylation array analysis workshop: April 20, 2012 UCL, London, UK. *Epigenetics* **7**(8)**:** 961-2

Mueller BU, Pabst T (2006) C/EBPalpha and the pathophysiology of acute myeloid leukemia. *Curr Opin Hematol* **13**(1)**:** 7-14

Musialik E, Bujko M, Kober P, Grygorowicz MA, Libura M, Przestrzelska M, Juszczynski P, Borg K, Florek I, Jakobczyk M, Baranowska A, Siedlecki JA (2014) Comparison of promoter DNA methylation and expression levels of

genes encoding CCAAT/enhancer binding proteins in AML patients. *Leuk Res* **38**(7)**:** 850-6

Myohanen SK, Baylin SB, Herman JG (1998) Hypermethylation can selectively silence individual p16ink4A alleles in neoplasia. *Cancer research* **58**(4)**:** 591-3

Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, Katz E, Dixon JM, Harrison DJ, Meehan RR (2012) Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Research* **22**(3)**:** 467-477

Nomdedeu J, Hoyos M, Carricondo M, Esteve J, Bussaglia E, Estivill C, Ribera JM, Duarte R, Salamero O, Gallardo D, Pedro C, Aventin A, Brunet S, Sierra J (2012) Adverse impact of IDH1 and IDH2 mutations in primary AML: experience of the Spanish CETLAM group. *Leuk Res* **36**(8)**:** 990-7

Nüsslein-Volhard C, Wieschaus E, Kluding H (1984) Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster. *Wilhelm Roux' Archiv* **193**(5)**:** 267-282

Ohlsson E, Schuster MB, Hasemann M, Porse BT (2016) The multifaceted functions of C/EBPalpha in normal and malignant haematopoiesis. *Leukemia* **30**(4)**:** 767-75

Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W, Berman DM, Jenuwein T, Pruitt K, Sharkis SJ, Watkins DN, Herman JG, Baylin SB (2007) A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nature genetics* **39**(2)**:** 237-42

Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**(3)**:** 247-57

Ooi SK, Bestor TH (2008) The colorful history of active DNA demethylation. *Cell* **133**(7)**:** 1145-8

Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**(7154)**:** 714-7

Orkin SH, Zon LI (2008) Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**(4)**:** 631-44

Pabst T, Eyholzer M, Fos J, Mueller BU (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**(8)**:** 1343-6

Pabst T, Mueller BU, Harakawa N, Schoch C, Haferlach T, Behre G, Hiddemann W, Zhang DE, Tenen DG (2001a) AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. *Nat Med* **7**(4)**:** 444-51

Pabst T, Mueller BU, Zhang P, Radomska HS, Narravula S, Schnittger S, Behre G, Hiddemann W, Tenen DG (2001b) Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-[alpha] (C/EBP[alpha]), in acute myeloid leukemia. *Nat Genet* **27**(3)**:** 263-270

Paschka P, Marcucci G, Ruppert AS, Mrozek K, Chen H, Kittles RA, Vukosavljevic T, Perrotti D, Vardiman JW, Carroll AJ, Kolitz JE, Larson RA, Bloomfield CD (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. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **24**(24)**:** 3904-11

Paschka P, Marcucci G, Ruppert AS, Whitman SP, Mrozek K, Maharry K, Langer C, Baldus CD, Zhao W, Powell BL, Baer MR, Carroll AJ, Caligiuri MA, Kolitz JE, Larson RA, Bloomfield CD (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. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **26**(28)**:** 4595-602

Paschka P, Schlenk RF, Gaidzik VI, Habdank M, Kronke J, Bullinger L, Spath D, Kayser S, Zucknick M, Gotze K, Horst HA, 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. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**(22)**:** 3636-43

Paschka P, Schlenk RF, Herzig J, Aulitzky T, Gaidzik VI, Bullinger L, Habdank M, Corbacioglu A, Spath D, Kohne C-H, Kundgen A, von Lilienfeld-Toal M, Held G, Horst HA, Gotze K, Bentz M, Krauter J, Ganser A, Dohner H, Dohner K (2011) ASXL1 Mutations Predict for Resistance to Chemotherapy and Inferior Outcome in Younger Adult Patients with Acute Myeloid Leukemia (AML): A Study of the German-Austrian AMLSG. *ASH Annual Meeting Abstracts* **118**(21)**:** 412-

Pasquet M, Bellanne-Chantelot C, Tavitian S, Prade N, Beaupain B, Larochelle O, Petit A, Rohrlich P, Ferrand C, Van Den Neste E, Poirel HA, Lamy T, Ouachee-Chardin M, Mansat-De Mas V, Corre J, Recher C, Plat G, Bachelerie F, Donadieu J, Delabesse E (2013) High frequency of GATA2 mutations in patients with mild chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia. *Blood* **121**(5)**:** 822-9

Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J, Van Vlierberghe P, Dolgalev I, Thomas S, Aminova O, Huberman K, Cheng J, Viale A, Socci ND, Heguy A, Cherry A, Vance G, Higgins RR, Ketterling RP, Gallagher RE, Litzow M, van den Brink MR, Lazarus HM, Rowe JM, Luger S, Ferrando A, Paietta E, Tallman MS, Melnick A, Abdel-Wahab O, Levine RL (2012) Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *The New England journal of medicine* **366**(12)**:** 1079-89

Patnaik MM, Itzykson R, Lasho TL, Kosmider O, Finke CM, Hanson CA, Knudson RA, Ketterling RP, Tefferi A, Solary E (2014) ASXL1 and SETBP1 mutations and their prognostic contribution in chronic myelomonocytic leukemia: a two center study of 466 patients. *Leukemia*

Patterson K, Molloy L, Qu W, Clark S (2011) DNA Methylation: Bisulphite Modification and Analysis. *Journal of Visualized Experiments : JoVE*(56)**:** 3170

Perez B, Kosmider O, Cassinat B, Renneville A, Lachenaud J, Kaltenbach S, Bertrand Y, Baruchel A, Chomienne C, Fontenay M, Preudhomme C, Cave H (2010) Genetic typing of CBL, ASXL1, RUNX1, TET2 and JAK2 in juvenile myelomonocytic leukaemia reveals a genetic profile distinct from chronic myelomonocytic leukaemia. *British journal of haematology* **151**(5)**:** 460-8

Perez C, Pascual M, Martin-Subero JI, Bellosillo B, Segura V, Delabesse E, Alvarez S, Larrayoz MJ, Rifon J, Cigudosa JC, Besses C, Calasanz MJ, Cross NC, Prosper F, Agirre X (2013) Aberrant DNA methylation profile of chronic and transformed classic Philadelphia-negative myeloproliferative neoplasms. *Haematologica* **98**(9)**:** 1414-20

Pike BL, Greiner TC, Wang X, Weisenburger DD, Hsu YH, Renaud G, Wolfsberg TG, Kim M, Weisenberger DJ, Siegmund KD, Ye W, Groshen S, Mehrian-Shai R, Delabie J, Chan WC, Laird PW, Hacia JG (2008) DNA methylation profiles in diffuse large B-cell lymphoma and their relationship to gene expression status. *Leukemia* **22**(5)**:** 1035-1043

Porse BT, Pedersen TÅ, Xu X, Lindberg B, Wewer UM, Friis-Hansen L, Nerlov C (2001) E2F Repression by C/EBPα Is Required for Adipogenesis and Granulopoiesis In Vivo. *Cell* **107**(2)**:** 247-258

Pratcorona M, Abbas S, Sanders MA, Koenders JE, Kavelaars FG, Erpelinck-Verschueren CA, Zeilemakers A, Lowenberg B, Valk PJ (2012) Acquired mutations in ASXL1 in acute myeloid leukemia: prevalence and prognostic value. *Haematologica* **97**(3)**:** 388-92

Preudhomme C, Sagot C, Boissel N, Cayuela JM, Tigaud I, de Botton 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**(8)**:** 2717-23

Price ME, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ, Robinson WP, Kobor MS (2013) Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* **6**(1)**:** 4

Qu Y, Lennartsson A, Gaidzik VI, Deneberg S, Karimi M, Bengtzen S, Hoglund M, Bullinger L, Dohner K, Lehmann S (2014) Differential methylation in CN-AML preferentially targets non-CGI regions and is dictated by DNMT3A mutational status and associated with predominant hypomethylation of HOX genes. *Epigenetics* **9**(8)

Ramji DP, Foka P (2002) CCAAT/enhancer-binding proteins: structure, function and regulation. *Biochem J* **365**(3)**:** 561-575

Rampal R, Alkalin A, Madzo J, Vasanthakumar A, Pronier E, Patel J, Li Y, Ahn J, Abdel-Wahab O, Shih A, Lu C, Ward Patrick S, Tsai Jennifer J, Hricik T, Tosello V, Tallman Jacob E, Zhao X, Daniels D, Dai Q, Ciminio L, Aifantis I, He C, Fuks F, Tallman Martin S, Ferrando A, Nimer S, Paietta E, Thompson Craig B, Licht Jonathan D, Mason Christopher E, Godley Lucy A, Melnick A, Figueroa Maria E, Levine Ross L (2014) DNA Hydroxymethylation Profiling Reveals that WT1 Mutations Result in Loss of TET2 Function in Acute Myeloid Leukemia. *Cell Reports* **9**(5)**:** 1841-1855

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A* **97**(10)**:** 5237-42

Rauch T, Li H, Wu X, Pfeifer GP (2006) MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer Res* **66**(16)**:** 7939-47

Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP (2009) A human B cell methylome at 100-base pair resolution. *Proceedings of the National Academy of Sciences of the United States of America* **106**(3)**:** 671-8

Renneville A, Abdelali RB, Chevret S, Nibourel O, Cheok M, Pautas C, Dulery R, Boyer T, Cayuela JM, Hayette S, Raffoux E, Farhat H, Boissel N, Terre C, Dombret H, Castaigne S, Preudhomme C (2014) Clinical impact of gene mutations and lesions detected by SNP-array karyotyping in acute myeloid leukemia patients in the context of gemtuzumab ozogamicin treatment: Results of the ALFA-0701 trial. *Oncotarget*

Renneville A, Boissel N, Gachard N, Naguib D, Bastard C, de Botton S, Nibourel O, Pautas C, Reman O, Thomas X, Gardin C, Terre C, Castaigne S, Preudhomme C, Dombret H (2009) 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**(21)**:** 5090-3

Renneville A, Boissel N, Nibourel O, Berthon C, Helevaut N, Gardin C, Cayuela JM, 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**(6)**:** 1247-54

Ribeiro AF, Pratcorona M, Erpelinck-Verschueren C, Rockova V, Sanders M, Abbas S, Figueroa ME, Zeilemaker A, Melnick A, Lowenberg B, Valk PJ, Delwel R (2012) Mutant DNMT3A: a marker of poor prognosis in acute myeloid leukemia. *Blood* **119**(24)**:** 5824-31

Ricci C, Spinelli O, Salmoiraghi S, Finazzi G, Carobbio A, Rambaldi A (2012) ASXL1 mutations in primary and secondary myelofibrosis. *British Journal of Haematology* **156**(3)**:** 404-7

Rideout WM, 3rd, Coetzee GA, Olumi AF, Jones PA (1990) 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science* **249**(4974)**:** 1288-90

Riggs AD (1975) X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* **14**(1)**:** 9-25

Robinson MD, Stirzaker C, Statham AL, Coolen MW, Song JZ, Nair SS, Strbenac D, Speed TP, Clark SJ (2010) Evaluation of affinity-based genome-wide DNA methylation data: effects of CpG density, amplification bias, and copy number variation. *Genome Research* **20**(12)**:** 1719-29

Roche-Lestienne C, Marceau A, Labis E, Nibourel O, Coiteux V, Guilhot J, Legros L, Nicolini F, Rousselot P, Gardembas M, Helevaut N, Frimat C, Mahon FX, Guilhot F, Preudhomme C (2011) Mutation analysis of TET2, IDH1, IDH2 and ASXL1 in chronic myeloid leukemia. *Leukemia* **25**(10)**:** 1661-4

Rocquain J, Carbuccia N, Trouplin V, Raynaud S, Murati A, Nezri M, Tadrist Z, Olschwang S, Vey N, Birnbaum D, Gelsi-Boyer V, Mozziconacci MJ (2010) Combined mutations of ASXL1, CBL, FLT3, IDH1, IDH2, JAK2, KRAS, NPM1, NRAS, RUNX1, TET2 and WT1 genes in myelodysplastic syndromes and acute myeloid leukemias. *BMC cancer* **10:** 401

Rucker FG, Schlenk RF, Bullinger L, Kayser S, Teleanu V, Kett H, Habdank M, Kugler CM, Holzmann K, Gaidzik VI, 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**(9)**:** 2114-21

Russell GJ, Walker PM, Elton RA, Subak-Sharpe JH (1976) Doublet frequency analysis of fractionated vertebrate nuclear DNA. *J Mol Biol* **108**(1)**:** 1-23

Russler-Germain DA, Spencer DH, Young MA, Lamprecht TL, Miller CA, Fulton R, Meyer MR, Erdmann-Gilmore P, Townsend RR, Wilson RK, Ley TJ (2014) The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer Cell* **25**(4)**:** 442-54

Sado T, Fenner MH, Tan SS, Tam P, Shioda T, Li E (2000) X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Developmental biology* **225**(2)**:** 294-303

Sado T, Okano M, Li E, Sasaki H (2004) De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. *Development* **131**(5)**:** 975-82

Saito Y, Nakahata S, Yamakawa N, Kaneda K, Ichihara E, Suekane A, Morishita K (2011) CD52 as a molecular target for immunotherapy to treat acute myeloid leukemia with high EVI1 expression. *Leukemia* **25**(6)**:** 921-31

Sakai T, Toguchida J, Ohtani N, Yandell DW, Rapaport JM, Dryja TP (1991) Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *American journal of human genetics* **48**(5)**:** 880-8

Samollow PB, Robinson ES, Ford AL, Vandeberg JL (1995) Developmental progression of Gpd expression from the inactive X chromosome of the Virginia opossum. *Dev Genet* **16**(4)**:** 367-78

Sanchez R, Zhou MM (2011) The PHD finger: a versatile epigenome reader. *Trends Biochem Sci* **36**(7)**:** 364-72

Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature reviews Genetics* **8**(6)**:** 424-36

Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M (2011) Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics : official journal of the DNA Methylation Society* **6**(6)**:** 692-702

Santana-Lemos BA, de Lima Lange AP, de Lira Benicio MT, Jose TD, Lucena-Araujo AR, Krause A, Thome CH, Rego EM (2011) The CEBPA gene is down-regulated in acute promyelocytic leukemia and its upstream promoter, but not the core promoter, is highly methylated. *Haematologica* **96**(4)**:** 617-20

Saxonov S, Berg P, Brutlag DL (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America* **103**(5)**:** 1412-1417

Scheuermann JC, de Ayala Alonso AG, Oktaba K, Ly-Hartig N, McGinty RK, Fraterman S, Wilm M, Muir TW, Muller J (2010) Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465**(7295)**:** 243-247

Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, Bergman Y, Simon I, Cedar H (2007) Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* **39**(2)**:** 232-236

Schnittger S, Eder C, Jeromin S, Alpermann T, Fasan A, Grossmann V, Kohlmann A, Illig T, Klopp N, Wichmann HE, Kreuzer KA, Schmid C, Staib P, Peceny R, Schmitz N, Kern W, Haferlach C, Haferlach T (2013) ASXL1 exon 12 mutations are frequent in AML with intermediate risk karyotype and are independently associated with an adverse outcome. *Leukemia* **27**(1)**:** 82-91

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**(5)**:** 796-804

Schnittger S, Kohl TM, Haferlach T, Kern W, Hiddemann W, Spiekermann K, Schoch C (2006) KIT-D816 mutations in AML1-ETO-positive AML are

associated with impaired event-free and overall survival. *Blood* **107**(5)**:** 1791-1799

Schnittger S, Schoch C, Kern W, Mecucci C, Tschulik C, Martelli MF, 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**(12)**:** 3733-9

Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G (2007) Genome regulation by polycomb and trithorax proteins. *Cell* **128**(4)**:** 735-45

Schwaab J, Schnittger S, Sotlar K, Walz C, Fabarius A, Pfirrmann M, Kohlmann A, Grossmann V, Meggendorfer M, Horny HP, Valent P, Jawhar M, Teichmann M, Metzgeroth G, Erben P, Ernst T, Hochhaus A, Haferlach T, Hofmann WK, Cross NC, Reiter A (2013) Comprehensive mutational profiling in advanced systemic mastocytosis. *Blood* **122**(14)**:** 2460-6

Scott LM, Civin CI, Rorth P, Friedman AD (1992) A novel temporal expression pattern of three C/EBP family members in differentiating myelomonocytic cells. *Blood* **80**(7)**:** 1725-35

Shen H, Laird PW (2013) Interplay between the cancer genome and epigenome. *Cell* **153**(1)**:** 38-55

Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, Ahmed S, Shu J, Chen X, Waterland RA, Issa JP (2007) Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet* **3**(10)**:** 2023-36

Shen Y, Zhu YM, Fan X, Shi JY, Wang QR, Yan XJ, Gu ZH, Wang YY, Chen B, Jiang CL, Yan H, Chen FF, Chen HM, Chen Z, Jin J, Chen SJ (2011) Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia. *Blood* **118**(20)**:** 5593-603

Shih AH, Abdel-Wahab O, Patel JP, Levine RL (2012) The role of mutations in epigenetic regulators in myeloid malignancies. *Nat Rev Cancer* **12**(9)**:** 599-612

Shih LY, Liang DC, Huang CF, Wu JH, Lin TL, Wang PN, Dunn P, Kuo MC, Tang TC (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**(4)**:** 604-9

Silverman LR, McKenzie DR, Peterson BL, Holland JF, Backstrom JT, Beach CL, Larson RA (2006) Further analysis of trials with azacitidine in patients with

myelodysplastic syndrome: studies 8421, 8921, and 9221 by the Cancer and Leukemia Group B. *J Clin Oncol* **24**(24)**:** 3895-903

Slomiany BA, D'Arigo KL, Kelly MM, Kurtz DT (2000) C/EBPalpha inhibits cell growth via direct repression of E2F-DP-mediated transcription. *Mol Cell Biol* **20**(16)**:** 5986-97

Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G (2011) Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature genetics* **43**(8)**:** 811-4

Smith ML, Hills RK, Grimwade D (2011) Independent prognostic variables in acute myeloid leukaemia. *Blood Rev* **25**(1)**:** 39-51

Snaddon J, Smith ML, Neat M, Cambal-Parrales M, Dixon-McIver A, Arch R, Amess JA, Rohatiner AZ, Lister TA, Fitzgibbon J (2003) Mutations of CEBPA in acute myeloid leukemia FAB types M1 and M2. *Genes, chromosomes & cancer* **37**(1)**:** 72-8

Sproul D, Kitchen R, Nestor C, Dixon JM, Sims A, Harrison D, Ramsahoye B, Meehan R (2012) Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. *Genome Biol* **13**(10)**:** R84

Sproul D, Nestor C, Culley J, Dickson JH, Dixon JM, Harrison DJ, Meehan RR, Sims AH, Ramsahoye BH (2011) Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America* **108**(11)**:** 4364-9

Stein BL, Williams DM, O'Keefe C, Rogers O, Ingersoll RG, Spivak JL, Verma A, Maciejewski JP, McDevitt MA, Moliterno AR (2011) Disruption of the ASXL1 gene is frequent in primary, post-essential thrombocytosis and post-polycythemia vera myelofibrosis, but not essential thrombocytosis or polycythemia vera: analysis of molecular genetics and clinical phenotypes. *Haematologica* **96**(10)**:** 1462-9

Stein EM, Tallman MS (2016) Emerging therapeutic drugs for AML. *Blood* **127**(1)**:** 71-78

Stein R, Razin A, Cedar H (1982) In vitro Methylation of the Hamster Adenine Phosphoribosyltransferase Gene Inhibits Its Expression in Mouse L Cells. *Proceedings of the National Academy of Sciences of the United States of America* **79**(11)**:** 3418-3422

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**(3)**:** 237-51

Stratton MR (2011) Exploring the genomes of cancer cells: Progress and promise. *Science* **331**(6024)**:** 1553-1558

Sugimoto Y, Muramatsu H, Makishima H, Prince C, Jankowska AM, Yoshida N, Xu Y, Nishio N, Hama A, Yagasaki H, Takahashi Y, Kato K, Manabe A, Kojima S, Maciejewski JP (2010) Spectrum of molecular defects in juvenile myelomonocytic leukaemia includes ASXL1 mutations. *Br J Haematol* **150**(1)**:** 83-7

Swerdlow SH, International Agency for Research on C, World Health O (2008) *WHO classification of tumours of haematopoietic and lymphoid tissues*. Lyon: International Agency for Research on Cancer

Szankasi P, Ho AK, Bahler DW, Efimova O, Kelley TW (2011) Combined testing for CCAAT/enhancer-binding protein alpha (CEBPA) mutations and promoter methylation in acute myeloid leukemia demonstrates shared phenotypic features. *Leukemia research* **35**(2)**:** 200-7

Szpurka H, Jankowska AM, Makishima H, Bodo J, Bejanyan N, Hsi ED, Sekeres MA, Maciejewski JP (2010) Spectrum of mutations in RARS-T patients includes TET2 and ASXL1 mutations. *Leuk Res* **34**(8)**:** 969-73

Tada Y, Brena RM, Hackanson B, Morrison C, Otterson GA, Plass C (2006) Epigenetic modulation of tumor suppressor CCAAT/enhancer binding protein alpha activity in lung cancer. *J Natl Cancer Inst* **98**(6)**:** 396-406

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**(5929)**:** 930-5

Tang JL, Hou HA, Chen CY, Liu CY, Chou WC, Tseng MH, Huang CF, Lee FY, Liu MC, Yao M, Huang SY, Ko BS, Hsu SC, Wu SJ, Tsay W, Chen YC, Lin LI, Tien HF (2009) AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations. *Blood* **114**(26)**:** 5352-61

Taskesen E, Bullinger L, Corbacioglu A, Sanders MA, Erpelinck CA, Wouters BJ, van der Poel-van de Luytgaarde SC, Damm F, Krauter J, Ganser A, Schlenk RF, Lowenberg B, Delwel R, Dohner H, Valk PJ, 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**(8)**:** 2469-75

Tefferi A, Guglielmelli P, Lasho TL, Rotunno G, Finke C, Mannarelli C, Belachew AA, Pancrazzi A, Wassie EA, Ketterling RP, Hanson CA, Pardanani A, Vannucchi AM (2014) CALR and ASXL1 mutations-based molecular prognostication in primary myelofibrosis: an international study of 570 patients. *Leukemia*

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**(10)**:** 4011-20

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**(12)**:** 4326-35

Thirlwell C, Eymard M, Feber A, Teschendorff A, Pearce K, Lechner M, Widschwendter M, Beck S (2010) Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip. *Methods* **52**(3)**:** 248-254

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 (2011a) Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia. *J Clin Oncol* **29**(21)**:** 2889-96

Thol F, Friesen I, Damm F, Yun H, Weissinger EM, Krauter J, Wagner K, Chaturvedi A, Sharma A, Wichmann M, Gohring G, Schumann C, Bug G, Ottmann O, Hofmann WK, Schlegelberger B, Heuser M, Ganser A (2011b) Prognostic significance of ASXL1 mutations in patients with myelodysplastic syndromes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **29**(18)**:** 2499-506

Togami K, Kitaura J, Uchida T, Inoue D, Nishimura K, Kawabata KC, Nagase R, Horikawa S, Izawa K, Fukuyama T, Nakahara F, Oki T, Harada Y, Harada H, Aburatani H, Kitamura T (2015) A C-terminal mutant of CCAAT-enhancer-binding protein alpha (C/EBPalpha-Cm) downregulates Csf1r, a potent accelerator in the progression of acute myeloid leukemia with C/EBPalpha-Cm. *Exp Hematol* **43**(4)**:** 300-8.e1

Traina F, Visconte V, Jankowska AM, Makishima H, O'Keefe CL, Elson P, Han Y, Hsieh FH, Sekeres MA, Mali RS, Kalaycio M, Lichtin AE, Advani AS, Duong HK, Copelan E, Kapur R, Olalla Saad ST, Maciejewski JP, Tiu RV (2012) Single Nucleotide Polymorphism Array Lesions, TET2, DNMT3A, ASXL1 and CBL Mutations Are Present in Systemic Mastocytosis. *PLoS One* **7**(8)**:** e43090

Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, Beverloo HB, Moorhouse MJ, van der Spek PJ, Lowenberg B, Delwel R (2004) Prognostically useful gene-expression profiles in acute myeloid leukemia. *The New England journal of medicine* **350**(16)**:** 1617-28

Van Vlierberghe P, Patel J, Abdel-Wahab O, Lobry C, Hedvat CV, Balbin M, Nicolas C, Payer AR, Fernandez HF, Tallman MS, Paietta E, Melnick A, Vandenberghe P, Speleman F, Aifantis I, Cools J, Levine R, Ferrando A (2011) PHF6 mutations in adult acute myeloid leukemia. *Leukemia* **25**(1)**:** 130-4

van Vliet MH, Burgmer P, de Quartel L, Brand JP, de Best LC, Vietor H, Lowenberg B, Valk PJ, van Beers EH (2013) Detection of CEBPA double mutants in acute myeloid leukemia using a custom gene expression array. *Genet Test Mol Biomarkers* **17**(5)**:** 395-400

Venolia L, Gartler SM (1983) Comparison of transformation efficiency of human active and inactive X-chromosomal DNA. *Nature* **302**(5903)**:** 82-3

Verhaak RG, Goudswaard CS, van Putten W, Bijl MA, Sanders MA, Hugens W, Uitterlinden AG, Erpelinck CA, Delwel R, Lowenberg B, Valk PJ (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**(12)**:** 3747-54

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 TA, 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. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **26**(33)**:** 5429-35

Voso MT, Santini V, Fabiani E, Fianchi L, Criscuolo M, Falconi G, Guidi F, Hohaus S, Leone G (2014) Why methylation is not a marker predictive of response to hypomethylating agents. *Haematologica* **99**(4)**:** 613-9

Wagner K, Damm F, Gohring G, Gorlich K, Heuser M, Schafer I, Ottmann O, Lubbert M, Heit W, Kanz L, Schlimok G, Raghavachar AA, Fiedler W, Kirchner

HH, 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. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**(14)**:** 2356-64

Wang H, Iakova P, Wilde M, Welm A, Goode T, Roesler WJ, Timchenko NA (2001) C/EBPalpha arrests cell proliferation through direct inhibition of Cdk2 and Cdk4. *Mol Cell* **8**(4)**:** 817-28

Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, Zhang Y (2004) Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**(7010)**:** 873-878

Wang J, Ai X, Gale RP, Xu Z, Qin T, Fang L, Zhang H, Pan L, Hu N, Zhang Y, Xiao Z (2013a) TET2, ASXL1 and EZH2 mutations in Chinese with myelodysplastic syndromes. *Leukemia research* **37**(3)**:** 305-11

Wang J, Li Z, He Y, Pan F, Chen S, Rhodes S, Nguyen L, Yuan J, Jiang L, Yang X, Weeks O, Liu Z, Zhou J, Ni H, Cai C-L, Xu M, Yang F-C (2014) Loss of Asxl1 leads to myelodysplastic syndrome–like disease in mice. *Blood* **123**(4)**:** 541-553

Wang N-d, Milton JF, Bradley A, Ou CN, Abdelsayed SV, Wilde MD, Taylor LR, Wilson DR, Darlington GJ (1995) Impaired Energy Homeostasis in C/EBPα Knockout Mice. *Science* **269**(5227)**:** 1108-1112

Wang X, Dai H, Wang Q, Xu Y, Wang Y, Sun A, Ruan J, Chen S, Wu D (2013b) EZH2 Mutations Are Related to Low Blast Percentage in Bone Marrow and -7/del(7q) in De Novo Acute Myeloid Leukemia. *PLoS One* **8**(4)**:** e61341

Wang Y, Krivtsov AV, Sinha AU, North TE, Goessling W, Feng Z, Zon LI, Armstrong SA (2010) The Wnt/β-Catenin Pathway Is Required for the Development of Leukemia Stem Cells in AML. *Science* **327**(5973)**:** 1650-1653

Warnecke PM, Stirzaker C, Melki JR, Millar DS, Paul CL, Clark SJ (1997) Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res* **25**(21)**:** 4422-6

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature genetics* **37**(8)**:** 853-62

Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature genetics* **39**(4)**:** 457-66

Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, Wartman LD, Lamprecht TL, Liu F, Xia J, Kandoth C, Fulton RS, McLellan MD, Dooling DJ, Wallis JW, Chen K, Harris CC, Schmidt HK, Kalicki-Veizer JM, Lu C, Zhang Q, Lin L, O'Laughlin MD, McMichael JF, Delehaunty KD, Fulton LA, Magrini VJ, McGrath SD, Demeter RT, Vickery TL, Hundal J, Cook LL, Swift GW, Reed JP, Alldredge PA, Wylie TN, Walker JR, Watson MA, Heath SE, Shannon WD, Varghese N, Nagarajan R, Payton JE, Baty JD, Kulkarni S, Klco JM, Tomasson MH, Westervelt P, Walter MJ, Graubert TA, DiPersio JF, Ding L, Mardis ER, Wilson RK (2012) The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**(2)**:** 264-78

Wen XM, Lin J, Yang J, Yao DM, Deng ZQ, Tang CY, Xiao GF, Yang L, Ma JC, Hu JB, Qian W, Qian J (2014) Double CEBPA mutations are prognostically favorable in non-M3 acute myeloid leukemia patients with wild-type NPM1 and FLT3-ITD. *International journal of clinical and experimental pathology* **7**(10)**:** 6832-40

Wheatley K, Burnett AK, Goldstone AH, Gray RG, Hann IM, Harrison CJ, Rees JK, Stevens RF, 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**(1)**:** 69-79

Whitman SP, Ruppert AS, Radmacher MD, Mrozek K, Paschka P, Langer C, Baldus CD, Wen J, Racke F, Powell BL, Kolitz JE, Larson RA, Caligiuri MA, Marcucci G, Bloomfield CD (2008) FLT3 D835/I836 mutations are associated with poor disease-free survival and a distinct gene-expression signature among younger adults with de novo cytogenetically normal acute myeloid leukemia lacking FLT3 internal tandem duplications. *Blood* **111**(3)**:** 1552-9

Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PAC, Rappsilber J, Helin K (2011) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**(7347)**:** 343-348

Wilop S, Fernandez AF, Jost E, Herman JG, Brummendorf TH, Esteller M, Galm O (2011) Array-based DNA methylation profiling in acute myeloid leukaemia. *Br J Haematol* **155**(1)**:** 65-72

Wong TN, Ramsingh G, Young AL, Miller CA, Touma W, Welch JS, Lamprecht TL, Shen D, Hundal J, Fulton RS, Heath S, Baty JD, Klco JM, Ding L, Mardis ER, Westervelt P, DiPersio JF, Walter MJ, Graubert TA, Ley TJ, Druley TE, Link

DC, Wilson RK (2015) Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature* **518**(7540)**:** 552-555

Wouters BJ, Delwel R (2016) Epigenetics and approaches to targeted epigenetic therapy in acute myeloid leukemia. *Blood* **127**(1)**:** 42-52

Wouters BJ, Jorda MA, Keeshan K, Louwers I, Erpelinck-Verschueren CA, Tielemans D, Langerak AW, He Y, Yashiro-Ohtani Y, Zhang P, Hetherington CJ, Verhaak RG, Valk PJ, Lowenberg B, Tenen DG, Pear WS, Delwel R (2007) Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood* **110**(10)**:** 3706-14

Wouters BJ, Lowenberg B, Erpelinck-Verschueren CA, van Putten WL, Valk PJ, 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**(13)**:** 3088-91

Yalcin A, Kreutz C, Pfeifer D, Abdelkarim M, Klaus G, Timmer J, Lubbert M, Hackanson B (2013) MeDIP coupled with a promoter tiling array as a platform to investigate global DNA methylation patterns in AML cells. *Leukemia research* **37**(1)**:** 102-11

Yamada Y, Watanabe H, Miura F, Soejima H, Uchiyama M, Iwasaka T, Mukai T, Sakaki Y, Ito T (2004) A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. *Genome Research* **14**(2)**:** 247-66

Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW, Liang WX, Mi JQ, Song HD, Li KQ, Chen Z, Chen SJ (2011) Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet* **43**(4)**:** 309-15

Yanai F, Ishii E, Kojima K, Hasegawa A, Azuma T, Hirose S, Suga N, Mitsudome A, Zaitsu M, Ishida Y, Shirakata Y, Sayama K, Hashimoto K, Yasukawa M (2003) Essential Roles of Perforin in Antigen-Specific Cytotoxicity Mediated by Human CD4+ T Lymphocytes: Analysis Using the Combination of Hereditary Perforin-Deficient Effector Cells and Fas-Deficient Target Cells. *The Journal of Immunology* **170**(4)**:** 2205-2213

Young MD, Willson TA, Wakefield MJ, Trounson E, Hilton DJ, Blewitt ME, Oshlack A, Majewski IJ (2011) ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Res*

Zhang D-E, Zhang P, Wang N-d, Hetherington CJ, Darlington GJ, Tenen DG (1997) Absence of granulocyte colony-stimulating factor signaling and

neutrophil development in CCAAT enhancer binding protein α-deficient mice. *Proceedings of the National Academy of Sciences* **94**(2)**:** 569-574

Zhang P, Iwasaki-Arai J, Iwasaki H, Fenyus ML, Dayaram T, Owens BM, Shigematsu H, Levantini E, Huettner CS, Lekstrom-Himes JA, Akashi K, Tenen DG (2004) Enhancement of Hematopoietic Stem Cell Repopulating Capacity and Self-Renewal in the Absence of the Transcription Factor C/EBPα. *Immunity* **21**(6)**:** 853-863

Zhu J, He F, Hu S, Yu J (2008) On the nature of human housekeeping genes. *Trends Genet* **24**(10)**:** 481-4

Ziller MJ, Müller F, Liao J, Zhang Y, Gu H, Bock C, Boyle P, Epstein CB, Bernstein BE, Lengauer T, Gnirke A, Meissner A (2011) Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types. *PLoS Genet* **7**(12)**:** e1002389

APPENDIX

Appendix Table 1 Primer sequences, annealing temperatures and WAVE analysis temperatures if appropriate

PCR	Primer	Sequence	Fragment Length (bp)	Annealing temperature for PCR (°C)	Temperatures for dHPLC (°C)
ASXL1ex12 A	F	5'-CCTAGGTCAGATCACCCAGTC-3'	563	63	59 60.7 64.2
	R	5'-AGCCCATCTGTGAGTCCAACTGT-3'			
ASXL1ex12 A2	F	5'-GGATCATCCCCACCACGGAGT-3'	232	63	65.7
	R	5'-CCACAGGCCTCACCACCATC-3'			
ASXL1ex12 B	F	5'-AGCTAGGAGAGAGGACCTGC-3'	565	62	59.3 60.2 61.5
	R	5'-CGATGGGATGGGTATCCAATG-3'			
ASXL1ex12 C	F	5'-CTCCTATGAGGGAAAGTGATACT-3'	518	62	57.9 58.9 60.6
	R	5'-CTTGGACAGTGGGGCAGATTG-3'			
ASXL1ex12 D	F	5'-ACAGCAGTGAGGCTGACACTAG-3'	544	63	59.5 60.3 61.5
	R	5'-GCCTCAATCCTGGCAAGACCAG-3'			
ASXL1ex12 E	F	5'-GAGCCATGGCTCGCTACGCAT-3'	433	64	58.5 60.1
	R	5'-CTGCTCTGGACCAAAGGAGATC-3'			
ASXL1ex12 E(MM)	F	5'-GAAGGAAGTCCGTGCTATGTCACAGGCCAG-3'	126	63	
	R	5'-CTGCTCTGGACCAAAGGAGATC-3'			
ASXL1ex12 F	F	5'-CTTACTACCTCGAGAACACCTCGT-3'	489	64	58.5 60.8 62.2
	R	5'-AAGAAGGCTCCAGAGGCTCAC-3'			

Appendix Table 1 Continued

PCR	Primer	Sequence	Fragment Length (bp)	Annealing temperature for PCR (°C)	Temperatures for dHPLC (°C)
ASXL1ex12 G	F	5'-GATGCCCTTTGTCATGGACTTGC-3'	323	64	59.2 60.3 63.1
	R	5'-TGCACGCACACTGGAGCGAGA-3'			
ASXL1ex12 H	F	5'-GTGCGAGCCACAGTGCATCAC-3'	219	64	55.6 59.6 62.5
	R	5'-CTAAATATACAATGTTTCCCATGGCCATA-3'			
CEBPA/1	F	5'-TCGCCATGCCGGGAGAACTCTAAC-3'	548	62	
	R	5'-AGCTGCTTGGCTTCATCCTCCT-3'			
CEBPA/2	F	5'-GCTGGTGATCAAGCAGGAGC-3'	332	62	
	R	5'-CCGCCACTCGCGCGGAGGTCG-3'			
CEBPA/3	F	5'-GGCAGCGCGCTCAAGGGGCTG-3'	424	64	
	R	5'-CACGGCTCGGGCAAGCCTCGAGAT-3'			
CEBPA/Core	F	5'-GGTTTGTAGGGTATAAAAGTTGGG-3'	168	60	
	R	5'-ACTCCATAAAAAAATTAAAATTCTCCC-3'			
CEBPA/DistFor	F	5'-GGGTTATTAATTATTGGGATTATGTTGAA-3'	314	58	
	R	5'-AAAAACCCTCAAATATCTCCTAT-3'			
CEBPA/DistRev	F	5'-GAGTTTTGGGAGTTTTTAAGTGTT-3'	262	58	
	R	5'-ACCCAAATAAAACTACTTCTTTAC-3'			
HLA-B (unconverted)	F	5'-TTTTAAGTTTTATTTTTGTGGGGTA-3'	300	Touchdown	
	R	5'-AAATCCCAACTAATAACTATTTTTCAA-3'			
HLA-B (converted)	F	5'-CCCAAAGTCCACTAACATTAGAA-3'	464	Touchdown	
	R	5'-GCTGAGAAAATAGCCTCAGAATA-3'			

Appendix Table 1 Continued

PCR	Primer	Sequence	Fragment Length (bp)	Annealing temperature for PCR (°C)	Temperatures for dHPLC (°C)
KIAA0323	F	5'-/5Biosg/GGGTTTTTTAGTTGTAGTTAGATGTG-3'	192	60	
	R	5'-ACTAAAAACAACAACCATACCTAC-3'			
	PyroSeqR	5'-ACCCCATATAAAACCCATCTTC-3'			
LY9	F	5'-/5Biosg/TGTTTTAGAGGGAGGGTTGTTTATA-3'	100	58	
	R	5'-AATCACAAATAAAACCCTAAATAAACTTA-3'			
	PyroSeqR	5'-TAAAACCTCTACCTACC-3'			
PRF1A	F	5'-AGTAGGGTTATTTTTTTGTTTTTGATGT-3'	150	60	
	R	5'-/5Biosg/CCTACCAATCCACACTACTAATACA-3'			
	PyroSeqF	5'-GTTATTTTTTTGTTTTTGATGTATA-3'			
PRF1B	F	5'-TAGGAAGTGTTGTGATTTATAAGATAAG-3'	163	60	
	R	5'-/5Biosg/CTTTAATATCAACACTTACAAAACCTTAA-3'			
	PyroSeqF	5'-TAAGATAAGATATTTGGGTTA-3'			
PRIC285	F	5'-/5Biosg/GTTTTTTTTTGTGTTATTAGTAGGAGAT-3'	313	58	
	R	5'-CCCCAACAACTAAACCAAATATT -3'			
	PyroSeqR	5'-AAAATACCCCCAAATAAAACTAACAA-3'			
SOC2	F	5'-/5Biosg/AGGTGGGAAGTAAAGAATAAGATGGA-3'	128	62	
	R	5'-CCAAACCTAAATCCCTAAAAAACCACTTT-3'			
	PyroSeqR	5'-CCTAAAAAACCACTTTCCT-3'			

Appendix Table 1 Continued

PCR	Primer	Sequence	Fragment Length (bp)	Annealing temperature for PCR (°C)	Temperatures for dHPLC (°C)
VAMP5	F	5'-GTGTTYGTTTATTAGGTAGAGGTGTTA-3'	281	59	
	R	5'-/5Biosg/CCCRCCTAAACCCTCACCATC-3'			
	PyroSeqF	5'- GTTTATYGTTTTYGATTTGATTTGG-3'			
WNT2	F	5'-GTGTATGAAATGATGGTAAGAGATGTT-3'	246	62	
	R	5'-/5Biosg/ATACATAATAATCTCCTTATCCCCTAACC-3'			
	PyroSeqF	5'-GGGAAGGGGGAATATYGTTGTATG-3'			

Abbreviations: bp, base pairs; dHPLC, denaturing HPLC; F, forward primer; PCR, polymerase chain reaction; PyroseqF or R, sequencing primer for pyrosequencing; R, reverse primer.

Appendix Table 2 List of samples with *CEBPA* mutations in all cohorts and classification of samples

Cohort	Patient No.	Mutation 1	Type of mutation	Location[1]	Mutation 2	Type of mutation	Location[1]	Classification
1	1	A66fs	FS	N	E316_L317insR	IF ins	C	Classic DM
1	2	G114fs	FS	N	Q312dup	IF ins	C	Classic DM
1	3	L78fs	FS	N	E309_V328dup	IF ins	C	Classic DM
1	4	K313dup	IF ins	C	R343fs	FS	C	Non-classic DM
1	5	A91fs	FS	N	S319delinsRL	IF indel	C	Classic DM
1	6	V314G (HOM)	IF ins	C				Homozygous non-classic C
1	7	I68fs	FS	N	R306_N307insRR	IF ins	C	Classic DM
1	8	H84fs	FS	N	K302_K304dup	IF ins	C	Classic DM
1	9	G54fs	FS	N	Q305_R306insQQ	IF ins	C	Classic DM
1	10	H24fs	FS	N	Q311_Q312insL	IF ins	C	Classic DM
2	11	S61fs	FS	N				Classic N
2	12	G54fs	FS	N	E167fs	FS	mid	Non-classic DM
2	13	I341V	M	C				Non-classic C
2	14	K313dup	IF ins	C				Classic C
2	15	Q215X	STOP	MID				Non-classic SM
2	16	K313dup	IF ins	C				Classic C
2	17	H24fs	FS	N	K304_Q305insL	IF ins	C	Classic DM
2	18	N321S (HOM)	M	C				Homozygous non-classic C
2	19	L253fs	FS	MID				Non-classic SM
2	20	H24fs	FS	N	Q209fs	FS	MID	Non-classic DM
2	21	A40fs	FS	N	E309_L317dup	IF ins	C	Classic DM
2	22	N356_C357del	IF del	C				Classic C
2	23	Y67X	STOP	N	R297_V308dup	IF ins	C	Classic DM
2	24	Q88fs	FS	N	K304_Q305insL	IF ins	C	Classic DM
2	25	P23fs	FS	N	N307_E309delinsK	IF indel	C	Classic DM
2	26	H24fs	FS	N	K313Nins14	IF ins	C	Classic DM
2	27	P112fs	FS	N	K304_Q305insL	IF ins	C	Classic DM

Appendix Table 2 continued

Cohort	Patient No.	Mutation 1	Type of mutation	Location[1]	Mutation 2	Type of mutation	Location[1]	Classification
2	28	A72fs	FS	N	K302_E309dup	IF ins	C	Classic DM
2	29	G96fs	FS	N	T310_Q311insHKA KQRNVET	IF ins	C	Classic DM
2	30	E59fs	FS	N	E309dup	IF ins	C	Classic DM
2	31	A113fs	FS	N	L315_E316insQ	IF ins	C	Classic DM
2	32	A79fs	FS	N	Q312dup	IF ins	C	Classic DM
2	33	I55fs	FS	N	K313Sins12	IF ins	C	Classic DM
2	34	H24fs	FS	N				Classic N
3	35	N292fs	FS	C				Non-classic C
3	36	P43fs	FS	N	Q312dup	IF ins	C	Classic DM
3	37	K275fs	FS	MID				Non-classic SM
3	38	L331Q	M	C				Non-classic C
3	39	G242S	M	MID				Non-classic SM
3	40	E50fs	FS	N	V314_L315ins13	IF ins	C	Classic DM
3	41	P180fs	FS	MID				Non-classic SM
3	42	R300_D301delinsQN [HOM]	IF indel	C				Homozygous classic C
3	43	A66fs	FS	N				Classic N
3	44	G123fs	FS	MID				Non-classic SM
3	45	G122fs	FS	MID				Non-classic SM
3	46	Y285S	M	C				Non-classic C
3	47	A44fs	FS	N	A295P	M	C	Non-classic DM
3	48	E166X	STOP	MID				Non-classic SM
3	49	G36fs	FS	N	Y181X	STOP	MID	Non-classic DM
3	50	V296E	M	C				Non-classic C
3	51	N307fs	FS	C				Non-classic C
3	52	H193_P196del	IF del	MID				Non-classic SM
3	53	G38fs	FS	N	K313fs	FS	C	Non-classic DM
3	54	I68fs	FS	N				Classic N
3	55	P233R	M	MID				Non-classic SM
3	56	E59X	STOP	N				Classic N

Appendix Table 2 continued

Cohort	Patient No.	Mutation 1	Type of mutation	Location[1]	Mutation 2	Type of mutation	Location[1]	Classification
3	57	H24fs,P23fs	FS	N				Classic N
3	58	P46fs	FS	N				Classic N
3	59	Q83fs	FS	N				Classic N
3	60	R300_D301delinsQY	IF indel	C				Classic C
3	61	R306_V314dup	IF ins	C	N321S	M	C	Non-classic DM
3	62	P183Q	M	MID				Non-classic SM
3	63	I68fs	FS	N	R297P	M	C	Non-classic DM
3	64	D107fs	FS	N	S299_T318dup	IF ins	C	Classic DM
3	65	T60fs	FS	N				Classic N
3	66	G242S	M	MID				Non-classic SM
3	67	P239_A240del	IF del	nonMID				Non-classic SM
3	68	R323del	IF del	C				Classic C
3	69	H24fs	FS	N	G233fs	FS	MID	Non-classic DM
3	70	D168fs	FS	MID				Non-classic SM
3	71	P187_P189del	IF del	MID				Non-classic SM
3	72	S266fs	FS	MID				Non-classic SM
3	73	V95fs	FS	N	A238fs	FS	MID	Non-classic DM
3	74	K276R	M	MID				Non-classic SM
3	75	H24fs	FS	N	Q311_Q312insL	IF ins	C	Classic DM
3	76	L78fs	FS	N	R300P	M	C	Non-classic DM
3	77	R289C	M	C				Non-classic SM
3	78	E316_R325dup [HOM]	IF dup	C				Homozygous classic C
3	79	L317Q [HOM]	M	C				Homozygous non-classic C

Abbreviations: FS, frameshift mutation; IF del, inframe deletion; IF indel, inframe insertion/deletion; IF ins, inframe insertion; M, missense mutation.

[1]Location: N, N terminus (amino acids 1-119); MID, from second ATG to DBD (amino acids 120-277); C, C terminus (amino acids 278-358).