High-Resolution Genome-Wide Single Nucleotide Polymorphism Mapping in Acute Myeloid Leukaemia

Raghavan, Manoj

High-Resolution Genome-Wide Single Nucleotide Polymorphism Mapping in Acute Myeloid Leukaemia

A thesis submitted for the degree of Doctor of Philosophy

Manoj Raghavan

Barts and the London School of Medicine and Dentistry,

Queen Mary, University of London

August 2008

Abstract

Investigation of the genetics of acute myeloid leukaemia (AML) has revealed the underlying basis of the disease and led to targets for therapy. High-resolution single nucleotide polymorphism (SNP) arrays detected regions of loss of heterozygosity and DNA copy number changes, augmenting the results of conventional cytogenetic analysis in AML. Fifteen out of 72 (20%) primary AML samples exhibited large regions of homozygosity that could not be accounted for by visible chromosomal abnormalities in the karyotype. Further analysis confirmed that these patterns were due to partial uniparental disomy (UPD). Remission marrow was available from five patients showing UPD in their leukemias, and in all cases the homozygosity was found to be restricted to the diagnostic leukemic clone. These cryptic non-random chromosomal abnormalities are characteristic of mitotic recombination. In 7 of 13 cases with UPD, concurrent homozygous mutations were identified at four distinct loci (*WT1*, *FLT3*, *CEBPA*, and *RUNX1*). This implies that mutation precedes mitotic recombination which acts as a "second hit" responsible for removal of the remaining wild-type allele. Clonal evolution from heterozygous to homozygous mutations by mitotic recombination would provide a mechanism for relapse of AML. Analysis of 27 paired diagnostic and relapsed AML samples demonstrated newly acquired segmental UPDs at relapse in 11 AML samples (40%). Six were segmental UPDs of chromosome 13q, which led to a change from heterozygosity to homozygosity for internal tandem duplication of *FLT3*. Three further AML samples had evidence of acquired segmental UPD of 13q in a subclone of the relapsed leukemia. One patient acquired segmental UPD of 19q which led to homozygosity for a *CEBPA* mutation 207 C→T. Finally, a single AML patient acquired segmental UPD of chromosome 4q, for which the candidate gene is unknown. In conclusion, the acquisition of segmental UPD and the resulting homozygous mutation is a common event associated with relapse of AML.

Dedication

To LiLing without whom, none of this would be possible, and my family for their support over the many years.

Acknowledgements

I am greatly indebted to Prof. Bryan Young for his enthusiasm and guidance throughout my research and to Prof Andrew Lister for his wise counsel and encouragement. I also thank Rosemary Gale and David Linch for allowing me access to DNA samples from the MRC AML 10 trial.

A great deal of thanks goes to the medical and nursing of the Bodley Scott Unit who obtained leukaemia samples for the tissue bank contributing to this work. All the clinical data is based on their work, and for that I am very grateful.

I would like to thank the staff of Medical Oncology who helped me learn so much for all their help in particular to Jude Fitzgibbon for innumerable informal conversations in the corridor. Finally, I thank the many patients without whom research into leukaemia would not be possible.

This work is funded from a project grant from Cancer Research UK.

Individual Contributions

All work in this thesis was performed by myself, Manoj Raghavan. Additional contributions were made from the following individuals, and a summary of my own and their contributions are outlined below.

Ethical committee submissions – Manoj Raghavan, Bryan Young

Information retrieval from clinical database and patient notes – Finlay Macdougall, Nigel Bainton, Manoj Raghavan

Morphological diagnoses – John Amess, Michael Jenner

Maintenance of CRUK medical oncology tissue bank – Keith Adams, Lindsey Goff, Sameena Iqbal

Cell cryopreservation – Sameena Iqbal

Provision of MRC AML 10 clinical trial DNA samples – Rosemary Gale, David Linch

DNA extraction – Manoj Raghavan, Gael Molloy, Jude Fitzgibbon

SNP array genotyping – Manoj Raghavan, Gael Molloy, Tracy Chaplin, Bryan Young

Cytogenetic analysis and FISH – Debra Lillington, Nicola Foot

Bisulfite sequencing – Silvana Debernardi

Mutation analysis – Manoj Raghavan, Lan-Lan Smith, Matt Smith, Ioannis Kakkas

Pyrosequencing - in conjunction with the Genome Centre, Barts and The London

Bioinformatics and data interpretation – Manoj Raghavan, Spyros Skoulakis, Jean-Baptiste Cazier, Bryan Young

GOLF software programming – Bryan Young

Published papers arising from this thesis

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

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

Gupta, M., Raghavan, M., Gale, R.E., Chelala, C., Allen, C., Molloy, G., Chaplin, T., Linch, D.C., Cazier, J.-B. & Young, B.D. (2008) Novel regions of acquired uniparental disomy discovered in acute myeloid leukemia. *Genes, Chromosomes and Cancer,* **47,** 729-818.

Raghavan, M., Smith, L.-L., Lillington, D.M., Chaplin, T., Kakkas, I., Molloy, G., Chelala, C., Cazier, J.-B., Cavenagh, J.D., Fitzgibbon, J., Lister, T.A. & Young, B.D. (2008) Segmental uniparental disomy is a commonly acquired genetic event in relapsed acute myeloid leukemia. *Blood,* **112,** 814-821.

Table of Contents

List of Tables and Figures

Chapter 1

Introduction

Acute myeloid leukaemia (AML) is a malignant proliferation of poorly differentiated myeloid cells, or blast cells. It is characterised by the accumulation of these cells in bone marrow and usually blood, often leading to cytopenias in other myeloid lineages. Blasts can also move into extracellular tissues causing splenomegaly, hepatomegaly and sometimes lymphadenopathy. The leukaemic cells have been described in many other tissues as chloromas or granulocytic sarcomas and in the central nervous system. Untreated the disease is uniformly fatal. In England and Wales, the annual incidence of myeloid leukaemia for all ages for 2004 was 6.0 and 4.7 per 100,000 population for males and females respectively (Office for National Statistics 2005, Office for National Statistics 2006). This incidence has been increasing for males having been only 5.5 per 100000 in 1992, most probably because the elderly population is expanding. For males aged 80-84 years the incidence has increased from 36.4 per 100000 in 1992 to 40.7 per 100000 in 2004, although the age standardised rate has remained unchanged over the last 10 years. AML is becoming an increasingly important disease of the elderly.

The initial descriptions of leukaemia in 1845 by Craigie, Bennett and Virchow were of suppuration or pus in the blood (Bennett 1845, Craigie 1845, Piller 2001). Bennett particularly noted that the alteration of the blood had occurred in the absence of inflammation and was able to give descriptions and drawings of white cells from light microscopy. By 1849, Virchow had described three cases and distinguished them into splenic and lymphatic forms due to the presence or absence of splenomegaly. Acute leukaemia was first used as a term in 1857 by

Nikolaus Friedrich, because of the short period between presentation and death. In 1868, Neumann described the origin of leukaemia as from the bone marrow and later added myelogenous leukaemia to the classification. It was the discovery of aniline dyes (by William Henry Perkin in 1856) that led Paul Ehrlich to develop stains that identified myeloid and lymphoid cells from their cytoplasmic and nuclear structures. He maintained that leukaemia was a disease of the haematopoietic system. Further evidence for this came from the identification of myeloblasts and lymphoblasts as immature cells in the granulocytic and lymphocytic series by Naegli in 1900.

During the twentieth century, progress in the diagnosis and treatment of AML accelerated due to improved cytological classification with cytochemistry, and the advent of chemotherapy inducing meaningful remissions. Subsequently, cytogenetics and molecular biology have greatly improved our understanding of the disease and refined the diagnostic classifications.

1.1 Aetiology

AML may be divided into *de novo* and secondary types. The causes of *de novo* AML are poorly understood. Many epidemiological studies have looked at potential agents predisposing to *de novo* AML. Because of the difficulties of confounding variables, subjective answers and the difficulties in finding corresponding control groups, the relative risk of leukaemia identified by these case control studies are slight. For example, there are studies into potentially carcinogenic agents such as hair dyes where there is a trend towards an association or a borderline significant association with AML (odds ratio 1.6, 95% confidence interval (CI) 1.1-2.4) (Mele, *et al* 1994, Rauscher, *et al* 2004). Another case control study found cigarette smoking associated with AML M2 (odds ratio 2.3, 95% CI 1.1-4.4) (Pogoda, *et al* 2002). As for all epidemiological case control studies, the real significance of these results will only be known if they can be replicated in other populations.

For secondary AML clear aetiological agents have been described. As demonstrated

after the Hiroshima and Nagasaki bombs (Folley, *et al* 1952), irradiation can lead to AML and myelodysplastic syndrome (MDS). The risk of AML was correlated inverse square of the distance from the explosion, with the risk rising rapidly with doses of radiation greater than 0.5 Gy. The incidence of AML was maximal 7 years after the explosions, but an increased incidence of AML continued up to 40 years later (Schull 1998). Latterly, there is also evidence for an increased incidence of AML in those exposed after the Chernobyl disaster (Gluzman, *et al* 2006). Radiotherapy treatment predisposes to AML/MDS (Michels, *et al* 1985), although most cancer patients are also exposed to cytotoxic chemotherapy that is mutagenic. Patients with polycythaemia vera (PV) treated with phosphorus-32 have an increased risk of AML, complicated by PV also increasing the risk of developing leukaemia (Parmentier 2003). A clearer indication of the risks of radiotherapy was given by a disease not associated with leukaemia, ankylosing spondylitis. It was treated with irradiation in the 1930s to the 1950s, but this population was found to have an increased incidence of leukaemia, with a dose-response relationship (Court-Brown and Doll 1957).

Chemical exposure can also lead to AML, of which the most important organic compound is benzene (Hayes, *et al* 1997). Studies of Chinese industrial workers show a dose dependent increase in levels of MDS and AML in those exposed to benzene, with a relative risk up to 7 times, in those exposed to 25 parts per million of benzene (Hayes, *et al* 1997). Two groups of cytotoxics are particularly associated with AML. Alkylating agents are associated AML/MDS with poor risk cytogenetics, such as deletion of chromosome 5, 5q, 7 or 7q, as is observed after high dose therapy (HDT) with cyclophosphamide (Micallef, *et al* 2000). In this study, after a median follow up of 6 years, 14% of 230 patients developed AML/MDS. However, retrospectively, many of these patients had the same cytogenetic abnormalities in a few cells prior to HDT, suggesting that chemotherapy given up to that point (which frequently included chlorambucil) had contributed to the development of AML. The other major group of cytotoxics that cause AML are topoisomerase II inhibitors, such as etoposide. These tend to be associated with balanced translocations, e.g. those involving the *MLL* gene at chromosome 11q23 (Super, *et al* 1993). In patients with acute promyelocytic leukaemia (APML) that is

secondary to topoisomerase II inhibitors, the t(15;17) breakpoint between *PML* and *RARA* occurs at sequence hotspots that are targeted by etoposide (Mistry, *et al* 2005).

AML can occur secondary to azathioprine use, as noted after solid organ transplants (Offman, *et al* 2004). DNA damage caused by azathioprine is due to mismatch repair, where an incorrect nucleotide is inserted during DNA replication. The AML cells of these patients are defective in mismatch repair as demonstrated by microsatellite instability (Offman, *et al* 2004).

AML is also associated with a number of other haematological diseases. MDS has a close association, with high-risk subtypes frequently transforming into AML. Many AML patients have trilineage dysplasia, and the distinction between AML and MDS is arbitrarily set with more than 20% myeloblasts in the bone marrow defining AML. Previously, the French American British classification (FAB) defined blasts of 20-30% as a subtype of MDS, known as refractory anaemia with excess blasts in transformation (RAEB-T) (Bennett, *et al* 1982). Factors influencing transformation to AML are part of the international prognostic scoring system (IPSS) (Greenberg, *et al* 1997). These include cytogenetic abnormalities, percentage of bone marrow myeloblasts and number of lineages with cytopenias.

Myeloproliferative diseases (MPDs), particularly chronic myeloid leukaemia (CML), can also transform into AML. CML transforms from a chronic phase into a blast crisis that resembles AML or acute lymphoblastic leukaemia (ALL), with some passing through an accelerated phase. CML is characterised by the Philadelphia chromosome, t(9;22) (Nowell and Hungerford 1960), that encodes the fusion protein BCR/ABL, the underlying genetic defect. Myelosuppressive treatment, e.g. with hydroxycarbamide, controls the high white blood cell and platelet count, but does not affect the number of Philadelphia positive cells. With hydroxycarbamide, 70% of patients transform to accelerated phase or blast crisis over 6 years (Hehlmann, *et al* 1993). Use of imatinib, a tyrosine kinase inhibitor targeting BCR/ABL, has reduced progression to accelerated phase or blast crisis to just 10% at 6 years (Druker, *et al* 2006). This has become the paradigm for targeted therapy in cancer, and demonstrates that targeting the underlying genetic defect may treat or prevent leukaemia.

Essential thrombocythaemia, and polycythaemia vera are frequently treated with

cytotoxic agents such as hydroxycarbamide, to reduce the platelet count or the packed cell volume. It has been difficult to disentangle whether the disease or the treatment predisposes to AML, although some of the treatments used historically, such as phosphorus-32 or busulphan, are leukaemogenic (Harrison 2005). About 20% of patients with myelofibrosis also transform into AML over 10 years, with a higher risk for those with a haemoglobin concentration of less than 10g/dl, constitutional symptoms or circulating blasts (Cervantes, *et al* 1998).

1.2 Genetic Predisposition

A small number of patients with AML are known to be associated with germline genetic abnormalities (Owen, *et al* 2008). Individuals with trisomy 21 or Down syndrome have an increased risk of several myeloid disorders, including transient abnormal myelopoiesis (TAM), MDS and AML, especially acute megakaryocytic leukaemia (AMKL) (Vyas and Roberts 2006). Patients with TAM and AMKL have mutations in *GATA-1*, a transcription factor involved in erythropoiesis and megkaryopoiesis (Ahmed, *et al* 2004).

Patients with neurofibromatosis (NF) and Noonan syndrome (NS) are predisposed to juvenile myelomonocytic leukaemia (JMML), AML and other malignancies (Lauchle, *et al* 2006). They have mutations in *NF1* and *PTPN11* respectively, both of which cause deregulation of the RAS signalling pathway. The oncogene, *RAS*, encodes a GTPase that is downstream of several cytokine growth factors and induces proliferation. In JMML, there is increased RAS pathway activity due to heterozygous mutations in *KRAS* or *PTPN11*, or homozygous mutations in *NF1* (Flotho, *et al* 2007).

All patients with congenital bone marrow failure syndromes have an increased risk of AML e.g. Diamond-Blackfan Anaemia (Freedman 2000), and congenital neutropenia syndromes such as Schwachman-Diamond syndrome (Rosenberg, *et al* 2006). Patients with severe congenital neutropenia have a 13% risk of developing AML or MDS while on treatment with recombinant granulocyte colony stimulating factor (GCSF) (Dale, *et al* 2003). It is unclear whether it is the disease or the treatment that predisposes to leukaemia, as the patients did not

live long enough before the advent of treatment with GCSF, but interestingly those with cyclical and idiopathic neutropenia did not get leukaemia.

Other inherited syndromes with defects in DNA repair such as Fanconi's anaemia (FA) or Bloom's syndrome are also associated with many cancers including AML. Patients with FA have chromosomal breaks due to mutations in genes in the FA pathway involved in homologous double-strand DNA repair. They have an incidence of AML and MDS between 10 to 15% (Mathew 2006). Bloom syndrome involves mutations of BLM helicase involved in the repair of double strand DNA breaks by homologous recombination. A registry of Bloom syndrome patients gives an incidence of AML of about 20% (German 1997).

There are also reports of familial AML or MDS, some associated with deletions of chromosome 5q or chromosome 7/7q, but no causative gene has been found in these cases (Gao, *et al* 2000, Grimwade, *et al* 1993, Hess 2001). Inherited single gene mutations have been described in familial AML involving *CEBPA* (Smith, *et al* 2004b). This is an inherited heterozygous mutation, where the affected family members have acquired a second somatic mutation at the time of developing AML in childhood or early adulthood. Familial platelet disorder with AML is an autosomal dominant inherited disorder that involves a heterozygous mutation and haploinsufficiency of *RUNX1* (Song, *et al* 1999). The patients have a low platelet count with a bleeding tendency, and may develop AML in the first to the 6th decade.

There is evidence from case control studies for inherited polymorphisms predisposing to AML, particularly in genes encoding metabolic enzymes and DNA repair proteins. The numbers of patients studied are relatively small but significant effects have been observed for predisposition to therapy related AML from DNA repair genes involved in the repair of double strand breaks, e.g. polymorphisms in the 5' untranslated region of *HLX1* and *RAD51* increased the relative risk of therapy related AML by 9.5 times (95% CI, 2.22-40.64) (Jawad, *et al* 2006). The same polymorphism of *RAD51* associated with a non-synonymous polymorphism of *XRCC3* also gave an increased risk (odds ratio, 8.11; 95% CI, 2.22–29.68) (Seedhouse, *et al* 2004). *XRCC1* is involved in base excision repair, and a non-synonymous polymorphism is associated with a reduced risk of developing therapy related AML (odds ratio 0.44; 95%

confidence interval, 0.20-0.93) (Seedhouse, *et al* 2002). As with all genetic association studies, replication in other populations will show how robust these associations really are (Kuptsova, *et al* 2007).

1.3 Diagnosis and Classification

AML typically presents with bone marrow failure and high white blood cell count due to the accumulation of myeloblasts. The symptoms are very variable but include fatigue and lethargy due to anaemia, bleeding due to thrombocytopenia and severe bacterial infection due to neutropenia. Tissue infiltration may result in organomegaly, particularly with monocytic and monoblastic leukaemias. The diagnosis of acute leukaemia is made on light microscopy of the blood smear and further classification of AML is based on morphology, immunophenotype and cytogenetics. As for all diseases, an accurate diagnosis is required to ensure the correct treatment is given in the light of the patient's likely prognosis. For AML, the main differential diagnosis is ALL, and subsequent subclassification of AML identifies the likely prognosis and possible lines of treatment.

1.4 Morphology

Morphological analysis is carried after staining the cells with a Romanowsky type stain and further elucidated using cytochemical techniques to look for myeloperoxidase (MPO) positive granules and non-specific esterases (NSE) to identify vesicles found in monocytic lineage cells. AML is characterised by the differentiation of the blast cells into various subtypes by the FAB classification (Table 1.1) (Bennett, *et al* 1976, Bennett, *et al* 1985). They are broadly divided into myeloid, monocytoid, erythroid and megakaryocytoid forms. The subtypes are generally treated the same, except for AML M3 or APML which is associated with t(15;17), producing the PML-RARA fusion protein (Goddard, *et al* 1991). This leukaemia is associated with hypofibrinogenemia leading to bleeding, which previously gave the condition a high mortality.

However, the promyelocytes readily differentiate with all-trans retinoic acid (ATRA), which is used as part of the treatment for APML. Several other morphological variants are associated with cytogenetic abnormalities, for example t(8;21) is associated with AML M2, and inv(16) with AML M4 with eosinophils (M4Eo). Other FAB types also have associations e.g. AML M6 often occurs in the context of MDS (Bain 1999). The importance of other factors than morphology in the classification of AML, particularly molecular cytogenetics and types of secondary AML, led to the development of the World Health Organisation (WHO) classification (Table 1.1) (Brunning, *et al* 2001).

Antibodies to the myeloid cell markers cytoplasmic MPO, CD33, CD13 and CD117 are used to confirm the diagnosis (Bain, *et al* 2002). By using flow cytometry, several hundred or thousand cells can be analysed in a short space of time. Some FAB types require further immunophenotyping to make the diagnosis i.e. M6 and M7. Acute erythroblastic leukaemia (M6) expresses glycophorin A, an erythroid cell marker but has a characteristic morphology because of the presence of megaloblastic looking blast cells. Acute megakaryocytic leukaemia (M7) expresses CD41, CD42 and CD61, although the morphology is often undifferentiated. AML M0 has by definition an undifferentiated morphology, so identification of the myeloid antigens and the absence of lymphoid antigens is required for diagnosis. These three FAB types all have a relatively poor prognosis in adults (Colita, *et al* 2001, Oki, *et al* 2006, Roumier, *et al* 2003).

Table 1.1 Classifications of AML.

FAB Classification:	**WHO Classification:**
M0: Undifferentiated	I. *AML with recurrent cytogenetic translocations*
M1: AML without maturation	AML with t(8;21)(q22;q22); *RUNX1(AML1 or CBFA2)/ RUNX1T1(ETO or MTG8)*
M2: AML with maturation	Acute promyelocytic leukaemia [AML with t(15;17)(q22;q12) and variants; PML/RAR]
M3: Acute promyelocytic leukaemia	AML with abnormal bone marrow eosinophils [inv(16)(p13q22) or t(16;16)(p13;q22) *CBFB/MYH11*]
M4: Myelomonocytic leukaemia	AML with 11q23 (*MLL*) abnormalities
M5: Monocytic leukaemia	II. *AML with multilineage dysplasia*
M6: Erythroleukaemia	With prior myelodysplastic syndrome
M7: Megakaryoblastic leukaemia	Without prior myelodysplastic syndrome
	III. *AML and myelodysplastic syndrome, therapy-related*
	Alkylating agent-related
	Epipodophyllotoxin-related
	Other types
	IV. *AML not otherwise categorized*
	AML minimally differentiated
	AML without maturation
	AML with maturation
	Acute myelomonocytic leukaemia
	Acute monocytic leukaemia
	Acute erythroid leukaemia
	Acute megakaryocytic leukaemia
	Acute basophilic leukaemia
	Acute panmyelosis with myelofibrosis

1.5 Cytogenetics

Not only are recurrent cytogenetic abnormalities associated with some morphological subtypes, but they are also the most important biological prognostic markers in AML. APML with t(15;17) and those with core binding factor (CBF) translocations i.e. t(8;21) and inv(16)(q13q22), have the best prognosis, with a 50-60% 5 year survival (Grimwade, *et al* 1998) (Table 1.2). Other cytogenetic abnormalities correspond to poor prognosis i.e. deletions of chromosome 5, 5q or 7, abnormalities of 3q, or a complex karyotype (more than 4 abnormalities) with about 10-15% 5 year survival. However, these make up only about 30% of AML. Most other cytogenetic abnormalities have an intermediate prognosis, including about 40% of patients with AML that have no karyotypic changes at all. Never the less, these prognostic abnormalities allow a refinement in the classification. The WHO classification uses a combination of cytogenetic abnormalities and the FAB classification (Table 1.1) in an attempt to create a more clinically relevant classification. Several large cohort clinical randomised trials have used cytogenetic profiles to risk stratify patients (Byrd, *et al* 2002, Grimwade, *et al* 1998, Slovak, *et al* 2000) (Table 1.2). There are minor variations in the definitions of good, intermediate and poor risk, with some studies including rare abnormalities that have a poor prognosis e.g. t(6;9), and others suggesting that some translocations involving *MLL* at 11q23 have a poor prognosis. These trials also disagree as to the prognosis of some cytogenetic abnormalities, e.g. del(5q) is defined as poor risk by the MRC and SWOG/ECOG, but is of intermediate risk according to CALGB (Table 1.2).

Table 1.2 Cytogenetic risk group definitions.

	MRC (Grimwade, *et al* 1998)	SWOG/ECOG (Slovak, *et al* 2000)	CALGB (Byrd, *et al* 2002)
Good	t(15;17), t(8;21), inv(16) or t(16;16)	inv(16)/t(16;16)/del(16 q), t(15;17) with/without secondary aberrations; t(8;21) lacking del(9q) or complex karyotypes	t(8;21), inv(16) or t(16;16)*
Standard / Intermediate	Neither good nor poor	Normal, +8, +6, -Y, del(12p)	Del(9q), normal karyotype, -Y, del(5q), del(7q), t(9;11), +11, del(11q), abn(12p), +13, del(20q), +21
Poor	-5 or del(5q), -7, abnormalities of 3q, complex karyotype (>4 abnormalities)	del(5q)/-5, -7/del(7q), abn 3q, 9q, 11q, 20q, 21q, 17p, t(6;9), t(9;22) and complex karyotypes (≥3 unrelated abnormalities)	Complex karyotype (≥3 abnormalities), inv(3), t(6;9), t(6;11), -7, +8 sole, +8 with one abnormality, t(11;19)(q23;p13.1)

*Did not include t(15;17) in analysis

Abbreviations: MRC, Medical Research Council; SWOG/ECOG, Southwest Oncology Group/ Eastern Cooperative Oncology Group; CALGB, Cancer and Leukaemia Group B.

1.6 Pathophysiology

1.6.1 Clonality

The clonal nature of leukaemia, including AML is now beyond doubt. Recurrent cytogenetic abnormalities, as first described for the Philadelphia chromosome in chronic

myeloid leukaemia, gave an indication that leukaemia was a clonal disorder (Nowell and Hungerford 1960). In AML, there are many recurrent clonal cytogenetic abnormalities as discussed above. Clonality has also been demonstrated using X chromosome inactivation patterns (XCIPs) e.g. using glucose 6 phosphate dehydrogenase alleles (Fialkow, *et al* 1967) or HUMARA, the human androgen receptor locus (Gale, *et al* 1993). Although it is normal for X chromosomes from either parent to inactivate in equal number, under physiological conditions one chromosome can be inactivated more frequently than the other, e.g. due to aging. This skewing can be demonstrated by examining at the XCIPs of T lymphocytes. Therefore, to determine the clonality of the leukaemia, the background XCIP must be assessed. Mutation analysis gives more reliable evidence of clonality. Gene sequencing has demonstrated acquired single gene defects in AML e.g. in tyrosine kinase genes such as *FLT3*. These mutations are only found in the leukaemic clone (Kottaridis, *et al* 2001).

Further studies have shown a likely stem cell origin for leukaemia, implying the whole clone is derived from a single stem cell (Bonnet and Dick 1997). These studies involve the transplantation of subsets of leukaemic cells into immune deficient transgenic mice (the non-obese diabetic severe combined immunodeficient or NOD-SCID mouse). Clonal subsets of human leukaemic cells sorted by flow cytometry with the immunophenotype CD34+ and CD38-, can be transplanted into NOD-SCID mice and after several weeks the human leukaemia is recapitulated in the mouse. Furthermore, the leukaemia can be flow sorted again based on the above immunophenotypic markers and transplanted to other NOD-SCID mice, demonstrating that the stem cells can self-renew.

1.7 The function of genetic abnormalities in AML

1.7.1 Translocations

Many of the recurrent cytogenetic abnormalities described in AML are balanced

translocations. The breakpoints of these chromosomes are non-random double strand DNA breaks, within the exons of two genes of each chromosome. The break is in-frame, producing a new functional protein from the fusion of the two genes. The importance of fusion genes is that they are oncogenic, and are probably the underlying mutation in the development of the leukaemia. The identification of the fused genes has led to the discovery of novel oncogenes and functional studies of the translated fusion proteins have demonstrated their oncogenic potential (de The, *et al* 1990, Djabali, *et al* 1992). Many of the genes involved in translocations in AML are regulators of transcription, such as CBFs, retinoic acid receptor alpha (*RARA*) or the mixed-lineage leukaemia gene (*MLL)*. All these genes are involved in normal haematopoiesis (Gilliland, *et al* 2004).

CBFs are heterodimeric complexes made from alpha and beta subunits. CBFbeta, encoded by *CBFB*, and CBFA2 encoded by *RUNX1*, interact and are required for normal haematopoiesis. During embryonic development in mice, *RUNX1* is expressed in haemangioblasts, and deficiency of either *RUNX1* or *CBFB* leads to a lack of erythropoiesis, myelopoiesis and fetal death (Okuda, *et al* 1996). RUNX family proteins share a conserved 128 amino acid region, RHD, homologous to the Runt protein in *Drosophila melanogaster* (Mikhail, *et al* 2006). The RHD domain binds the alpha to the beta subunit, which then allows RHD to bind consensus DNA sequences in the promoter regions of several cytokines involved in haematopoiesis, including granulocyte-monocyte colony stimulating factor and interleukin-3 (Otto, *et al* 2003). Both *RUNX1* and *CBFB* are involved in several translocations in AML and ALL. *CBFB* fuses with *MYH11* by the inv(16) translocation, and produces an acute myelomonocytic leukaemia with eosinophilic differentiation (AML M4Eo). *MYH11* encodes a smooth muscle heavy chain protein. In yeast cells, CBFB-MYH11 binds to RUNX1 with other nuclear corepressors (NCoRs) e.g. mSin3A, and prevents transcription through a repression domain in MYH11 (Lutterbach, *et al* 1999). Additionally in mice, CBFB-MYH11 prevents normal haematopoiesis by inhibiting the CBF complex and preventing the differentiation of myeloid and lymphoid precursors (Kundu, *et al* 2002). The most common translocation involving *RUNX1* in AML is t(8;21)(q22;q22), which creates a fusion with *RUNX1T1*, another

transcription factor (also known as *ETO* or *MTG8*), and produces AML with mature myeloid blasts (AML M2). *RUNX1T1* encodes a zinc finger nuclear protein homologous to the *Drosophila nervy* gene (Gelmetti, *et al* 1998). The fusion protein RUNX1-RUNX1T1 recruits NCoRs and histone deacetylases (HD) that prevent DNA transcription (Gelmetti, *et al* 1998, Wang, *et al* 1998). The RHD domain of the RUNX1 part of the fusion protein then targets cytokine promoter sequences for transcription repression, thus blocking myeloid and erythroid differentiation (Peterson and Zhang 2004). It would appear that CBF leukaemias have similar mechanisms for transcription repression, and they have other similarities, both having being associated with a relatively good prognosis and being associated with *CKIT* mutations (Paschka, *et al* 2006).

The PML/RARA fusion protein is found in 99% of APML, and usually arises from the translocation t(15;17)(q22;q21) (de The, *et al* 1990, Goddard, *et al* 1991, Pandolfi 2001). A common motif between CBF leukaemias and RARA fusion leukaemias is their ability to reduce transcription through NCoRs and HDs. Retinoic acid receptors (RAR) interact with retinoid-X-receptors and NCOR2 (also known as SMRT), which then recruit HDs to repress transcription of target genes that have retinoic acid response elements (Chen and Evans 1995, Heinzel, *et al* 1997). Interaction of RAR with its ligand, retinoic acid (RA), allows dissociation of the NCoR/HD complex, permitting transcription. PML/RARA also binds RXR and NCoR/HD, leading to transcription repression that is reversed by pharmacological levels of RA (Grignani, *et al* 1998, Minucci and Pelicci 2007). At a cellular level, the effects of *PML/RARA* lead to a differentiation block at the promyelocyte stage, but RA allows neutrophil differentiation (Grignani, *et al* 1993). This leads to hyperleucocytosis due to an accumulation of differentiated neutrophils when APML is treated with ATRA. A rare form of APML involving the translocation t(11;17), produces the fusion gene PLZF/RARA. It also binds NCoR and HD causing transcription repression, but has little or no response to RA (Grignani, *et al* 1998).

MLL is a homologue of the *Drosophila trithorax* gene, required for embryonic and haematopoietic development (Ayton and Cleary 2001). Like *trithorax*, it regulates HOX gene expression in embryonic development, including haematopoietic development (Hess, *et al*

1997). *MLL* is very large, coding for a protein with 3,969 amino acids and is 430kD. It contains several different functional protein subunits that are formed by post-translational cleavage to produce large multimeric complexes (Nakamura, *et al* 2002). An example of one of the subunits is the SET domain at the C terminal, which codes for a histone H3 lysine-4 methyltransferase that methylates histones at the promoter of *HOXA9*. *MLL* is the most promiscuous gene involved in translocations in acute leukaemia, with many different partner genes (DiMartino and Cleary 1999). The breakpoint of *MLL* is consistently between exons 5 and 8, leading to retention of motifs for DNA binding at the N terminal, but loss of the motifs involved in histone methylation at the C terminal. Histone modification is still important in the function of *MLL* fusion genes because the histone methylation gene, *hDOT1L*, is required for the MLL-AF10 fusion protein to promote leukaemogenesis by targeting *HOXA9* (Okada, *et al* 2005). At least one of the functions of the N-terminal motifs in MLL fusion proteins is to interact with a known tumour suppressor gene (TSG), *MEN1*, at HOX gene promoters, leading to leukaemogenesis in mice (Yokoyama, *et al* 2005). These mechanisms lead to alterations in HOX gene expression in AML with *MLL* translocations, as has been observed using expression microarrays in primary AML samples (Debernardi, *et al* 2003).

1.7.2 Single gene mutations

Table 1.3 Common single gene mutations in AML

Mutation	Mutation location and type	Predominant AML subtypes	Reference
RUNX1	Runt domain, point mutation	AML M0	(Roumier, *et al* 2003)
MLL	PTD	NK, +11	(Rege-Cambrin, *et al* 2005, Whitman, *et al* 2007a)
FLT3	JM region, ITD	NK, t(15;17), t(6;9)	(Kottaridis, *et al* 2001) (Thiede, *et al* 2002)
	Loop /TKD, point mutation	NK, inv(16)	(Mead, *et al* 2007)
NPM1	Exon12, 4bp insertion	NK	(Falini, *et al* 2005)
C-KIT	Loop /TKD, point mutation	inv(16), t(8;21)	(Paschka, *et al* 2006)
NRAS	Codons 12, 13 and 61, point mutations	t(3;5), inv3/t(3;3), inv(16)	(Bowen, *et al* 2005) (Bacher, *et al* 2006)
KRAS	Codon 61, point mutation	inv(16)	(Bowen, *et al* 2005)
WT1	Exons 7-9	NK	(King-Underwood, *et al* 1996, Summers, *et al* 2007)

Abbreviations: NK, normal karyotype; PTD, partial tandem duplication; ITD, internal tandem duplication; JM, juxtamembrane; TKD, tyrosine kinase domain.

Some of the genes involved in chromosomal translocations may also acquire mutations contributing to leukaemogenesis when there are no chromosomal rearrangements (

Table 1.3). Acquired point mutations of *RUNX1* are found in undifferentiated AML (AML M0), and occasionally in other FAB types (Preudhomme, *et al* 2000, Roumier, *et al* 2003, Smith, *et al* 2005). Most are found in the Runt domain, involved in DNA binding. They lead to decreased DNA binding and sometimes decreased binding to CBFbeta (Matheny, *et al* 2007). Mouse models of these *RUNX1* mutations cause abnormal haematopoiesis when the mutation homozygous, but some produce only a mild haematopoietic defect when heterozygous (Matheny, *et al* 2007). In primary AML, these mutations and others that cause deletion or early truncation of *RUNX1*, are frequently heterozygous, suggesting haploinsufficiency of *RUNX1* expression is important in the pathogenesis of AML (Mikhail, *et al* 2006). However, some mutations are homozygous, implying a gene dosage effect for the mutation (Silva, *et al* 2003).

MLL partial tandem duplications (PTD) occur in about 5% of AML and more in those with a normal karyotype or trisomy 11 (Basecke, *et al* 2006). They cause an in-frame repetition of a pair of fused exons e.g. exon 9 with exon 3 (e9/e3), but the functional abnormality is uncertain. *MLL* PTD retains the C terminus of the protein and there is evidence of lack of expression of the wild type *MLL* allele, in contrast to *MLL* translocations (Whitman, *et al* 2005). Therefore, it retains histone methyl transferase activity and can affect HOX gene expression, as demonstrated by the overexpression of *HOXA* genes in mouse models with *MLL* PTD (Dorrance, *et al* 2006). *MLL* PTD at diagnosis confers a poor prognosis although this may be ameliorated by more intense chemotherapy (Whitman, *et al* 2007a).

Other genes involved in oncogenesis are also found mutated in AML. *RAS* encodes a GTPase and is an oncogene commonly mutated in cancer. It was originally found in bladder cancer as a homologue of a gene in murine sarcoma viruses (Parada, *et al* 1982), and is involved in several cell-signalling pathways. Amongst the most important pathways it affects, are the mitogen-activated protein (MAP) kinase pathway involved in proliferation and the phosphoinositol-3 kinase (PI3K)/Akt pathway that promotes cell survival (Schubbert, *et al* 2007). The mutations affect the GTPase activity of the protein, preventing its negative regulation by GTPase activating proteins. Mutations in the *RAS* pathway are found in a number of myeloid malignancies, especially juvenile myelomonocytic leukaemia (JMML) in which

mutations of *RAS*, *NF1* and *PTPN11* are frequent. Three isoforms of RAS have been identified, *HRAS*, *KRAS* and *NRAS*. Mutations of *NRAS* are found in approximately 10% of AML patients and *KRAS* mutations in 5% (Bacher, *et al* 2006, Bowen, *et al* 2005).

RAS GTPases are downstream of several growth factor receptor kinases such as the granulocyte colony stimulating receptor and the erythropoietin receptor. FLT3 (fms-like tyrosine kinase 3) and c-KIT (cellular homolog of the feline sarcoma viral oncogene v-kit) receptors are type 3 receptor tyrosine kinases (RTKs) that can both be mutated in AML (Kitamura and Hirotab 2004, Parcells, *et al* 2006). Type 3 RTKs are a family of transmembrane proteins involved in cell signalling; other members include platelet derived growth factor receptors alpha and beta, and CSF1-R. They have an extracellular domain that forms homodimers when its ligand attaches, and an intracellular tyrosine kinase that phosphorylates downstream targets including PI3K and RAS. Both are involved with the proliferation of haematopoietic progenitors, when the appropriate cytokine attaches, FLT3 ligand to FLT3 receptor or stem cell factor (SCF) to CKIT (Kitamura and Hirotab 2004, Wodnar-Filipowicz 2003). Somatic mutations of *FLT3* occur in about 25% of AML patients (Kottaridis, *et al* 2001, Thiede, *et al* 2002). The most common mutation is an internal tandem duplication (ITD) that affects the juxtamembrane region, leading to constitutive activation of the kinase (Griffith, *et al* 2004). This activates the RAS signalling pathway, and explains why it is rare for *RAS* and *FLT3* mutations to coincide in the same patient (Bowen, *et al* 2005). *FLT3* ITD mutations confer a poor prognosis, emphasising their importance in the pathogenesis of AML. Other mutations of *FLT3* are found in AML, especially in the activation loop or tyrosine kinase domain (TKD). These also cause constitutive activation of the kinase, but their prognostic importance is controversial (Mead, *et al* 2007, Whitman, *et al* 2007b). Mutations of *KIT*, which encodes the c-KIT receptor, are found in mast cell neoplasms (Kitamura and Hirotab 2004), and t(8;21) and inv(16) AML. In the latter, they may confer a worse prognosis (Paschka, *et al* 2006). Mutations occur in the activation loop of the receptor, at an analogous location to TKD mutations of *FLT3*. They also cause constitutive activation of the tyrosine kinase.

CEBPA is a transcription factor gene important in the development of haematopoiesis.

It affects the differentiation of granulocytes and monocytes from multilineage precursors (Laslo, *et al* 2006), so that mice deficient in *CEBPA* have no neutrophils. Mutations of *CEBPA* occur in fewer than 10% of AML patients, and mostly affect those with a normal karyotype (Leroy, *et al* 2005). The mutations are in two regions of the gene, at the carboxyl terminal affecting DNA binding, or at the amino terminal leading to a truncated 30 kD form (Snaddon, *et al* 2003). The latter is a dominant negative mutation, so a heterozygous mutation is apparently sufficient for leukaemogenesis, but not infrequently both alleles are mutated. In patients with a familial *CEBPA* mutation, the heterozygous mutation is inherited in an autosomal dominant fashion, but a second somatic *CEBPA* mutation is acquired at diagnosis of AML, leading to a biallelic mutation (Smith, *et al* 2004b).

More recently, a mutation of *NPM1* (nucleophosmin) has been found to be prevalent in normal karyotype AML (Falini, *et al* 2005). *NPM1* is a fusion partner in several other haematological malignancies, including *PML/NPM* found in rare forms of APML, and *NPM/ALK* in anaplastic large cell lymphoma. It is a transcription factor involved in the nuclear-cytoplasmic shuttling of proteins and nucleic acids (Grisendi, *et al* 2006). Mutation of exon 12 leads to cytoplasmic rather than nuclear expression in AML cells because of defects in its shuttling ability. The heterozygous mutation has a dominant negative effect because it forms heterodimers with the wild type protein. At least one of the effects of the mutation is to delocalise ARF (also known as p21), a growth inhibitor. In the absence of *FLT3* ITD mutations, *NPM1* mutations confer a relatively good prognosis in patients with normal karyotype AML (Falini, *et al* 2005).

Small studies into *WT1* have demonstrated mutations in about 10% of AML patients (King-Underwood, *et al* 1996, Summers, *et al* 2007), moreover there is an association with refractory disease. *WT1* was originally described as a TSG associated with Wilm tumours, but was soon discovered to be critical for renal development because inherited mutations lead to renal disease including Denys-Drash syndrome (Wagner, *et al* 2003). Although it encodes a transcription factor with a zinc finger region, it has many isoforms due to alternate splicing and is expressed in so many tissues that it is unclear how it functions. A recent paper has reported

51 mutations in 470 AML samples from the MRC AML10 and 12 studies, confirming the results of the smaller studies, and has found most of the mutations affect the C terminal DNA binding domain (Virappane, *et al* 2008).

As for all cancers, multiple mutations are required for development of the malignancy (Kinzler and Vogelstein 2002). Translocations only lead to AML in transgenic mouse models after several months, presumably after further mutations have taken place (Cuenco, *et al* 2000). Other mutations, such as *FLT3* ITD, only led to a myeloproliferative disease rather than AML in mouse models (Lee, *et al* 2005). The familial leukaemias such as familial *CEBPA* mutations and FPD/AML show long latencies before AML is apparent, consistent with the multi-step theory of cancer or leukaemia development.

Some mutations have a tendency to occur together, such as *NPM1* and *FLT3* ITD (Falini, *et al* 2005), and inv(16) or t(8;21) and *C-KIT* (Paschka, *et al* 2006). Other mutations are mutually exclusive such as *RAS* and *FLT3*, presumably because they act along the same pathway (Bowen, *et al* 2005). It has been proposed that there are two types of mutation, referred to as class I (proliferative) and class II (impairing differentiation), which are both required to cause AML (Kelly and Gilliland 2002). Examples of class 1 mutations would be *FLT3, CKIT* or *RAS*. Examples of class 2 mutations would be translocations such as t(8;21) or mutations such as *CEBPA*.

Although haematopoietic cells are thought to have fewer mutations than solid tumours, it is still likely there are several genetic events, be they other mutations, epigenetic or in non-coding regions that may contribute to the leukaemia, or at least alter the phenotype. Therefore, this theory is likely to be a simplification of the pathogenesis of AML. In addition, the function of many of these genes is more complex than simply preventing differentiation or being proliferative. For example, NPM protein has anti-apoptotic and proliferative effects as well as preventing differentiation (Grisendi, *et al* 2006).

1.8 Therapy of AML

Despite an increased understanding of the biology of AML in the last 20 years, the therapeutics of AML has changed relatively little in that time. Most patients with AML are in an elderly age group and their prognosis from this risk factor alone would dictate that they should be treated supportively rather than with curative intent. Other patients may have co-morbidities e.g. ischaemic heart disease, that weigh against the benefits of intensive chemotherapy. Supportive care includes blood product support, antibiotics when infections arise and some degree of myelosuppressive therapy, if the patient is symptomatic from a rising leukaemic blast count. Appropriate myelosuppresssive drugs include 6-mercaptopurine (6-MP), low dose cytarabine arabinoside (araC) or hydroxyurea. In a large prospective randomised trial between low dose araC and hydroxyurea for elderly AML (MRC AML14), low dose araC had an improved overall survival (OS) (odds ratio 0.6; 95% confidence interval, 0.44-0.81; P=0.0009), although there was only a 13% OS for all patients at 1 year (Burnett, *et al* 2007).

In higher doses 6-MP and araC are used with curative intent. After the Second World War, thiopurines such as 6-MP were amongst the first cytotoxic agents in development to have an effect on AML (Burchenal, *et al* 1953, Scott 1957). By the 1960s the purine analogue, araC, and anthracycline, daunorubicin, were used as single agent therapy for AML, and could produce a small number of morphological complete remissions (CR) (Malpas and Scott 1968). Their combined use led to 60-65% of patients going into CR (Crowther, *et al* 1970, Yates, *et al* 1973). The protocol of Yates et al, DA 3+7 (three daily doses of daunorubicin, and 7 days of araC) has become the template for the modern standard of induction chemotherapy.

The most important chemotherapeutic agents continue to be anthracyclines and cytarabine arabinoside. The initial induction course of chemotherapy causes myelosuppresion for 20 to 30 days, with the aim of producing CR. This can be achieved in adults under the age of 50-55 in 73-85% of patients (Grimwade, *et al* 1998, Rohatiner, *et al* 2000), but this falls to 55-60% in those over 60 years of age (Grimwade, *et al* 2001, Milligan, *et al* 2006).

Remission is consolidated with further courses of chemotherapy, e.g. high dose araC, or

it may involve high dose chemotherapy (HDT) with autologous stem cell support. Debate surrounds the dose of araC required, although a higher dose is associated with a prolonged survival (Bradstock, *et al* 2005, Mayer, *et al* 1994). Trials at Barts using historical controls show an improvement in recurrence free survival (RFS) for HDT (hazard ratio 0.5 P=0.005) but no significant improvement in OS in comparison with historical controls having chemotherapy alone (Rohatiner, *et al* 2000). The national prospective randomised trial, MRC AML10, gave similar results for HDT, with an improvement in RFS (54% v 40%, P=0.04) but not in OS (Burnett, *et al* 1998). In patients at high risk of relapse, consolidation with allogeneic stem cell transplantation, providing a graft versus leukaemia effect, may be considered. However, the high treatment related mortality (TRM) may negate the improved RFS. In recent years, four large prospective trials have reported the results of randomisations between those patients who have an available related matched donor for allogeneic transplant and those without (Burnett, *et al* 2002, Cornelissen, *et al* 2007, Jourdan, *et al* 2005, Suciu, *et al* 2003). In MRC AML10 there was no significant improvement in OS for those autografted against those without an allogeneic donor (56% versus 50% at 7 years, P=0.01), although there may be an advantage in OS for intermediate risk patients (Burnett, *et al* 2002). A meta-analysis of these four trials for OS demonstrated a benefit for those with a donor, especially for those with intermediate and poor risk cytogenetics, and for those below the age of 40 years (Cornelissen, *et al* 2007). These results imply that good risk patients should be treated with chemotherapy rather than HDT, but there is a place for allogeneic transplantation for intermediate and poor risk patients. Trials are currently underway using reduced intensity instead of ablative conditioning for allogeneic transplant, as this may be benefit older patients for whom the TRM of ablative conditioning is too high (Davies, *et al* 2006, Mohty, *et al* 2005). These trials have already shown a long-term benefit for some patients.

Consolidation is designed to reduce the disease load, because although these patients are in morphological and clinical remission after induction, leukaemia cells are still present. The minimum levels of discrimination of minimal residual disease by morphology or fluorescent *in situ* hybridisation (FISH) is approximately 1 in 100 cells; by immunophenotyping it is 1 in 10^3-

10^4 cells (van Dongen, *et al* 2005), and the best discrimination by polymerase chain reaction (PCR) is about 1 in 10^4-10^5 (Grimwade 2002). Therefore, even below an undetectable level, leukaemic cells are present. One of the aims of new treatments has been to improve consolidation therapy to prevent relapse from the residual leukaemia.

Novel targeted therapies are now being used based on the biological understanding of AML. ATRA used in conjunction with chemotherapy has improved the outcome of APML (Tallman, *et al* 1997) (although the discovery of ATRA preceded our understanding of the biology of APML). Other targeted therapies are in clinical trials. Gemtuzumab ozogamicin is a conjugate of an antibody against CD33 and a cytotoxic antibiotic. It has produced CRs in some patients in phase 2 trials (Sievers, *et al* 2001), and is now in phase 3 trials, including the UK national AML trial for patients below the age of 60 years, AML15. The FLT3 receptor inhibitor, CEP701, has also been used in AML15, because it has shown clinical responses in phase 2 trials (Knapper, *et al* 2006, Smith, *et al* 2004a).

Although there is reason to hope that these new treatments may improve the outcome of AML, it is new biological targets may only come about from an improved understanding of the biology of AML. One of the most significant steps forward in this understanding has come from the use of microarray technologies.

1.9 Genomics and the study of AML

The Human Genome Project (HGP) has allowed detailed annotation of the whole of the human genome allowing a huge amount of data to be mined (International Human Genome Sequencing Consortium 2001). Several large-scale projects have subsequently been developed from the HGP, including the Cancer Genome Project (Greenman, *et al* 2007), the HapMap (International HapMap Consortium 2005) and more recently and more recently a project advanced by the Human Genome Structural Variation Working Group (Human Genome Structural Variation Working Group 2007).

Technological advances have been at least part of the reason these advances have been

possible. These technologies have also been put to use in providing high-throughput techniques that can analyse many parts of the genome on a single sample. Microarrays were the first example of this, initially spotting hundreds or thousands of unique nucleotide sequences at a time, to which a DNA or RNA sample could be hybridised (Ramsay 1998). The signal amplitude produced by each spot represents the quantity of each unique sequence in the sample. Others have developed technologies to increase the number of nucleotides that can be analysed on a slide or chip, e.g. Affymetrix uses photolithography, and Illumina uses micro bead technologies.

Many of the initial uses of microarrays were in acute leukaemia, analysing the transcriptome, in other words mRNA expression (Golub, *et al* 1999). Further developments have seen their use in analysing DNA copy number using the arrays as an alternative to comparative genetic hybridisation (CGH), the technology being called array CGH (Wicker, *et al* 2007). Microarrays have also been developed to show homozygosity or heterozygosity for the alleles of single nucleotide polymorphisms (SNPs) (Affymetrix 2003). These SNP arrays have become useful to find regions of loss of heterozygosity in cancer.

1.9.1 Gene expression profiling

The level of expression of mRNA of individual genes reflects the phenotype of the leukaemic cell, as it is the translated genes that are involved in the functions of the cell. Several genes have been described that were highly expressed in AML cells, e.g. *WT1* (Inoue, *et al* 1994), *EVI1* (Barjesteh van Waalwijk van Doorn-Khosrovani, *et al* 2003), *BAALC* (Tanner, *et al* 2001) and MN1 (Heuser, *et al* 2006). *WT1* was originally found mutated in association with Wilm's tumours, and was subsequently found to be over expressed in neuroectodermal and some haematological malignancies. *EVI1* is involved in translocations of chromosome 3q in AML and MDS, and was subsequently found to be over expressed in AML. *BAALC* is on the long arm of chromosome 8, and was found over expressed in some AML patients with trisomy 8 using representational difference analysis, comparing them with AML patients with a normal

karyotype (Tanner, *et al* 2001). MN1 is involved in a rare t(12;22) translocation in AML (Heuser, *et al* 2006).

While the function of these genes is currently being investigated, clinical studies have shown high expression of these genes is associated with a poor outcome. A high *WT1* expression was associated with a 3 year OS of 13% versus 38% (P=0.038) in 129 patients (Bergmann, *et al* 1997) in one study, although another could find no association in 125 patients (Schmid, *et al* 1997). High expression of WT1 is currently being investigated as a marker for minimal residual disease after chemotherapy in the UK national clinical trial, AML15. In a study of 309 patients, those with high *BAALC* expression under the age of 60 years had a 3 year OS of 36% v 54% (P = .001) (Baldus, *et al* 2006). High *EVI1* expression was associated with poor risk cytogenetics, but also predicted a poor OS in patients with intermediate risk cytogenetics (P=0.05), although patient numbers were small (Barjesteh van Waalwijk van Doorn-Khosrovani, *et al* 2003). High *MN1* expression predicted a poor OS in 142 patients with AML (38.1% versus 58.8% at 3 years, P = .03) (Heuser, *et al* 2006).

If single gene expression can give prognostic information in AML, then the expression of many genes may be able to improve the classification of AML. Expression arrays measure the mRNA expression of thousands of genes across the genome simultaneously. Initial studies demonstrated it was possible to differentiate AML from ALL (Golub, *et al* 1999), and subsequently patterns have been found that can distinguish AML phenotypes and cytogenetic subtypes (Debernardi, *et al* 2003). Because many of these subgroups have prognostic importance, the next step was to show that expression patterns can correlate with current prognostic subgroups and define new prognostic groups (Valk, *et al* 2004). This is leading to international collaborations to use expression profiles as a diagnostic tool in AML and other haematological malignancies in the MILE study (Mills 2005).

Whilst molecular classification has been very successful, finding the biological importance of expression profiles has been more difficult. Although the level of mRNA expression does not necessarily correlate with protein expression, for the genes important in AML their correlation seems to be good (Kern, *et al* 2003). A few genes have been identified

from genome wide expression studies, including *MN1*, which was highly expressed in the profiles of patients with a poor response to chemotherapy (Heuser, *et al* 2006). There are several complications in discovering individually important genes from expression profiling. The number of genes in an array means to make statistically significant correlations requires large numbers of samples. Influences on the level of expression of many genes are multifactorial, leading to variability in their expression. There are also factors independent of the leukaemia that may affect expression profiles e.g. delayed sample collection leading to increased leukaemic cell apoptosis. From a biological point of view, mRNA expression takes no account of different splice variants, e.g. the expression of exon 9a from *RUNX1T1* in AML with t(8;21) has been shown to be leukaemogenic (Yan, *et al* 2006). Expression of mRNA also does not reveal the effects of post-translational protein modifications e.g. glycosylation.

Despite these problems, there are robust patterns of expression that suggest groups of genes that are involved in leukaemogenesis. AML patients with translocations involving MLL have an increased expression of some HOX A and B genes, in a similar pattern to some patients with normal karyotype AML (Debernardi, *et al* 2003). These patients also seem to be associated with *NPM1* mutations (Alcalay, *et al* 2005). As well as patterns, expression of individual genes is leading to an understanding of some of the downstream effects of mutations. Increased CXCR4 expression is seen in a subset of patients that includes those with *FLT3* ITD mutations, and confers a poor prognosis (Rombouts, *et al* 2004). The homeobox gene, *CDX2* has been found to be highly expressed in almost all AML patients, and can produce a transplantable AML mouse model, suggesting that this is part of a common pathway in AML (Scholl, *et al* 2007).

In addition to the deregulation of expressed genes, the effect of non-coding RNA transcripts on the leukaemia phenotype has been described. MicroRNAs are short hairpin shaped oligonucleotides that prevent the translation of specific mRNAs (Calin and Croce 2006). They are coded in intergenic and intronic regions as long precursor forms of about 60-100 nucleotides, before being spliced into their active short form of about 20-22 nucleotides. One microRNA, mir-181a has been associated with acute monocytic and myelomonocytic leukaemia

(Debernardi, *et al* 2007). As most of the genome is untranslated there are likely to be many more non-coding regions that are important in leukaemogenesis.

1.9.2 DNA arrays

CGH uses a comparison of pooled normal DNA to tumour DNA hybridised to normal chromosomes to see gains and deletions (Kallioniemi, *et al* 1994). Its advantage compared to conventional cytogenetic analysis is the ability to discriminate relatively small changes that may be missed when looking at chromosomal bands. Array CGH (aCGH) is an even higher resolution version of this process (Strefford, *et al* 2006). In this case, the comparison is made on hybridisation to thousands of spotted probes across the genome on a microarray slide. The probes are usually bacterial artificial clones (BACs) that are from 1Mb to a few thousand bases in size, considerably smaller than a chromosome band. More recently, the resolution has increased so that tiling arrays for individual chromosomes are using oligonucleotide probes (Wicker, *et al* 2007). It has been used predominantly in solid tumours where metaphase cytogenetic analysis is much more difficult. A recent study using relatively low resolution aCGH has identified recurrent regions of gain and deletion in complex karyotype AML (Rucker, *et al* 2006). CDX2 was discovered as a potential important gene in AML from its amplification on chromosome 13 by aCGH (Scholl, *et al* 2007).

1.10 SNP array technologies

Single nucleotide polymorphisms (SNPs) are the commonest genetic variation in the human genome, occurring about every 200bp (Sachidanandam, *et al* 2001). They are mostly biallelic so the two alleles may form transitions between purines i.e. adenine and guanine, or pyrimidines i.e. thymine and cytosine. Alternatively, transversions can occur between purine and pyrimidines. They are located throughout the genome, although they are more frequent in GC rich regions (Sachidanandam, *et al* 2001); this may be due to nucleotide excision repair

correcting methylated cytosines (CpG motifs), which can be converted to thymine.

Most SNPs are in intergenic regions and whilst they are useful markers (see association studies below), they are unlikely to have effects on cellular function. However, SNPs in gene promoter regions have the potential to alter the expression of their associated gene, and SNPs within exons alter the coding of the gene itself. SNPs in exons can be synonymous, altering the codon but not the amino acid, or non-synonymous, altering the amino acid structure of the gene and producing a polymorphic protein. In pharmacogenetics, non-synonymous polymorphisms are responsible for variation in the function of drug-metabolising enzymes (Ahmadi, *et al* 2005, Marsh and McLeod 2004). Recently, even synonymous SNPs have been shown to alter the function of proteins, e.g. synonymous *MDR1* variants have altered ability to pump fluorescent substrates out of cells (Kimchi-Sarfaty, *et al* 2007). Synonymous codons are not used equally in the genome, and it is hypothesised that alternate tRNA molecules complementary to these codons may alter protein folding.

1.10.1 Association studies

The database for SNPs is maintained in an open access repository known as dbSNP (National Center for Biotechnology Information 2007). There are currently over 11.75 million SNPs documented in the database from build 128. Each SNP is recorded with data about its flanking sequence, its genomic location and population frequency of each allele where available.

The HapMap Project (International HapMap Consortium 2005) has produced a database of the distribution of SNPs in various populations and the way in which linkage disequilibrium causes them to be inherited as haplotypes blocks. Various ethnic groups have inherited different haplotypes, because of the migration patterns of peoples since the original evolution of *Homo sapiens*. The HapMap Project has analysed African (Yoruban, from West Africa), Caucasian (Americans of northern European ancestry), and east Asian (Japanese and Chinese) individuals. One of the main aims of the genome-wide approach to genotyping has been to look for inherited

haplotypes that may be associated with diseases that have polygenic inheritance, e.g. non-insulin dependent diabetes mellitus. High-throughput techniques using microarrays are now able to look at several thousand SNP allelotypes at once (Lindblad-Toh, *et al* 2000), and have been used to identify novel loci associated with disease, e.g. in breast cancer (Easton, *et al* 2007) and prostate cancer (Gudmundsson, *et al* 2007, Yeager, *et al* 2007). These studies require large populations of patients with disease and controls, because of the statistical significance required with such large numbers of SNPs, and critically they require repetition in other populations to reduce the risk of false positives.

The haplotype approach means the number of SNPs in these arrays can be reduced when performing association studies, as many SNPs in the same haplotype block will be in linkage. Therefore, single tagging SNPs can be used to represent a whole haplotype. For example, tagging SNPs have been used to produce a set of SNPs to represent variation in drug metabolism enzymes and other genes that affect pharmacokinetics (Ahmadi, *et al* 2005).

Another use for high-resolution SNP arrays is for linkage analysis, to identify genes in families with rare inherited disorders. The rare autosomal recessive disease harlequin ichthyosis, a fatal skin disease, is such an example. Linkage analysis and a common region of homozygosity identified a mutation in ABCA12 at chromosome 2q35 (Kelsell, *et al* 2005).

1.10.2 Cancer studies

The identification of somatically acquired regions of homozygosity i.e. loss of heterozygosity (LOH), is important in cancer. Regions of LOH have been assumed to contain TSGs, which would be mutated and then undergo gene conversion, as is consistent with Knudson's two hit hypothesis (Knudson 1971). Most genetic abnormalities in cancer are somatically acquired, involving amplifications, deletions or smaller single gene mutations. SNP arrays can be analysed to look for LOH in the tumour compared to germline DNA. Prior to this technology, LOH was identified by the PCR of microsatellite markers, which consist of short tandem repeats polymorphic in their length. There are far fewer microsatellites compared with

SNPs, limiting their resolution and genome coverage.

An additional advantage of SNP arrays is to give a measure of DNA copy number. Calling a SNP homozygous or heterozygous requires a quantitative assessment of one allele against the other. The signal values used can also measure the total quantity of DNA in a similar manner to aCGH (Huang, *et al* 2004). Therefore, SNP arrays have an advantage over aCGH, in simultaneously identifying deletions and LOH.

Complicating the detection of somatically acquired gains and deletions is the recognition of structural polymorphisms in the human genome i.e. insertion-deletions (indels) (Conrad, *et al* 2006) and copy number variants (CNVs) (Sebat, *et al* 2004). The median size of these variants is 228kb and they are distributed across the genome, affecting many transcribed genes (Redon, *et al* 2006). There is at least one report associating a copy number polymorphism with disease, in a type of glomerulonephritis (Aitman, *et al* 2006). Many of these discoveries have been made using high-resolution SNP arrays, which emphasises the need to eliminate germ line variants when searching for somatic aberrations.

Initial studies used arrays with approximately 1500 SNPs to identify LOH in small cell lung cancer, using paired normal lung tissue as a control, and CGH to confirm these as deletions (Lindblad-Toh, *et al* 2000). Subsequently, copy number was also analysed with LOH in cancer cell lines with paired B lymphoblastoid cell lines (Zhao, *et al* 2004). This study described 35 amplifications and deletions, as well as two cases of copy number neutral LOH in chromosome 9 and 13, each confirmed by real time PCR. This was the first description of acquired uniparental disomy (UPD) using a SNP array. Analyses of other tumours including bladder (Hoque, *et al* 2003), lung (Janne, *et al* 2004) and breast (Schubert, *et al* 2002) similarly demonstrate the highly disordered nature of chromosomes in epithelial solid tumours.

Data from the Mitelman Database of Chromosomal Aberrations in Cancer (National Cancer Institute 2006) confirm that solid tumours often have a much more disordered karyotype than leukaemias (Mitelman, *et al* 2004). Unlike most cases of solid tumours, leukaemias usually have conventional cytogenetic analysis performed at diagnosis, allowing a comparison with array based chromosomal analysis including SNP arrays. Both of these factors give an

advantage to analysing AML by SNP arrays. Firstly, confirmation of the karyotype by SNP array demonstrates the validity of using this platform (Chapter 3). Secondly, abnormalities detected by SNP arrays that are cryptic to karyotype analysis can identify mechanisms of disease pathogenesis. This is particularly important when 40% of AML have no karyotypic abnormality. Analysis of paediatric precursor B cell ALL with arrays of 500,000 SNPs have identified microdeletions of less than 1Mb that contain genes involved in B cell development, e.g. *PAX5* (Mullighan, *et al* 2007). In AML, acquired UPD has been discovered in 15-20% of cases (Gorletta, *et al* 2005, Raghavan, *et al* 2005), and the region of LOH in many cases contains homozygous mutated genes involved in the pathogenesis of AML, e.g. *CEBPA* (Fitzgibbon, *et al* 2005) (Chapter 4). The application of SNP arrays in haematological malignancy to identify novel genetic aberrations has now been made in myeloma (Walker, *et al* 2006), follicular lymphoma (recurrent UPD of 6p) (Fitzgibbon, *et al* 2007) and juvenile myelomonocytic leukaemia (UPD of 17q resulting in a homozygous NF1 mutation) (Flotho, *et al* 2007, Stephens, *et al* 2006). Additionally, acquired UPD has been discovered recurrently in epithelial tumours e.g. basal cell carcinomas have UPD 9q with the region of LOH containing a homozygously mutated *PTCH* gene (Teh, *et al* 2005). The following chapters outline the development of the use of SNP arrays to analyse AML for regions of LOH and DNA copy number, the discovery of acquired UPD as a recurrent genetic aberration in AML, and the consequences of UPD for the pathogenesis and subsequent progression of the disease.

Chapter 2

Materials and Methods

2.1 Ethics approval

Approval had previously been obtained for the collection and cryopreservation of leukaemia samples, blood and bone marrow (study P/02/067 approved by East London and City Local Research Ethics Committee). Excess samples were obtained with consent for storage as part of the clinical management of the patients. Samples were stored in the Medical Oncology Tissue Bank, Barts and the London School of Medicine and Dentistry Human Tissue Resource Centre (Human Tissue Authority deemed license 12199). For this study, ethical approval was obtained for SNP genotyping of the blood and marrow samples and genotyping germ line material from remission samples (study P1/03/310 approved by East London and City Local Research Ethics Committee).

2.2 Cryopreservation of leucocytes

AML bone marrow and blood samples were previously taken from patients at diagnosis and at subsequent times during the course of their disease. Some samples were obtained by leucopheresis.

Leucocytes were Ficoll separated and cryopreserved in liquid nitrogen. In detail, an

equal volume of RPMI-1640 medium (Moore, *et al* 1967) was added to the leucocyte sample (in ACDA, acid citrate dextrose solution, or EDTA, ethylene diamine tetra-acetic acid). The sample was divided into 10ml aliquots and added carefully to 5 mls of Lymphoprep (AXIS-SHIELD PoC AS Norway) (Sodium Diatrizoate 9.1% (w/v) Polysaccharide 5.7% (w/v)) as to maintain the layers. It was centrifuged at 1500 rpm for 25 mins with no brake. The buffy coat was taken off into 10mls RPMI and centrifuged at 1300 rpm for 10 mins. The supernatant was discarded and the pellet resuspended in RPMI in a cryovial. Freeze mix was added drop wise. The vial was stored in liquid nitrogen.

2.3 DNA extraction

Genomic DNA was extracted using standard phenol-chloroform methods. In detail, the samples retrieved from cryopreservation were incubated for 1-2 mins at 37°C. The suspension of cells was taken into a 2 ml tube and centrifuged at 13,000 rpm in a bench centrifuge (Eppendorf 5417R) for 5 mins at 4°C. The supernatant was removed and the pellet of cells resuspended in 1-2 ml (or more) of digestion buffer (100 mM NaCl, 10 mM Tris HCl pH 8, 25 mM EDTA pH 8, 0.5% SDS) depending on the size of the pellet. If necessary, the suspension was transferred to a 15 ml falcon tube. 50-100 µl of proteinase K (in a buffer of 50 mM Tris pH 8 and 1.5 mM Ca acetate and kept on ice) was added to make about 0.1 mg/ml. The mixture was vortexed and incubated in a water bath at 56°C for 12-18 hours overnight.

The digested cells had an equal volume of phenol (ultrapure, equilibrated pH8.0) (USB Corporation, Cleveland, OH, USA) added to them. The mixture was shaken and centrifuged for 10 mins (3000 rpm for falcon tubes, 10,000 rpm for bench centrifuge). The upper aqueous layer containing DNA was removed and put into a fresh tube and the remaining organic layer containing protein discarded. An equal volume of chloroform and phenol (each 50% by volume) was added and the solution mixed and spun as before. Again, the aqueous layer was put in a fresh tube avoiding the cloudy interface. If the solution was still cloudy, a further wash with phenol-chloroform was performed; otherwise, an equal volume of chloroform only was added.

The solution was mixed and spun as before and once more the aqueous layer taken into a 15 ml falcon tube. One-tenth-volume Sodium citrate 0.3 M and twice the volume of 100% ethanol was added. The solution was gently agitated to allow the strands of DNA to precipitate and coalesce. The DNA was removed with a sterile inoculation loop and washed with a few drops of 75% ethanol. After drying a little in air, the DNA was placed in an appropriate volume of reduced EDTA TE buffer (0.1 mM EDTA, 10 mM Tris HCl, pH 8.0). EDTA can inhibit some of the enzymic reactions in the GeneChip mapping array protocol, hence the lower concentration of EDTA. The DNA was allowed to dissolve overnight at 4°C. The concentration of DNA and quality was measured by a spectrophotometer (ND-1000, Nanodrop, Wilmington, DE, USA), as expressed by 260 nm/280 nm and 230 nm/260 nm ratios. Ideally, the 260/280 ratio would be 1.8 showing the least protein contamination, and the 230/260 ratio greater than 2 showing the least organic solvent contamination.

2.3.1 DNA extraction from organic fraction after RNA isolation using Trizol reagent

Samples that had previously had RNA extracted using the Trizol reagent (Invitrogen) had the organic fraction stored at -20 to -70°C. The sample was retrieved from the freezer and allowed to defrost at room temperature (RT). The vial was centrifuged at 2000 g for 15 mins at 4°C and any remaining aqueous top layer removed. Ethanol 100% was added at 0.3 ml/ml of the original volume of Trizol added. The solution was mixed and allowed to stand at RT for 3 mins before centrifuging at 2000 g for 5 mins at 4°C. A small pellet remained and the pink phenol-ethanol solution containing protein was removed carefully, and stored at -20°C. Sodium citrate 0.1 M in 10% ethanol was added at 1 ml/ml of Trizol. The solution was mixed and allowed to stand at RT for 30 mins before being spun as before. Again, the supernatant was removed (and discarded) leaving the pellet. The pellet was washed once more with sodium citrate and ethanol as above. Ethanol 75% at 2 ml was then added, and the solution mixed and allowed to stand for 20 mins at RT. After centrifuging again as above, the ethanol was carefully

removed and the pellet allowed to dry in air for 10 mins. The DNA was dissolved in 20-100 μg of reduced EDTA TE buffer overnight at 4°C and the concentration measured as above.

2.4 GeneChip array assay (Matsuzaki, *et al* 2004)

2.4.1 Preparation of reduced complexity samples

For each individual assayed, 250 ng of genomic DNA was digested with 10 U of Xba I (New England BioLabs). It was made up to 20 μL with 2 μl of 10x NEBuffer 2 (New England Biolabs) (NEBuffer 2 contains 500 mM NaCl, 100 mM Tris-HCl, 100 mM $MgCl_2$ and 10 mM dithiothreitol pH7.9 at 25°C), 2 μl 10x bovine serum albumin and molecular biology grade water. The mixture was prepared on ice in a pre-PCR designated laminar airflow hood to prevent contamination with DNA template, and DNA added in a PCR staging hood. The mixture was incubated for 2 h at 37°C. Following heat inactivation at 70°C for 20 min, 0.25 μM of adaptor (5'phosphate-CTAGAGATCAGGCGTCTGTCGTGCTCATAA-3', and 5'-ATTATGAGCACGACAGACGCCTGATCT-3' synthesised by QIAGEN) was ligated to the digested DNA with T4 DNA Ligase (New England BioLabs) in 25 μL for 2 h at 16°C. The ligation reaction was stopped by heating to 70°C for 20 min, and then diluted fourfold with water. For each sample, four PCRs were run using 10 μL of the diluted ligation reaction (25 ng of starting DNA) in 100 μL volumes containing 0.75 μM of primer (5'phosphate-CTAGAGATCAGGCGTCTGTCGTGCTCATAA-3'), 0.25 mM dNTPs, 2.5 mM $MgCl_2$, 10 U AmpliTaq Gold (Applied Biosystems), and PCR Buffer (Applied Biosystems). Thirty-five cycles of PCRs were done in a MJ DNA Engine Tetrad Cycler (MJ Research). The cycling program was 95°C denaturation for 20 sec, 59°C annealing for 15 sec, and 72°C extension for 15 sec. As a check, 3 μL of PCR products were visualized on 2% TBE agarose gels to confirm the size range of amplicons. However, it was soon clear that if sufficient quantity of DNA was obtained after purification, then this check was unnecessary.

The PCR products were purified using QIAquick spin columns (QIAGEN). 500 µl of Buffer PB was added to each 100 µl PCR sample, which was applied to the column set in a collection tube. The column was centrifuged at 13000 rpm at RT (the speed and temperature of all centrifugations for the QIA quick protocol) for 1 min and the flow through discarded. 750 µl of buffer PE was added to the column and centrifuged as before, discarding the flow through again. The column was centrifuged an additional 5 mins. The column was placed in a fresh 1.5ml tube and 55 µl of Buffer EB was applied to the centre of the membrane of the column. The column was incubated at room temperature for 5 mins and centrifuged for 2 mins. The eluate was then applied to the second column and allowed to stand as before for 5 mins before centrifugation for 2 mins again. This was repeated until the PCR products from all 4 columns were eluted. This protocol was found to be slow and inefficient at obtaining sufficient purified PCR product. Therefore, a new protocol was devised using Ultrafree-MC, PLTK Ultracel-PL Membrane, 30 kDa columns (Millipore, Billerica, MA). PCR products from the four reactions were combined and applied to the column and centrifuged for 15 mins at 2000 g at RT. The flow through was discarded, 250 µl of Ultrapure water was added and the tube centrifuged for 10 mins more. If more than 5 µl of liquid remained in the column, then the tube was centrifuged for 1 or 2 mins more until it reduced. 30 µl of EB buffer was applied to the membrane and allowed to stand for 1 min before being taken off into a clean 1.5ml tube.

PCR yields, based on absorbance readings at 260 nm, were typically 15-20 µg. If the yield was less than 12µg, it was found that the call rate subsequently obtained would be low (<85%), so one would not proceed with that sample. To allow efficient hybridization to the 25-mer oligonucleotides on the array, PCR amplicons were fragmented with DNAse I (Amersham Biosciences). Here, 0.24 U of DNAse I was added to 20 µg of purified PCR amplicons in a 55 µL volume containing 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, and 1 mM dithiothreitol for 30 min at 37°C, followed by heat inactivation at 95°C for 15 min. Fragmentation products were visualized on 2% TBE agarose gels. Occasionally, fragmentation would be inadequate and so the fragmentation was repeated, but this time only incubating for 5 mins before inactivation. The 3' ends of the fragmented amplicons were

biotinylated by adding 143 μM of a proprietary DNA labelling reagent (Affymetrix) using Terminal Deoxynucleotidyl Transferase (Promega) in a 70 μL volume containing 100 mM cacodylic acid (pH 6.8), 0.1 mM dithiothreitol, and 1 mM $CoCl_2$ for 2 h at 37°C, followed by heat inactivation at 95°C for 15 min.

2.4.2 Genotyping by Allele-Specific Hybridization

The fragmented and biotinylated PCR amplicons were combined with 11.5 μg/ml human Cot-1 (Invitrogen) and 115 μg/ml herring sperm (Promega) DNAs. The DNAs were added to a hybridization solution containing 2.69 M tetramethylamonium chloride (TMACl), 56 mM MES, 5% DMSO, 2.5 X Denhardt's solution, and 0.0115% Tween-20 in a final volume of 260 μL. The hybridization solution was heated to 95°C for 10 min, then placed on ice. Next, 200 μL of the hybridization solution was injected into cartridges housing the oligonucleotide arrays (Affymetrix GeneChip 10K Mapping Array version Xba131 or Xba142). Hybridization was carried out at 48°C for 16–18 h in a rotisserie rotating at 60 rpm. Following the overnight hybridization, the arrays were washed with 6X SSPE and 0.01% Tween-20 at 25°C, then more stringently washed with 0.6X SSPE and 0.01% Tween-20 at 45°C. Hybridization signals were generated in a three-step signal amplification process: 10 μg/ml streptavidin (Pierce) was added to the biotinylated targets hybridized to the oligonucleotide probes, and washed with 6X SSPE and 0.01% Tween-20 at 25°C, followed by the addition of 5 μg/ml biotinylated goat anti-streptavidin (Vector) to increase the effective number of biotin molecules on the target; finally, streptavidin R-phycoerythrin (SAPE) conjugate (Molecular Probes) was added and washed extensively with 6X SSPE and 0.01% Tween-20 at 30°C. The Streptavidin, Antibody, and SAPE were added to arrays in 6X SSPE, 1X Denhardt's solution, and 0.01% Tween-20 at 25°C for 10 min. The washing and staining procedures were performed using Affymetrix fluidics stations. Arrays were scanned using the GeneArray (Agilent) or GCS3000 (Affymetrix) scanners. Scan images were processed to get hybridization signal intensity values using either

Micro Array Suite (MAS) v5 software (Affymetrix) or GCOS v1 software (Affymetrix). The genotype-calling algorithm as previously described (Liu, *et al* 2003) was implemented in GenoTyping Tools (GTT; Affymetrix) and GDAS v2 (Affymetrix) analysis software. Default algorithm parameters, that is, a discrimination score cut-off of 0.08 and call zones of 0.8, were used to make all of the genotype calls.

2.4.3 Mutation analysis

Details of the PCR conditions and primers for mutation analysis AML samples is given in Table 2.1. Conditions are derived from previous publications for *CEBPA* (Snaddon, *et al* 2003) and specific exons of *FLT3* (exons 14-15 and 20) (Abu-Duhier, *et al* 2001, Kiyoi, *et al* 1999).

Table 2.1 Details of primers and reaction conditions for PCRs.

Exon	Primer sequence (5'-3')	PCR mixture	PCR Reaction Conditions					
			Initial denaturation	Denaturation	Annealing	Extension	No. of cycles	Final extension
FLT3								
14-15F	gcaatttaggtatgaaagccagc	100-200 ng of genomic DNA, 0.4 μM primers, 200 μM each dNTP, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$, 5% DMSO, 0.5 U Pic Taq (Cancer Research UK)	94°C 180 s	94°C 30 s	56°C 60 s	72°C 120 s	35	72°C 600 s
14-15R	ctttcagcattttgacggcaacc							
20F	ccaggaacgtgcttgtca	100-200 ng of genomic DNA, 0.4 μM primers, 200 μM each dNTP, 10x Buffer (ProMega, WI) diluted to 10 mM Tris-HCl (pH 9), 50 mM KCl, 1.5 mM $MgCl_2$, 1% Triton X-100, 0.5 U Pic Taq (Cancer Research UK)	94°C 180 s	94°C 60 s	56°C 30 s	72°C 120 s	35	72°C 600 s
20R	tcaaaaatgcaccacagtgag							

CEBPA								
PP1Fb	tcgccatgccgggagaactctaac	1.5 U Pic Taq (Cancer Research UK), 5% DMSO, 1x Buffer (Invitrogen), 2.5 mM $MgCl_2$, 200 µM each dNTP, 1 µM primers	95°C 120 s	95°C 60 s	62°C 40 s	72°C 60 s	40	72°C 600 s
CEBPA Ac	cctgctgccggctgtgctggaac							
PP4Fb	cttcaacgacgagttcctggccga							
PP4Rb	agctgcttggcttcatcctcct							
PP5Fb	ccgctggtgatcaagcagga							
PP5Rb	ccggtactcgttgctgttct							
CEBPA Cc	caaggccaagaagtcggtggaca							
PP6Rb	cacggtctgggcaagcctcgagat							
CCND3								
1F	ccctgcctgttcgctgcccga	50-100 ng/µl template, 0.4µM primers, 200µM each dNTP, 1x buffer (Promega), 1.25 U Pic Taq (Cancer Research UK)	96°C 120 s	96°C 30 s	58°C 30s	72°C 30 s	35	72°C 600 s
1R	tccagacccgggagcggtgg							
2F	gctccagacccagcagtga				55°C 30s			
2R	gggagacaatagctgtcggg				55°C 30s			
3F	ccactgaccccacccttte							
3R	gcaaaaatgtgcatagggctc				57°C 30s			
4F	ctcatggttccctttcctcct							
4R	atgaatggagaggctgctgc				52°C 30s			
5F	cactcttctcccatgttccca							
5R	actccagagggcctctccag							
PU.1								
1F	tcacccagggctcctgtagctca	0.5 U Pic Taq (Cancer Research UK), 5% DMSO, 1x Buffer (Invitrogen), 2.5 mM $MgCl_2$, 200 µM each dNTP, 0.4 µM primers	95°C 300 s	95°C 40 s	62°C 40 s	72°C 120 s	35	72°C 600 s
1R	tcgtgggcaggcaggcaggcgtcc							
2F	tgatggggaccagcgtgcggggt							
2R	tctctccagaccccaggaccaggc							
3F	actataacctttcctgccctgcc							
3R	agcctgtgtcagcttcctgtgaag							
4F	tgcactccttctctccccagctgacc							
4R	acacacacgcgactcggtggcgtg							
5F	ccgggcccctgtgcgtacgcaagg							
5R	ccgggagcgtcctccctgtgtccg							

RUNX1								
3F	gcctgtcctcccaccaccctctc	0.5 U Pic Taq (Cancer Research UK), 5% DMSO, 1x Buffer (Invitrogen), 2.5 mM $MgCl_2$, 200 µM each dNTP, 0.4 µM primers	94°C 120 s	94°C 30 s	60°C 30s	72°C 30 s	35	72°C 600 s
3R	agctgcttgctgaagatccg							
4F	gtgggtttgttgccatgaaacg				58°C 30s			
4R	catccctgatgtctgcatttgtcc							
5F	cccaaggaatctgagacatggtcc				58°C 30s			
5R	tgttcaggccaccaacctcattc							
WT1								
7F	gacctacgtgaatgttcacatg	50-100 ng/µl template, 0.4µM primers, 200 µM each dNTP, 1x buffer (Promega), 1.25 U Pic Taq (Cancer Research UK)	92°C 120 s	92°C 45 s	60°C 60s	72°C 90 s	30	72°C 600 s
7R	acaacacctggatcagacct							
8F	cctttaatgagatccccttttcc				58°C 60s			
8R	ggggaaatgtggggtgtttcc							
9F	tgcagacattgcaggcatggcagg				58°C 60s			
9R	gcactattccttctcaactgag							
10F	acttcactcgggccttgatag				50°C 60s			
10R	gtggagagtcatacttgaaag							

Chapter 3

Analysis of copy number and chromosomal abnormalities in AML by SNP array

The purpose of this chapter is to explore the correlation between karyotype and abnormalities detected using SNP arrays. The potential abnormalities that may be detected are changes in copy number i.e. gains and deletions, and loss of heterozygosity (LOH), which may include the former, but also include recombination events and gene conversion. Large events should be confirmed by conventional cytogenetic analysis but small events may be undetectable.

3.1 Materials and Methods

Fifty-seven AML patient diagnostic DNA samples were genotyped using the Affymetrix 10K SNP mapping array as described in Chapter 2 . Clinical characteristics of these AML samples are given in Table 3.1 and in detail in the Appendix. Where available, paired remission bone marrow was used for germ line comparison. Analysis was performed with GDAS software as previously described and Genome Orientated Laboratory File (GOLF) software (http://bioinformatics.cancerresearchuk.org/cazier01/Golf.html) to look for copy number changes and LOH. Conventional cytogenetics had been performed on these samples at diagnosis. The microarray data has been deposited in the National Cancer Biotechnology Information (NCBI) Gene Expression Omnibus (GEO)(2005) and are accessible through GEO Series accession number GSE7490. The numbering system for patient samples is described in

Appendix A, and relates to the numbering of samples in GEO series GSE7490.and GSE7210.

Table 3.1 Characteristics of the 72 AML patients genotyped.

Karyotype	**Number of patients**
Normal	41
t(8;21)	5
t(15;17)	4
inv 16	3
11q23 translocation	2
-7	3
+8	3
Others	11
FAB type	
M0	1
M1	24
M2	16
M3	4
M4	16
M5	10
M6	1
M7	0
Male: Female	39/33
Median age	50

3.2 Results

3.2.1 Analysis of data from the SNP arrays

3.2.1.1 Use of LOH scores from GDAS software

The array uses hybridisation signals from 5 perfectly matched 25bp oligonucleotides, each displaced by 1 nucleotide, and 5 mismatched 25bp oligonucleotides. This is repeated for the alternate allele. The system also uses the antisense strand, so in total there are 40

oligonucleotides for each SNP. The allelotype is calculated from the relative allele signal from the sense and antisense strands.

The algorithm used is called Modified Partitioning Around Medoids (MPAM). Initially a discrimination score (DS) is calculated for each pair of SNP alleles:

$$DS = (P-M)/(P+M)$$

Where P= perfect match intensity

M= mismatch intensity

The median DS for each allele A and B, and for sense and antisense is calculated, and then the DS for the SNP is derived. If this is high enough, the SNP is passed and the relative allele score (RAS) is calculated. If not, the SNP is assigned a no call. The RAS is calculated from the average mismatch, $\overline{M}$

$$\overline{M} = (MA - MB)/2$$

Where MA= mismatch allele A intensity

MB= mismatch allele B intensity

Then relative intensity values for each allele A and B are calculated

$$A = \mathrm{Max}(PA - \overline{M}, 0)$$

$$B = \mathrm{Max}(PB - \overline{M}, 0)$$

Where PA= Perfect match A intensity

PB= Perfect match PB intensity

Finally, an overall RAS value is given by

$$RAS = A/(A+B)$$

This is calculated for the sense and anti sense strands, RAS1 and RAS2 respectively, and plotted on a graph of RAS1 against RAS2. Figure 3.1 shows an example where two heterozygous SNPs are plotted. The medoid values (shown by the ovals indicating when a call is made) were determined from a set of 100 individuals of Caucasian, African-American and East Asian origin. If the RAS values fall within the appropriate medoid, then the appropriate call is made, if not then the SNP is assigned a no call.

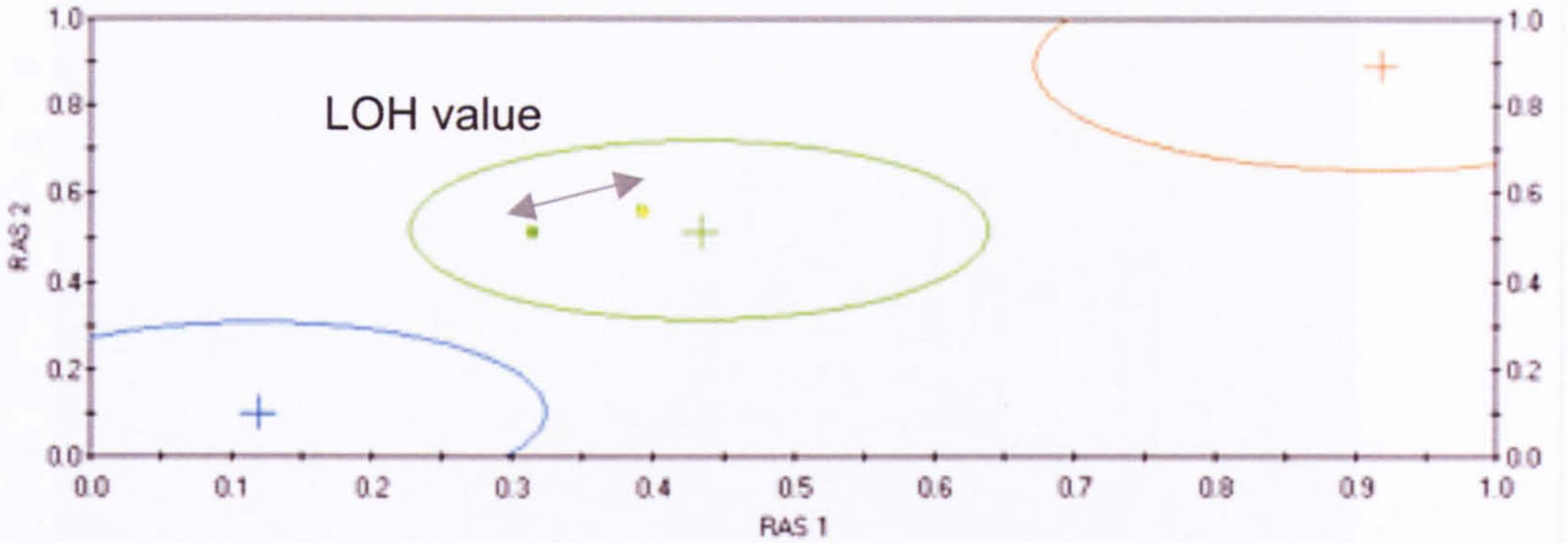


Figure 3.1 MPAM mapping analysis window.

An example of 2 heterozygous calls is shown, marked with a green and yellow point. The LOH value is derived from the distance between the two points. The centres of the medoids are marked with crosses.

LOH can be determined by a change from heterozygous to homozygous between the germ line and leukaemia sample. This implies that there is a change in RAS values for the value to fall in a different medoid. The distance in RAS values between the two SNPs may be used as a surrogate for LOH, i.e. the difference of two squares between the two points.

Figure 3.2 shows an example (patient sample 10) in which there is a deletion of the long arm of chromosome 7 (full karyotype 46,XX,dic(7;22)(q11.2;q10), +8[10]). The values for non-informative homozygous SNPs do not change, so their LOH value is low, less than 0.2. Heterozygous SNPs that do not change, across the short arm of chromosome 7 also have low LOH values. However, across the deleted region, heterozygous SNPs have LOH values between 0.3 and 1. There is variation in LOH values between individual SNPs, but taken as a whole they show the region of deletion has LOH.

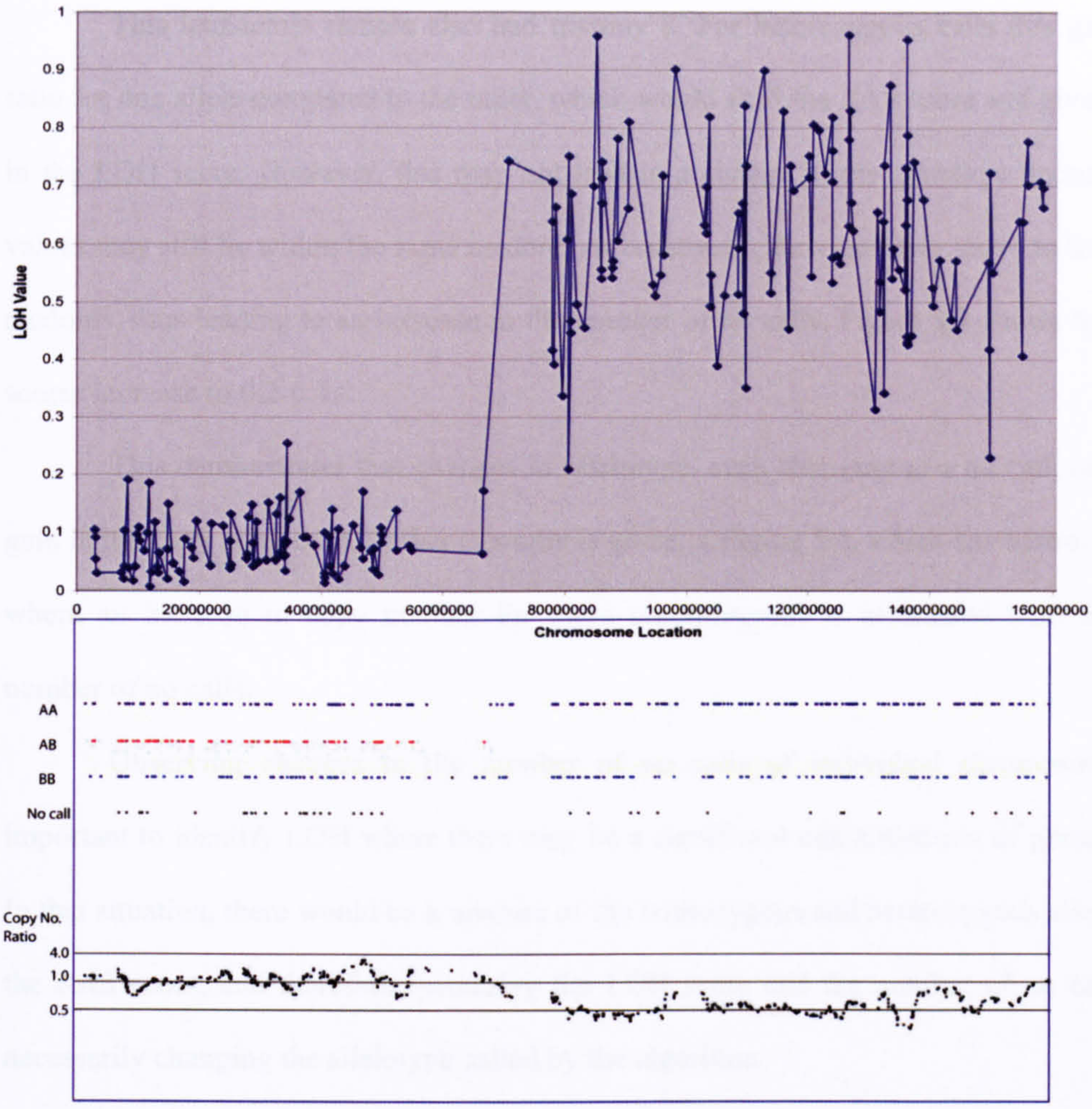


Figure 3.2 Increased LOH values across the long arm of chromosome 7, due to deletion (patient sample 10).

The lower diagram shows SNP allele calls (arbitrarily called A or B) along the chromosome in the leukaemia sample. Blue points represent homozygous calls and red points heterozygous calls (AA on the top line, BB on the bottom line and AB on the middle line. Grey points below the BB line are no calls). Increased LOH values between 0.5 and 1.0 are seen along the region of homozygosity (upper graph). Only informative calls are shown on the upper graph i.e. those heterozygous in the germline. Failed SNPs are also eliminated from this graph. The copy number ratio between the leukaemia and germline is shown on the lower graph.

This leukaemia sample also had trisomy 8. For heterozygous calls this gave a 2 to 1 ratio for one allele compared to the other, which would shift the RAS score and give an increase in the LOH score. However, this may not lead to a change in the genotype because the RAS values may still lie within the same medoid. Alternatively, the value may move to lie in between medoids, thus leading to an increase in the number of no calls. Figure 3.3 shows that the LOH scores increase to 0.2-0.4.

This demonstrates that changes in allelotype, even changing to a no call, can be due to gain in the copy number. A further example is given in Figure 3.4, which has trisomy 11 and 13, where an increase in copy number for these chromosomes is associated with an increased number of no calls.

Observing changes in the number of no calls of individual chromosomes is also important to identify LOH where there may be a significant contamination of germ line tissue. In that situation, there would be a mixture of the homozygous and heterozygous alleles, altering the RAS score, and therefore increasing the LOH score and the number of no calls, but not necessarily changing the allelotype called by the algorithm.

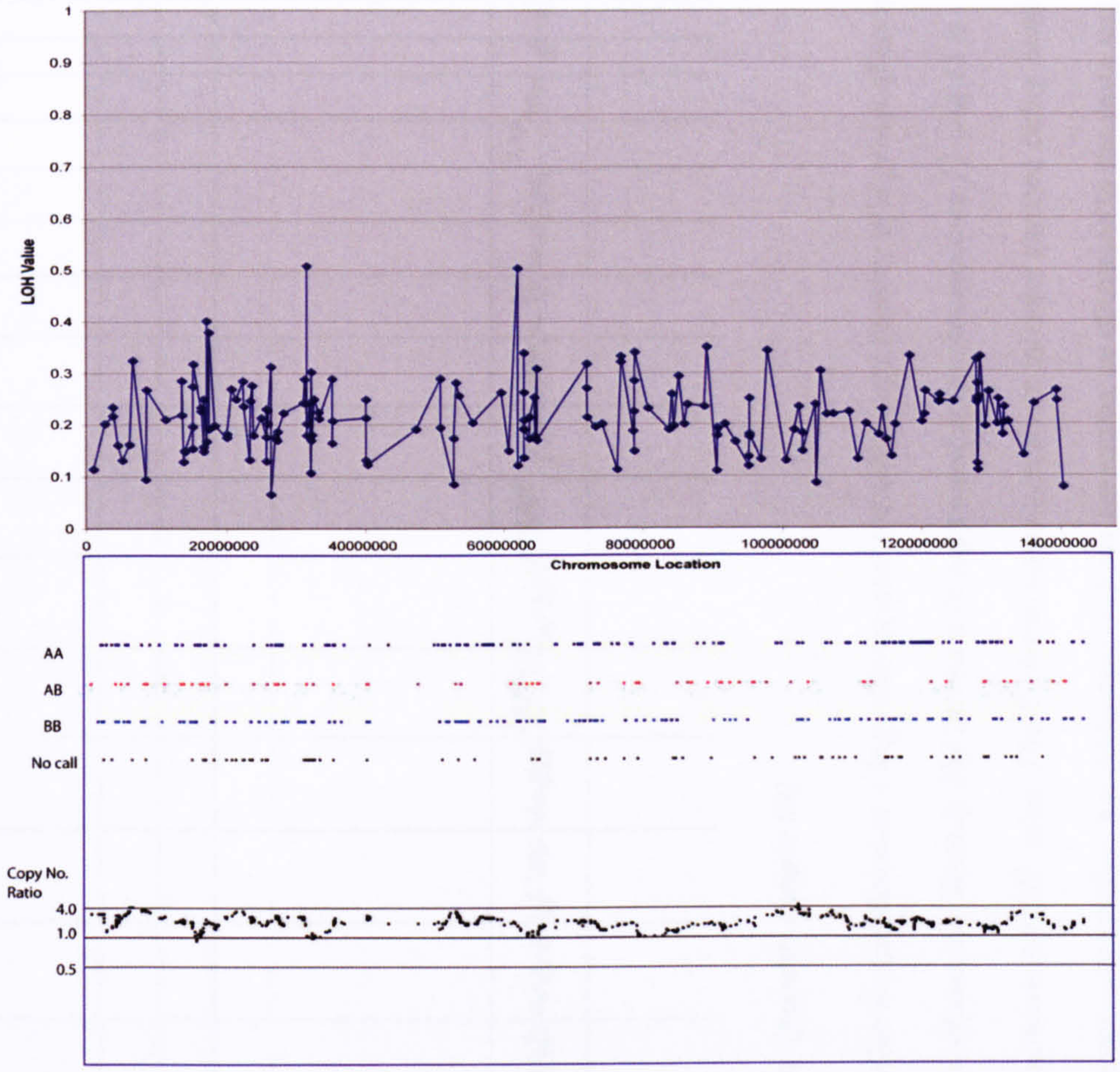


Figure 3.3 Moderately increased LOH values across chromosome 8 due to trisomy 8 (patient sample 10). Note the LOH values are less than 0.5, which is lower than the LOH values obtained from the deleted chromosome arm in Figure 3.2.

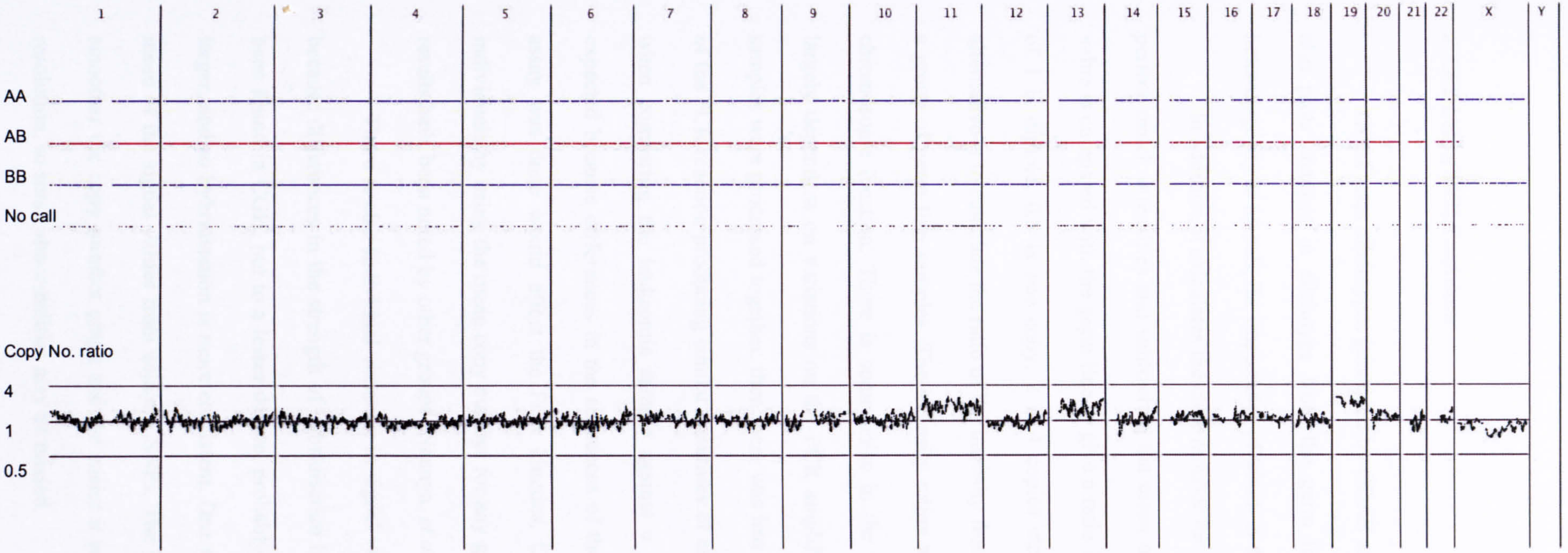


Figure 3.4 Allelotype compared with copy number in a leukaemia (patient sample 69)

The karyotype of the leukaemia is 48,XY,+11,+13. The copy number is calculated as a log ratio against the mean copy number of a group of normal human DNAs. The middle line of the copy number graph is a ratio of one, i.e. diploid. The increase in copy number on chromosome 11 and 13 is associated with a higher number of no calls, consistent with an increase in LOH value. The X chromosome acts as an internal control, being homozygous and having a lower copy number. Chromosome 19 gives an aberrant copy number, not consistent with the karyotype, as discussed in the main text.

3.2.1.2 Use of GOLF software

GOLF uses allelotypes generated by GDAS and plots them by chromosome location. It also plots changes in allelotype from the germ line, indicating LOH, and changes from heterozygosity to no call, the importance of which is described above.

In addition, it calculates the copy number for an individual SNP, taking the mean of the perfect match intensities and subtracting the mean of the mismatch intensities. This absolute value is compared with the germ line to give a ratio indicating the actual copy number (a ratio of 1 is diploid, 0.5 is one copy, 2 is 4 copies etc.). If no germline is available, then the alternative is to calculate the ratio of the intensity from the sample to the average intensity from a group of germ line samples. The intensity ratios are then plotted on a logarithmic scale by chromosome location. There is some noise in the measurement of copy number, which is largely dependent on variations on the PCR amplification step. If germ line and leukaemia samples were processed together, then there was less noise, presumably due to similar kinetics of the PCR reaction producing similar quantities of each DNA fragment. There was more noise when comparing the leukaemia sample against a group of normal DNAs; this would be expected because differences in the sequences of the genome and the occasion on which the assay was done would affect the PCR kinetics. Combining the results of several normal individuals by using the mean copy number for any given SNP would reduce the noise. Similar results have been noted by other groups (Nannya, *et al* 2005).

There is also individual variation in signal value between adjacent SNPs. This may be because differences in the strength of hybridisation between different oligomers. This has also been found in CGH, but to a lesser degree, probably because the sizes of the BACs are much larger, and so hybridisation is more consistent. One method to reduce the variation is to take a mean of the signal values from adjacent SNPs. The larger number of adjacent SNPs used, the smoother the copy number graph and the easier it is to interpret. However, this method loses resolution, so small abnormalities may be missed.

GOLF also gives a broad indication of regions of homozygosity across the genome. It looks for regions of homozygosity defined by the number of contiguous homozygous SNPs, and compares the copy number of this region to the average across the rest the genome. This facility can produce a map of LOH, as was used in Chapter 5 (Figure 5.1, Figure 5.2 and Figure 5.3).

3.2.2 Analysis of numerical chromosomal abnormalities

Large regions of LOH that correlated with chromosome arm deletions and a decrease in copy number could be inferred from the relative allele score (RAS) of the SNPs in that region (Liu, *et al* 2003). Thus, in Figure 3.2, deletion of chromosome 7 can be seen from homozygosity for the whole of the chromosome and a reduction in copy number as compared with control DNA.

Gains were observed as a change in copy number as well a change in allelotype. Figure 3.3 shows an example in which there is trisomy 8. The copy number is increased along the whole length of the chromosome and there is a loss of heterozygous calls to no calls. If trisomic chromosomes consisted of 3 copies of the same parental chromosome then the chromosome would be completely homozygous. However, this result demonstrates that one of the parental chromosomes has been duplicated, and the other retained. For heterozygous calls, there is therefore a 2:1 ratio of alleles. Depending on the signal intensity this led to retention of heterozygosity, a no call or homozygosity.

Most of the samples had a clear correlation between the copy number and karyotypic changes. AMLs with known balanced translocations had no evidence of a change in copy number at the region of translocation confirming there was no loss or gain of genomic material at the break point, at least at this resolution. Most normal karyotype AMLs showed no change in copy number and no large regions of LOH. Some of the AMLs did show large regions of homozygosity without a change in copy number, but these shall be discussed in Chapter 4 .

Those AMLs that had chromosomal deletions had associated regions of LOH and a fall in copy number, as is detailed in Table 3.3. Only one case (patient sample 70) who had the

karyotype 45,XX,t(6,9)(q22.3;q34)/45,XX,t(6,9)(q22.3;q34),del(7)(q35) showed no evidence of a change in chromosome 7, despite having a blast percentage of 94%. Cytogenetic analysis relies on the analysis of the clone that will grow in culture for 24 hours, so it is likely that the majority clone did not have a deletion of chromosome 7 and did not grow, explaining its absence on the SNP array. Many karyotypic changes were not as simple as delineated in the karyotype. For example, GSM173409 had monosomy 9 by cytogenetic analysis, but only the long arm was deleted on SNP array, with the telomeric end of the chromosome also remaining (Table 3.3). Similarly, patient sample 34 had monosomy 7, but retained the proximal part of the long arm of chromosome 7 on the array. These results suggest that apparent whole chromosome deletions have more complex rearrangements than is revealed by conventional cytogenetic analysis. In a similar vein, patient sample sample 31 had a complicated rearrangement of chromosome 9, but from the array, the result was a deletion with LOH of a small part of the short arm of chromosome 9. The region is delineated by the heterozygous SNPs rs2383159 (at chromosome base pair location 20881855) and rs1758734 (base pair location 25253918). Amongst several genes within this region is *CDKN2A* or *p16*, and *CDKN2B* or *p14*, both TSGs involved the cell cycle, and both have been described to be epigenetically suppressed in AML (Christiansen, *et al* 2003, Linggi, *et al* 2002).

Several deletions were not identified by conventional cytogenetics at all. Two small interstitial deletions were observed at 5p (base pair 15872615-23071052) in patient sample 78 and 7q (base pair 140586305-143189876) in patient sample 23 (see Table 3.3). The latter involves the T cell receptor (TCR) beta variable chain, so may be due to homologous recombination. TCR gene rearrangements are known to occur in some subtypes of AML (Chapiro, *et al* 2006, Dupret, *et al* 2005). Gain of 1p was a recurrent abnormality detected by SNP array but not identified by karyotypic analysis (patient samples 16, 37, 38 and 78). The cytogenetics literature lists 87 examples of add(1)(p?) (National Cancer Institute 2006), suggesting it could be a common abnormality, but not easily identified by cytogenetic analysis.

Abnormalities of chromosome 19 were difficult to interpret. Sample 53 gives an example of trisomy 19 with an increase in copy number. However, GSM173413 shows an

anomalous fall in copy number associated with trisomy 19. Several other AML samples show apparent gain or deletion in the context of diploid chromosome 19 on cytogenetic analysis. Problems with the interpretation of copy number of chromosome 19 have previously been noted with CGH (Kallioniemi, *et al* 1994). The complete sequence of chromosome 19 shows it is the densest number of genes of any chromosome with large numbers of repetitive sequences (Grimwood, *et al* 2004). On the array, chromosome 19 has the lowest density of SNPs. Together, these problems mean interpretation of copy number must made with caution. Similar problems can occur with the smaller chromosomes, particularly 21 and 22. A smaller number of SNPs means that if the copy number of only a few of them is raised or decreased due to noise, the mean copy number of the chromosome is altered, and the copy number may be incorrectly interpreted as amplified or deleted.

3.2.2.1 Single SNP changes

Several cases, when compared to their remission sample, had single SNP loci that had LOH, and some also appeared to gain heterozygosity. It would be valuable to find genuine small regions of LOH, as they would give the location of putative oncogenes. However, gains of heterozygosity would be difficult to explain; it would suggest the nucleotide had mutated. The rate of reproducibility of allelotypes on the array is 99.96% (Affymetrix 2003), which would mean one may expect 4 or more SNPs to be discrepant due to noise on any given array. Repeating the assay for two samples (N and M) showed a different result for some of these discrepant SNPs i.e. homozygosity instead of heterozygosity, implying one of the results was wrong.

Table 3.2 Single SNP changes. The Affymetrix (Affy) array genotypes were assayed again by Pyrosequencing (Pyro).

Paired patient samples		SNP ID	Germ Line / Remission		Leukaemia	
Remission	Diagnosis		Affy	Pyro	Affy	Pyro
44	10	rs1944130	C/T	C/C	C/C	C/C
28	27	rs910952	A/A, G/A*	A/A	A/A	A/A
89	16	rs910950	G/A	A/A	A/A	A/A
95	9	rs1977697	A/G	G/G	G/G	G/G
89	16	rs1977697	A/G	G/G	G/G	G/G
105	14	rs1977697	A/G	G/G	G/G	G/G
96	32	rs2123077	T/C	T/C	T/T, T/C*	T/C
89	16	rs278341	T/C	T/C	T/T	T/C
28	27	rs864179	T/C, C/C*	C/C	T/C, C/C*	C/C
96	32	rs864179	T/C	C/C	T/C, C/C*	C/C
105	14	rs864179	T/C	C/C	C/C	C/C
41	40	rs278341	T/C	C/C	C/C	C/C

*Repeated array results

In order to confirm (or refute) these genotypes, Pyrosequencing was performed (Table 3.2). The technology used in Pyrosequencing relies on the release of a pyrophosphate molecule when a dinucleotide triphosphate (dNTP) is incorporated into the DNA sequence by DNA polymerase. The dNTPs are added to the reaction one at a time, so it is clear which particular nucleotide (A, T, C or G) is incorporated. The pyrophosphate is converted to adenine triphosphate (ATP) by ATP sulfurylase, and ATP produces fluorescence by being the substrate for luciferase. The fluorescence is proportional to the amount of nucleotide incorporated. Sequencing this way is a more accurate way of analysing the SNP change than the Affymetrix

hybridisation technology. In most cases the discrepant Affymetrix result was heterozygous and all were transversions rather than transitions. I conclude that single SNP changes cannot be relied upon to determine LOH.

There are a number of explanations for the failure of individual SNPs to give the correct genotype. There were a few situations where an array with a good call rate (>95%) had only 3-4 SNPs that were not concordant. If the tertiary structure of a particular sequence has more likelihood of being degraded by the DNAase at the fragmentation step then this would change the balance of transcripts. The hybridisation process may have problems with non-specific binding, as it is at a relatively low temperature (47°C), and the washing of the array is done at a much lower temperature i.e. room temperature. It needs these values so that many probes with different optimal hybridisation temperatures may bind specifically. However, it may be at the expense of occasional non-specific binding.

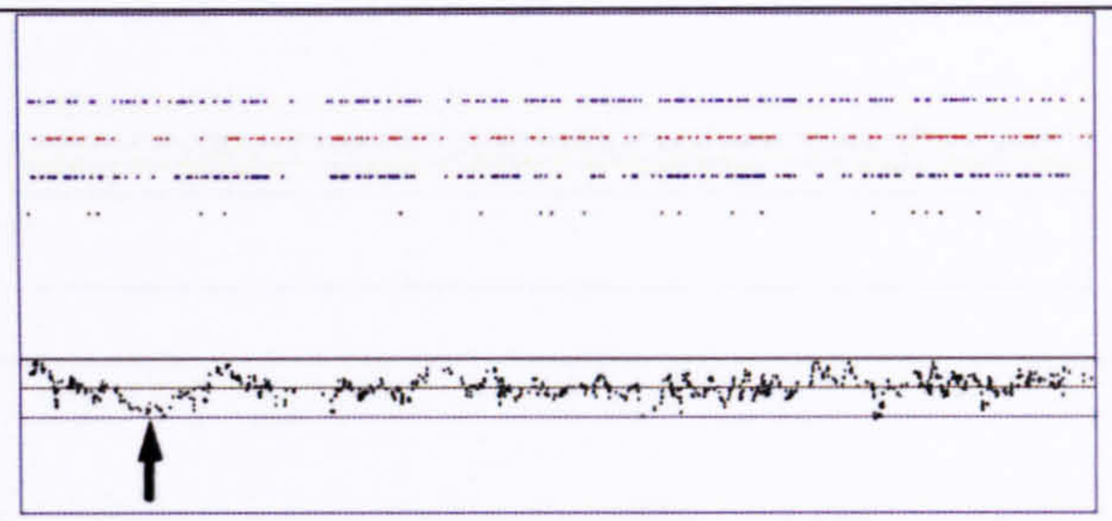
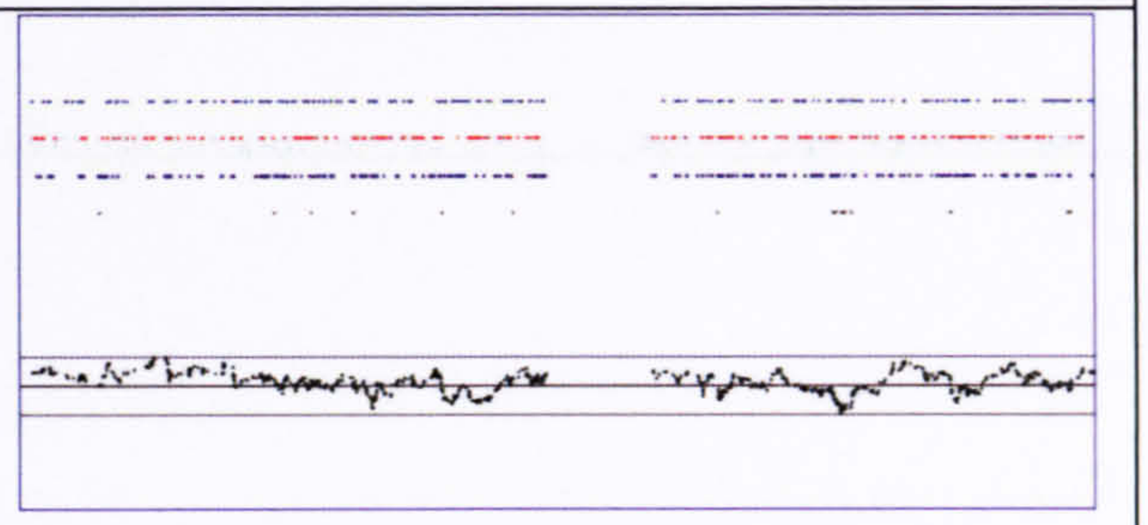
3.3 Summary and conclusions

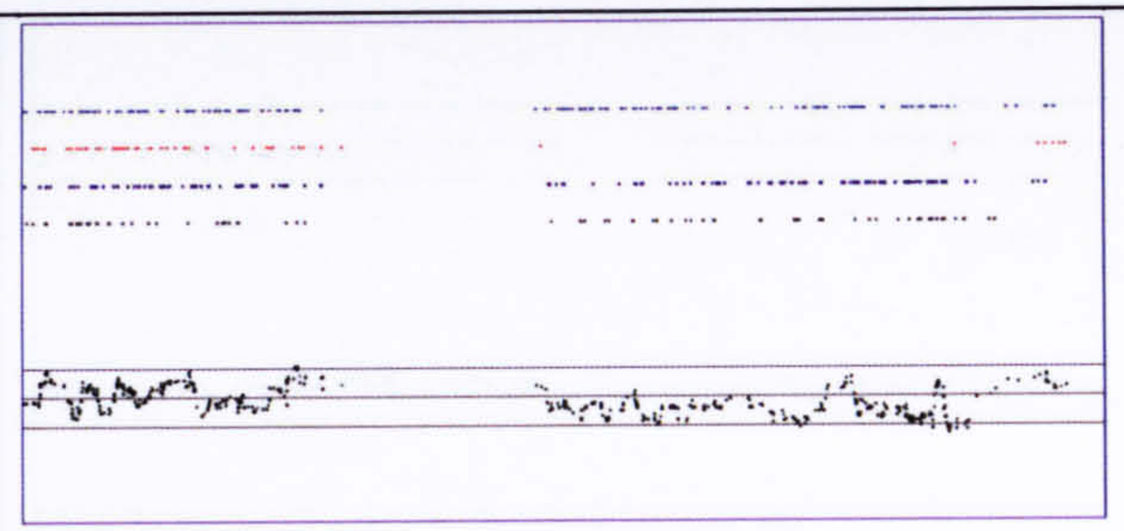
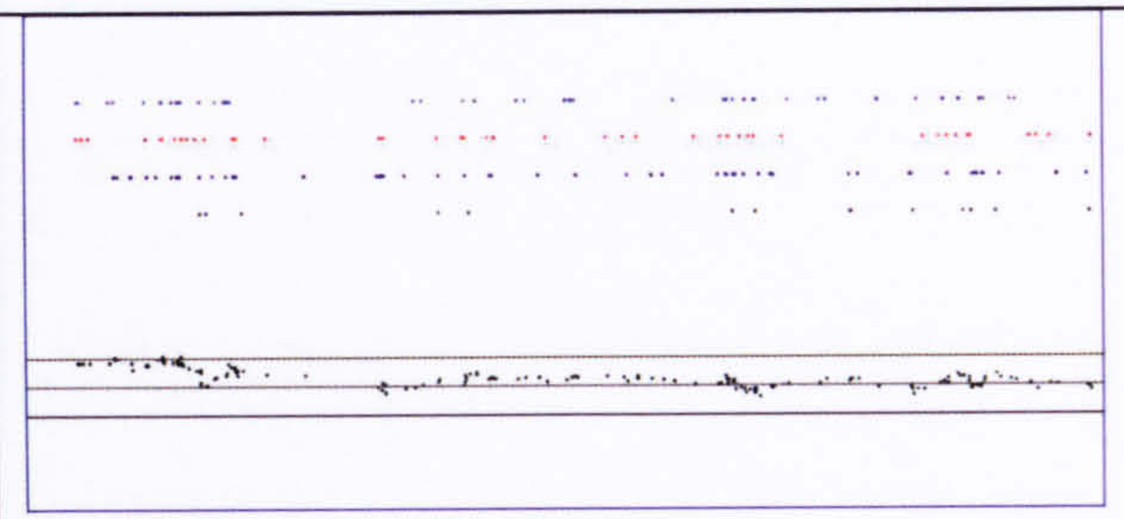
The Affymetrix 10K SNP arrays are able to provide accurate genotype information that is able to detect regions of LOH that are only a few megabases in length. The copy number data is consistent with conventional karyotypic analysis, but because of the improvement in resolution also provides additional data on the breakpoints of deletions, and clarifies interstitial deletions and complex rearrangements. By understanding the how the technology works, the change in allelotypes, from heterozygosity to homozygosity or no call can be used with copy number information to interpret genetic aberrations in the leukaemia.

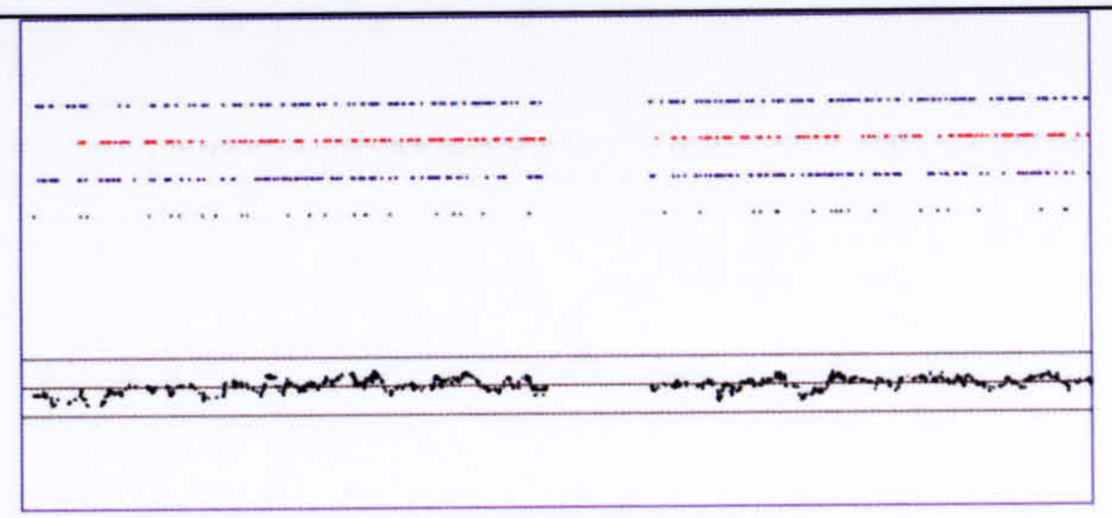
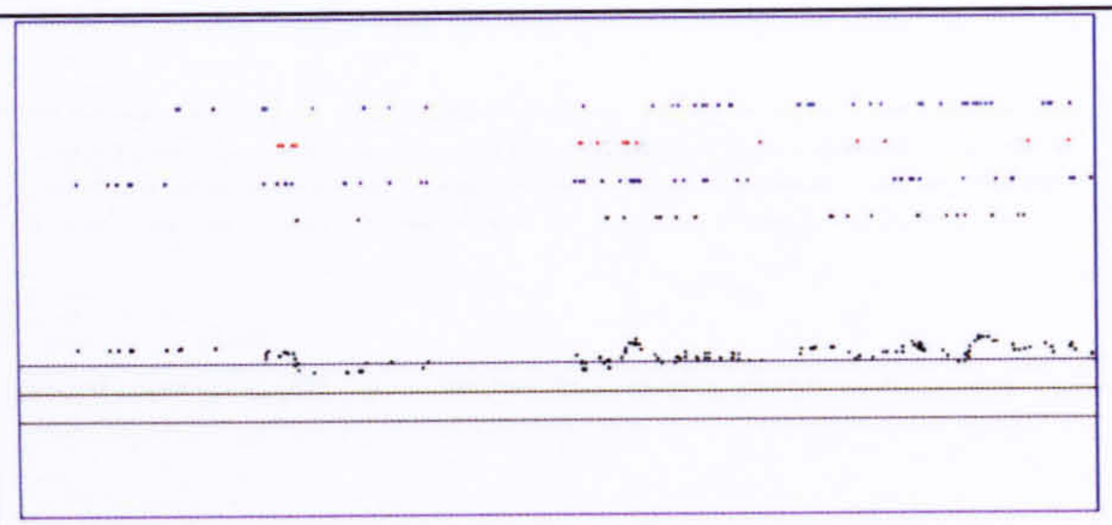
Many of the limitations in interpreting the data are due to the relatively low resolution of the 10K array, compared to current systems using hundreds of thousands of SNPs. Arrays of this size can identify genes within regions of only a few hundred kilobases (Mullighan, *et al* 2007). However, one of the main advantage of the SNP array is to identify regions of LOH in the absence copy number change, due to somatic recombination events, and these shall be discussed in the next chapter.

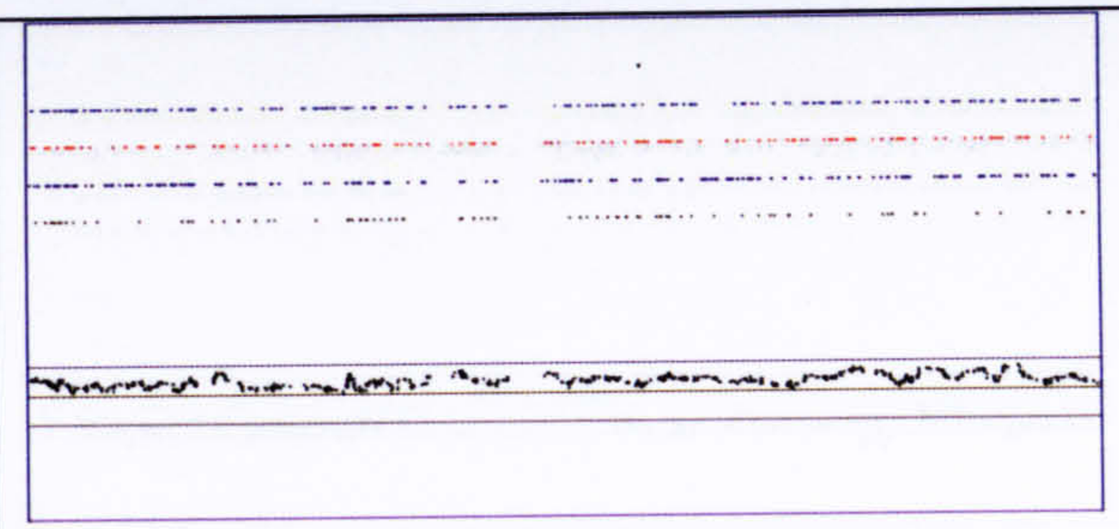
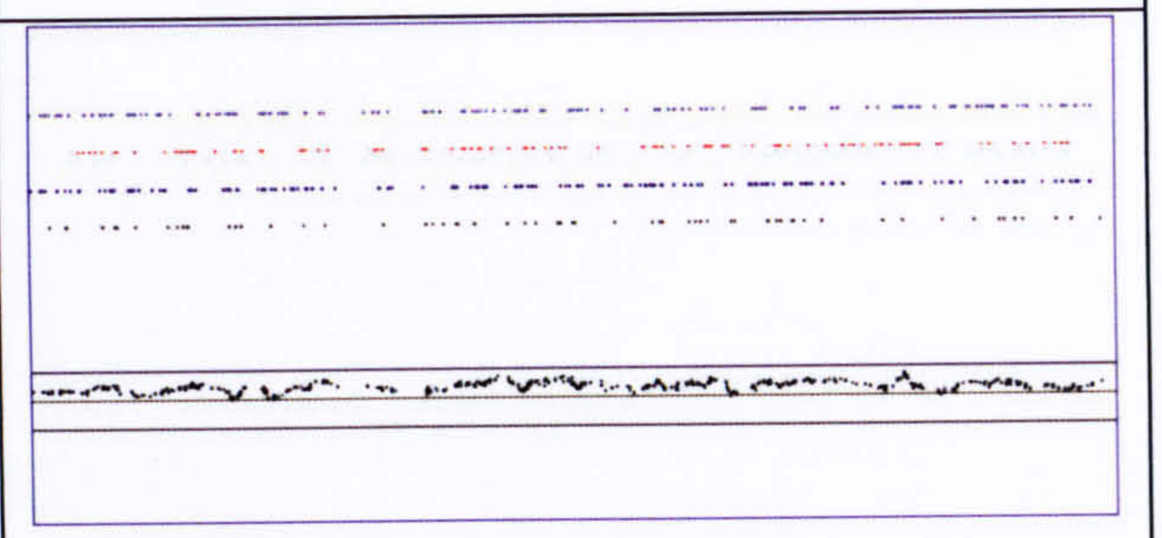
Table 3.3 Copy number changes in other leukaemias.

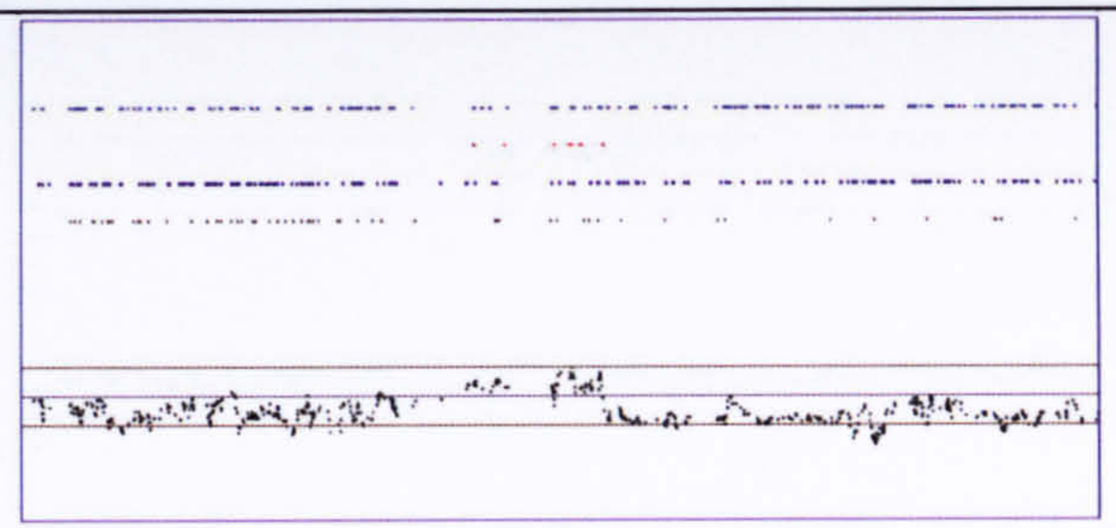
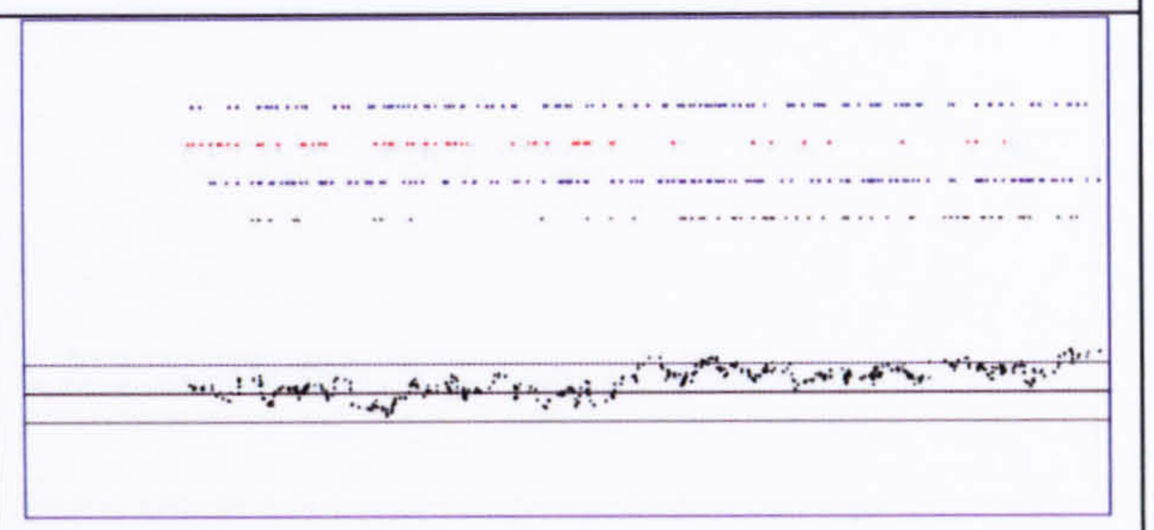
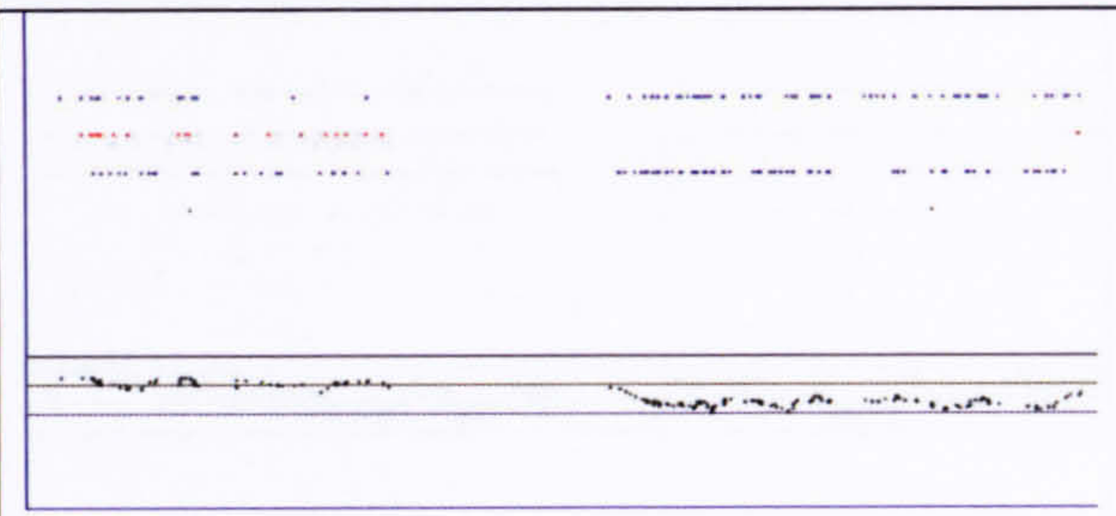
The copy number graph is a ratio against germ line DNA derived from remission bone marrow, where available, or against the mean copy number of a group of normal human DNAs. A running window of a mean of 10 SNPs or fewer was used in plotting the copy number graph to reduce variation in individual SNP copy number.

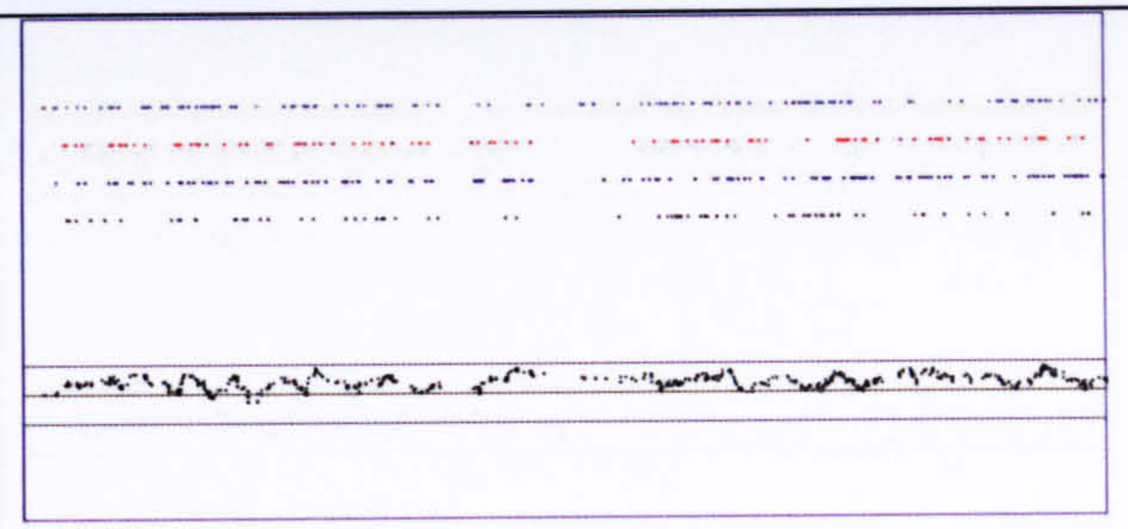
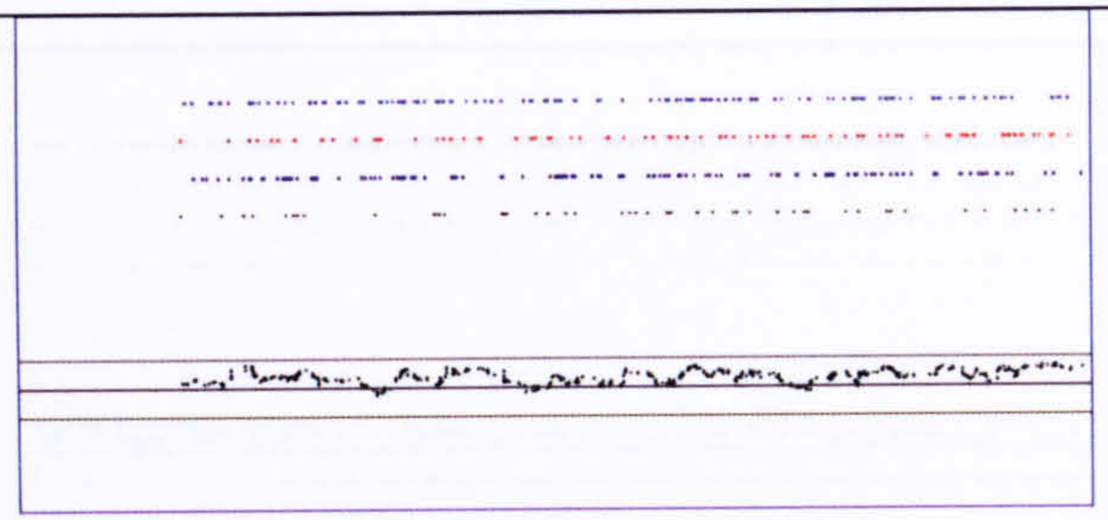
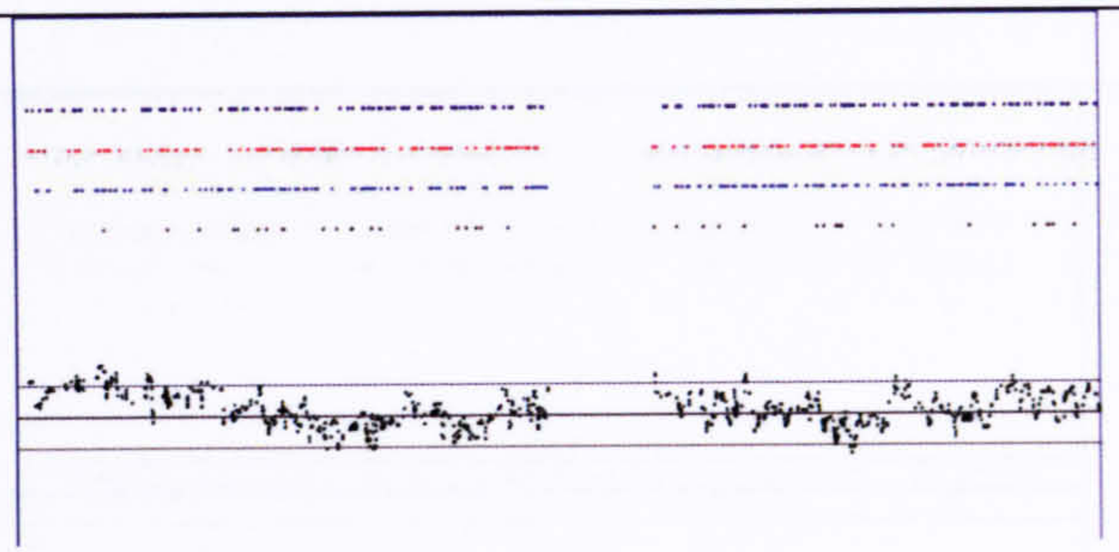
Patient sample	Karyotype	Copy Number Changes	Chromosome LOH and copy number map	Control
78	46,XY	5p interstitial deletion		Normal DNA
		1p gain		

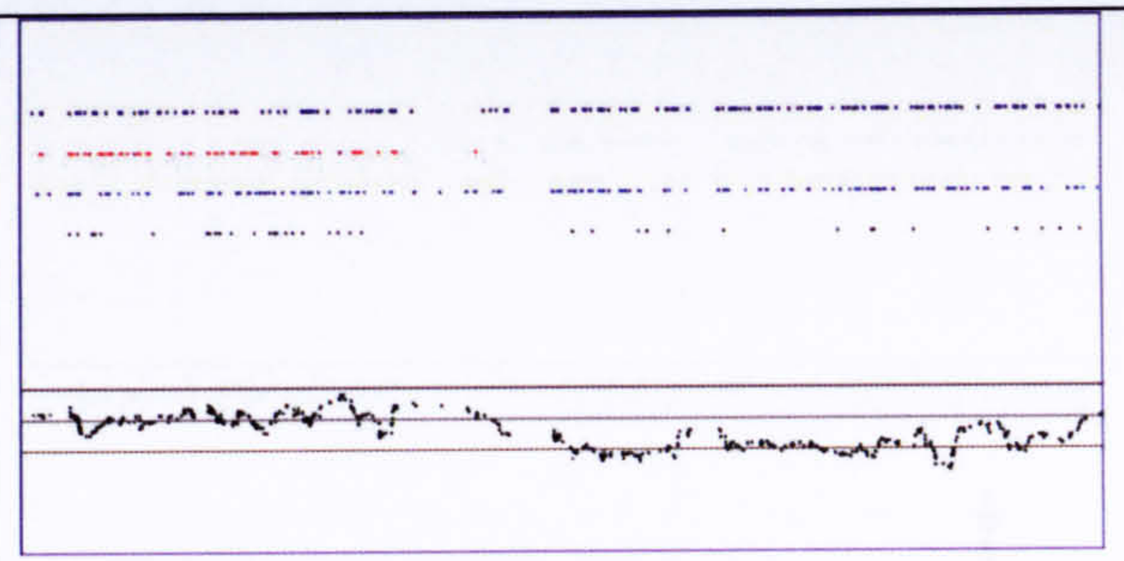
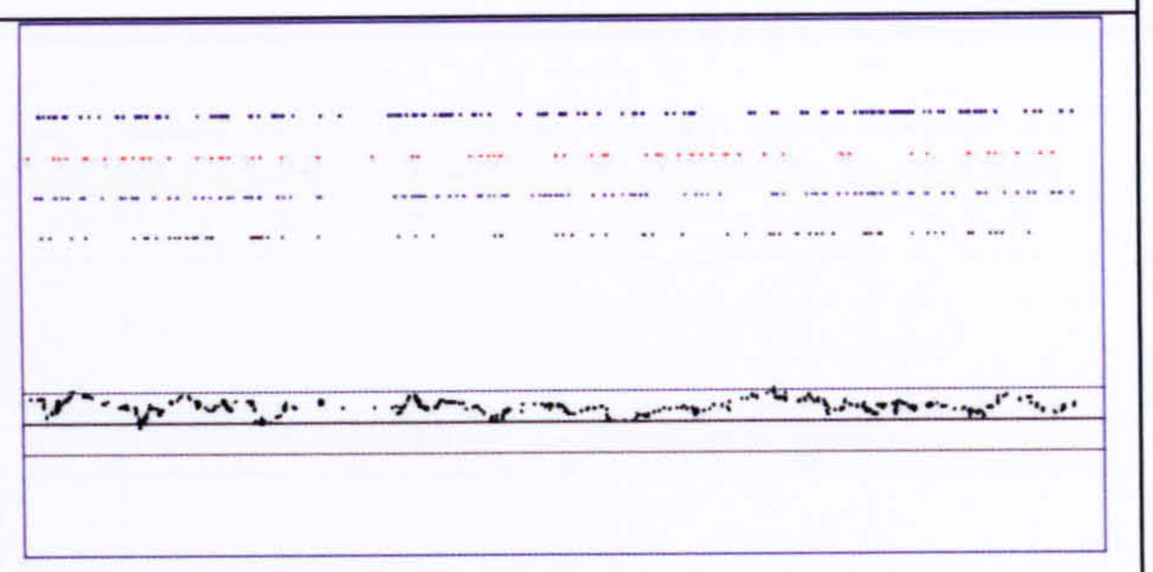
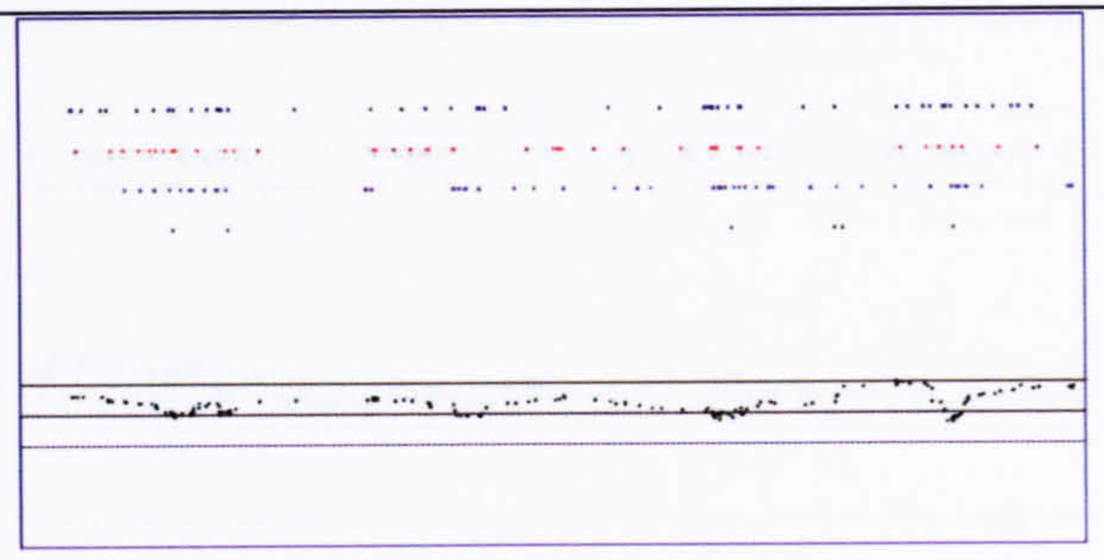
31	46,XY,del(9)(q13q34),del(12)(p13),add(17)(p13)[7] /46,XY,add(7)(p15),del(9)(q13q34),del(12)(p13)	9q deletion except telomere		Normal DNA
		17p gain		

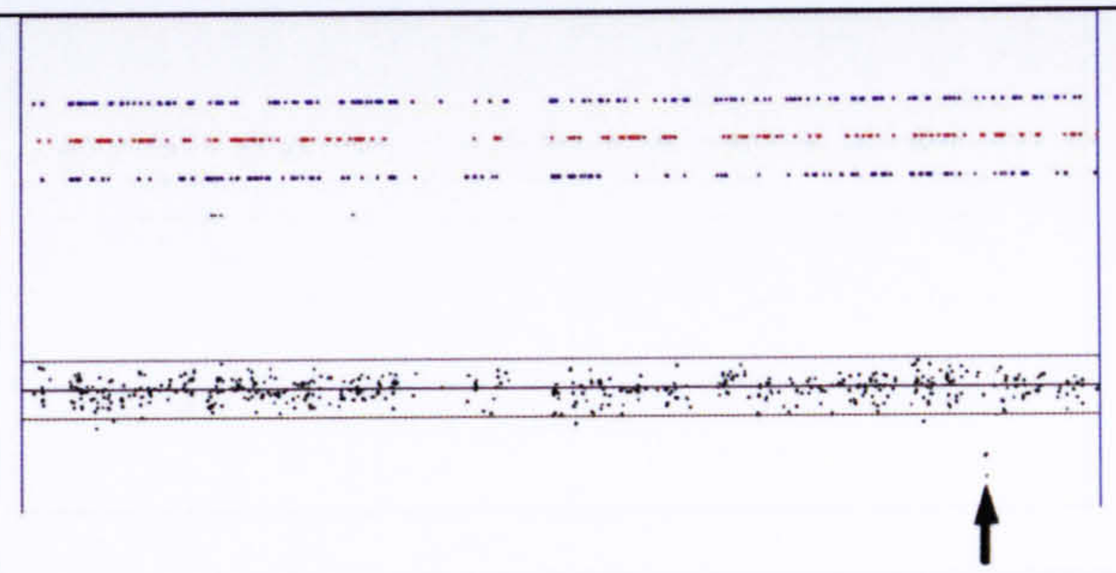
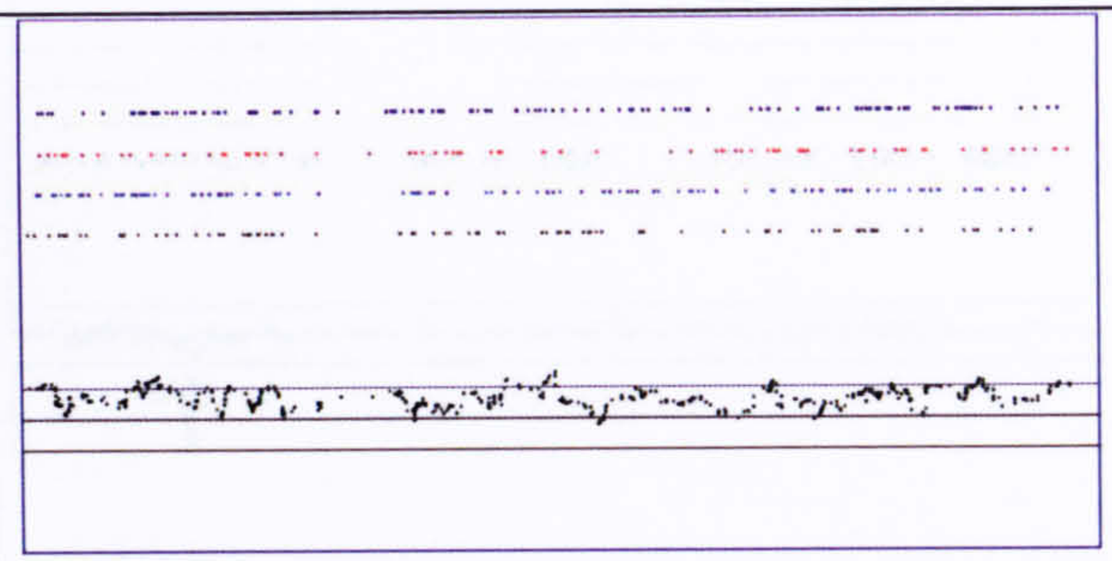
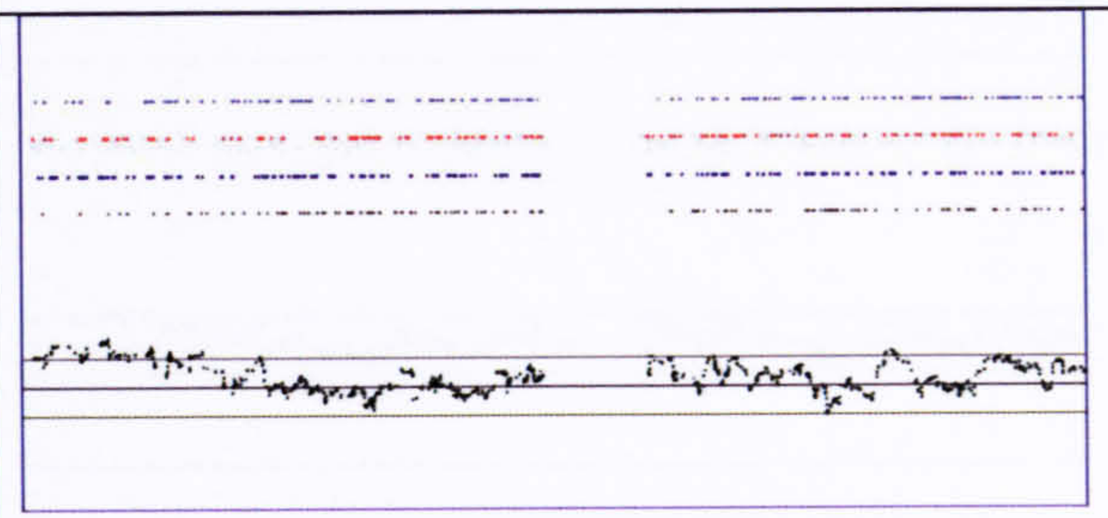
53	46,XY, +19 and del 1 x *ABL* homologue	1p deletion		Germ line
		19 gain		

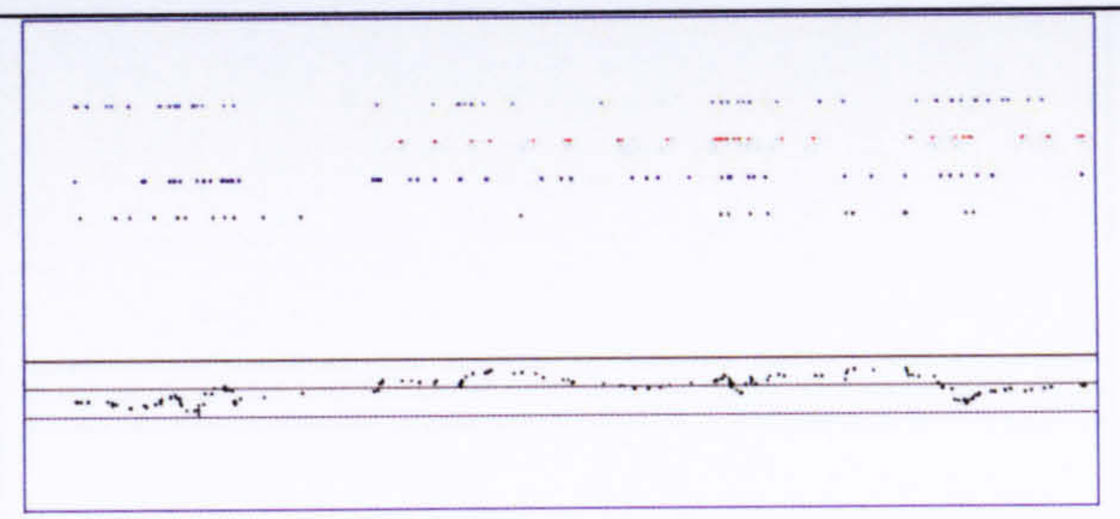
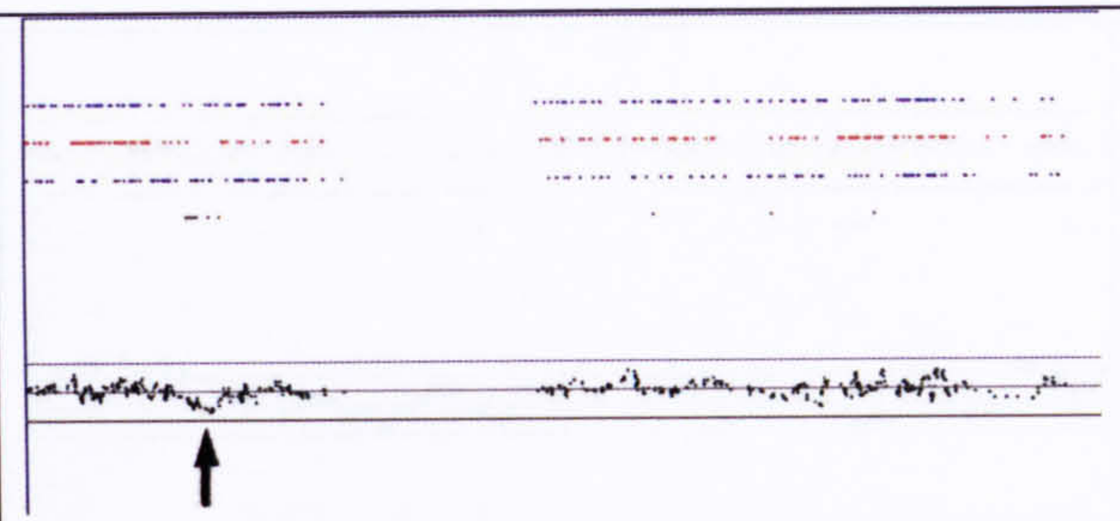
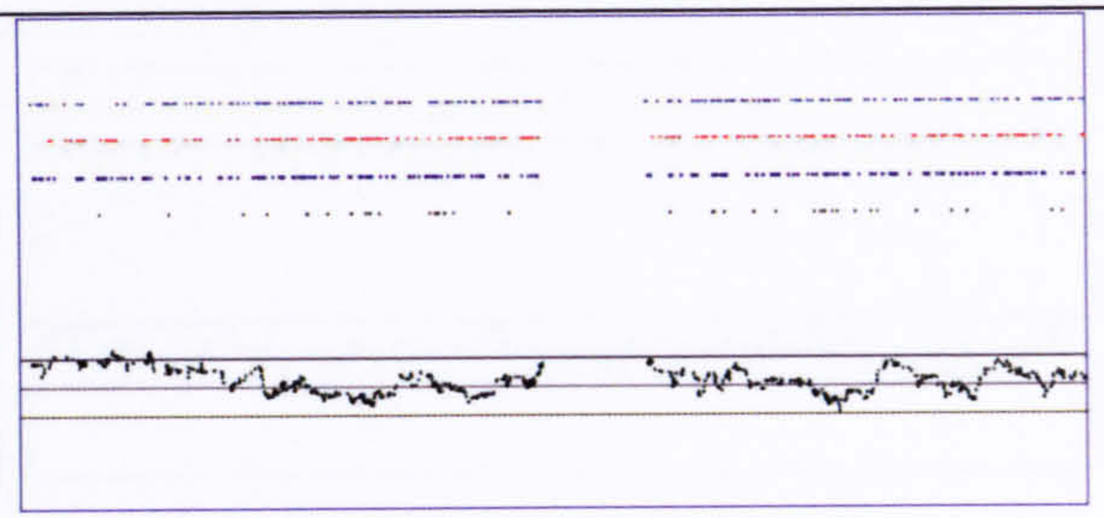
69	48,XY,+3,+10	3 gain		Germ line
		10 gain		

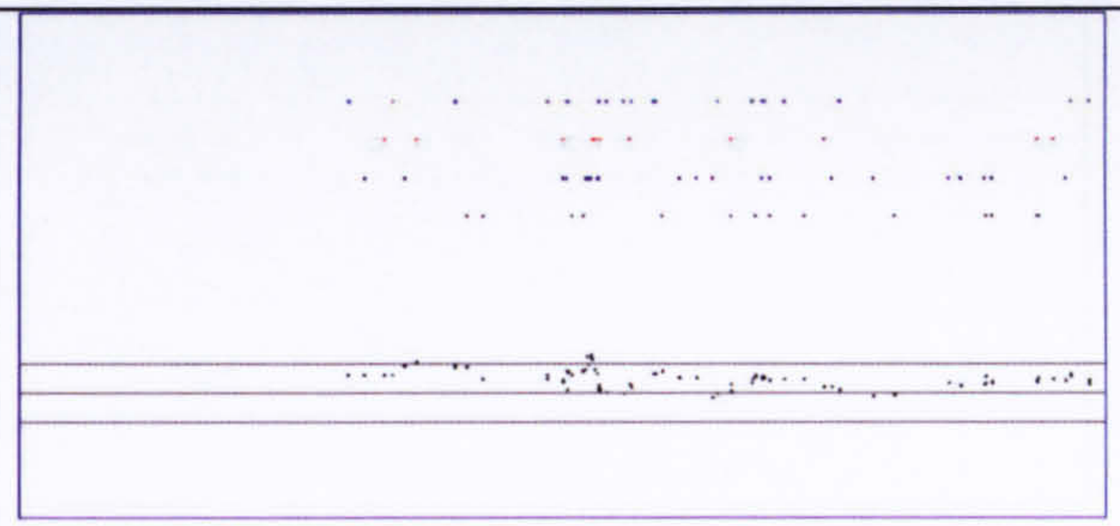
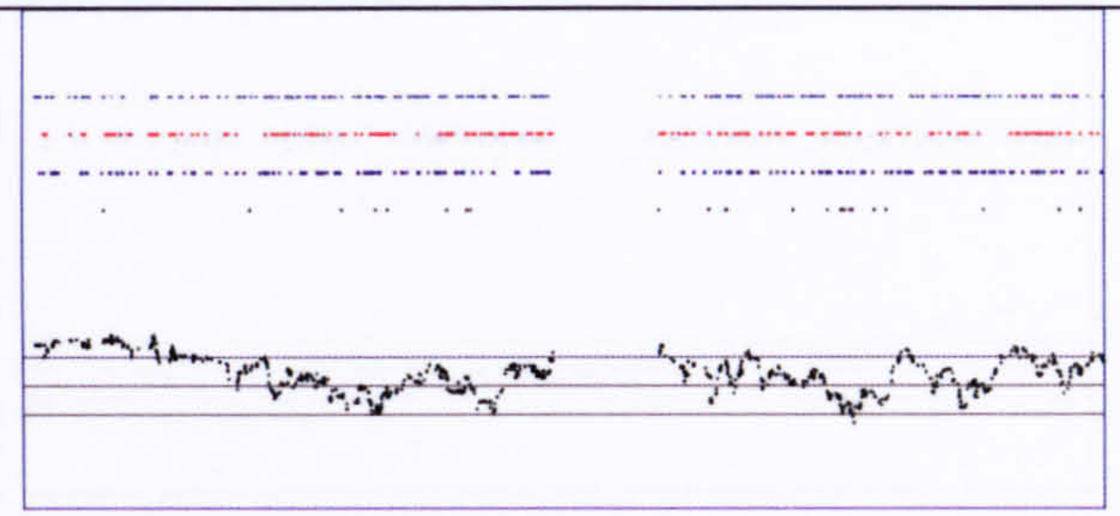
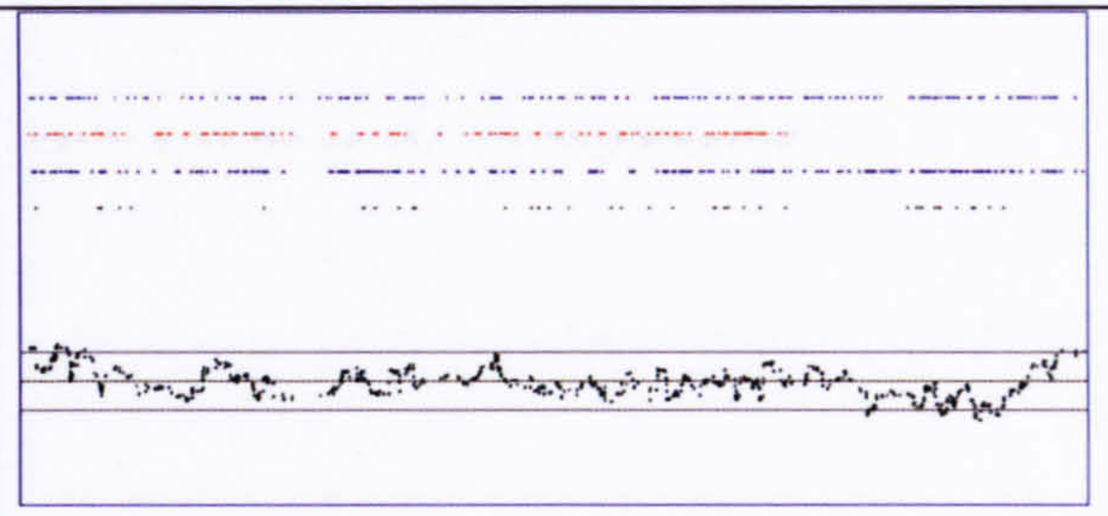
34	45,XY,-7,+mar. ish ider(7)add(7)(wcp7+,CEP7+,D75486-)	7 deletion, Except proximal part of q arm		Normal DNA
		13q gain		
GSM173395	46,XX,del(16)(q12.1)/46,XX	16q deletion		Normal DNA

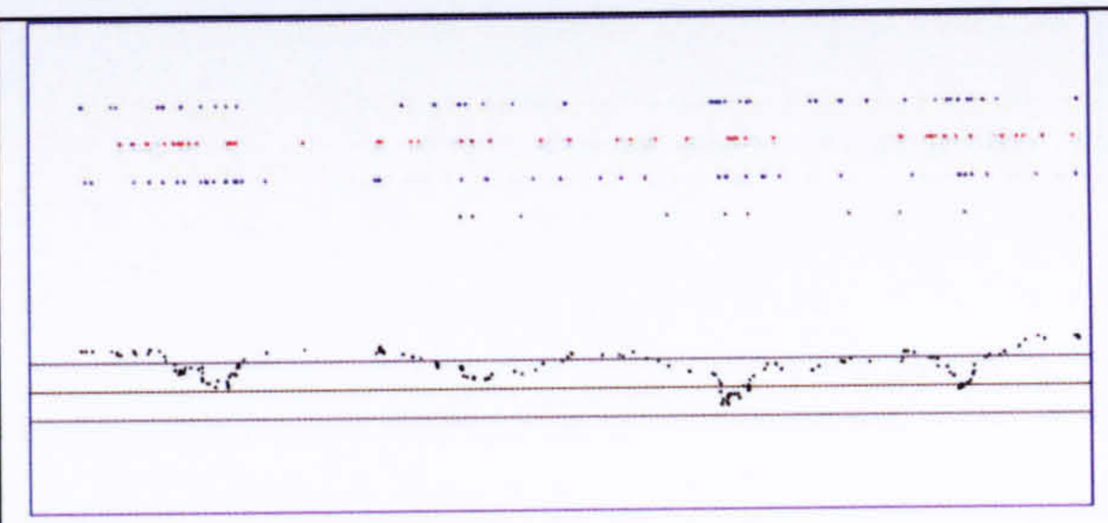
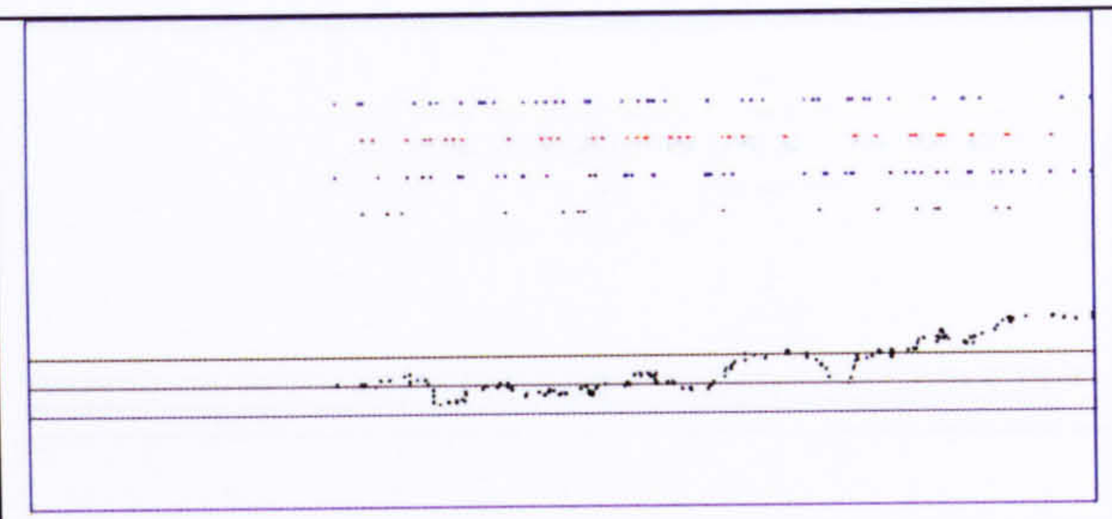
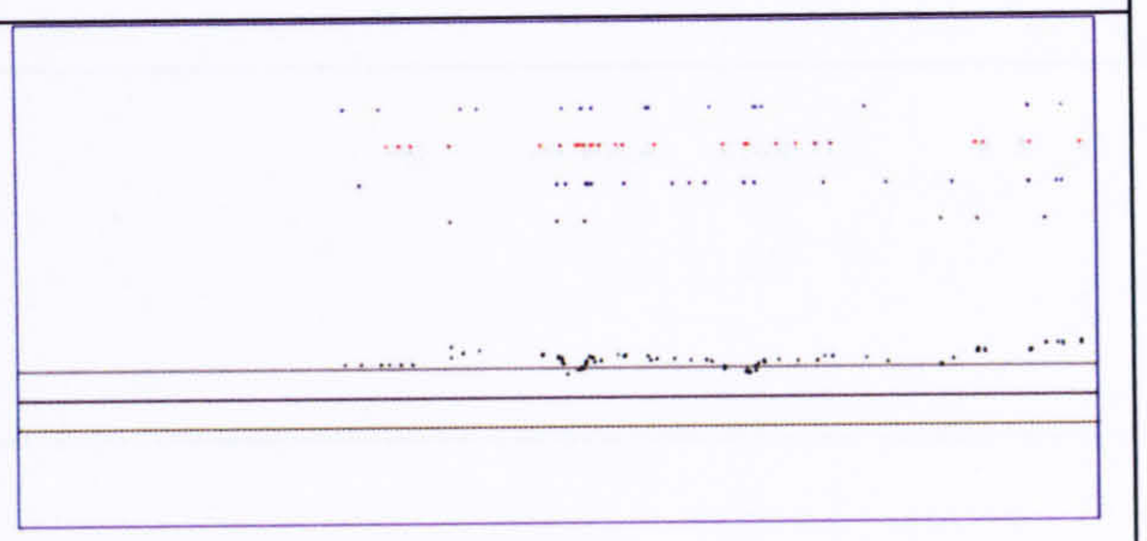
68	48,XY,+11,+13	11 gain		Normal DNA
		13 gain		
16	46,X,-X,t(8,21)(q22;q22)	1p gain		Germline

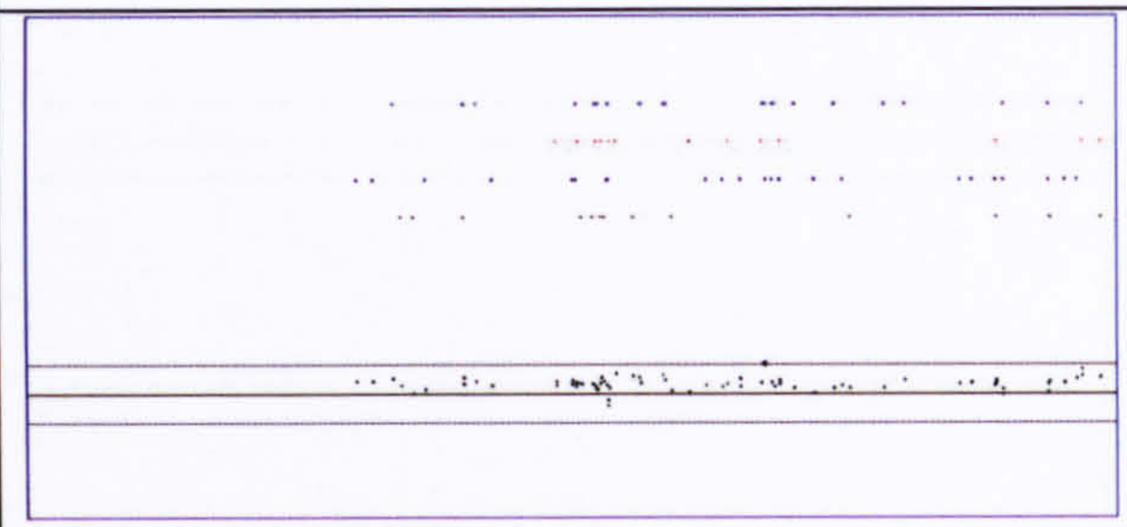
10	46,XX,dic(7;22), +8[10]	7q deletion		Germ line
		8 gain		
		17 gain		

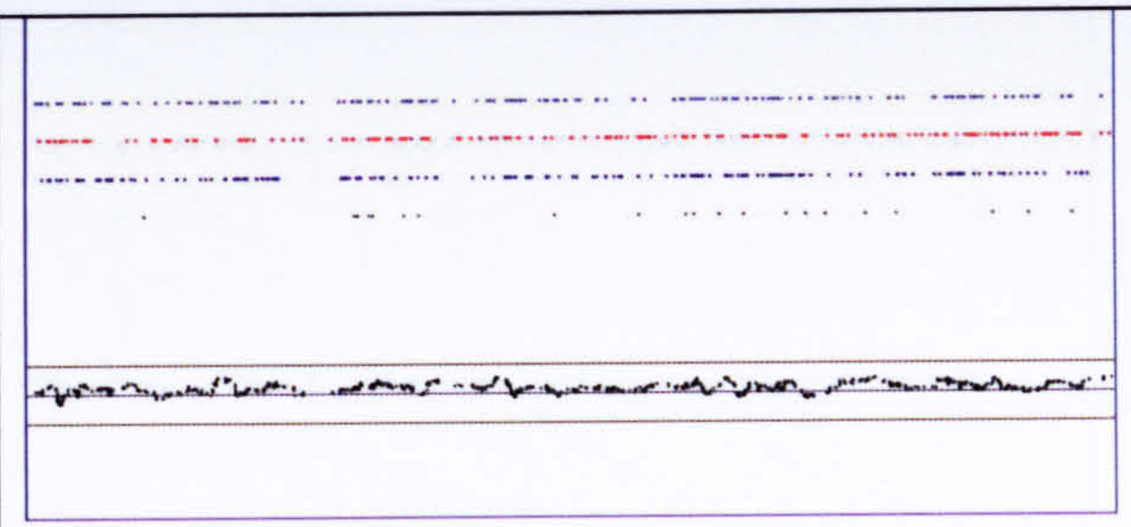
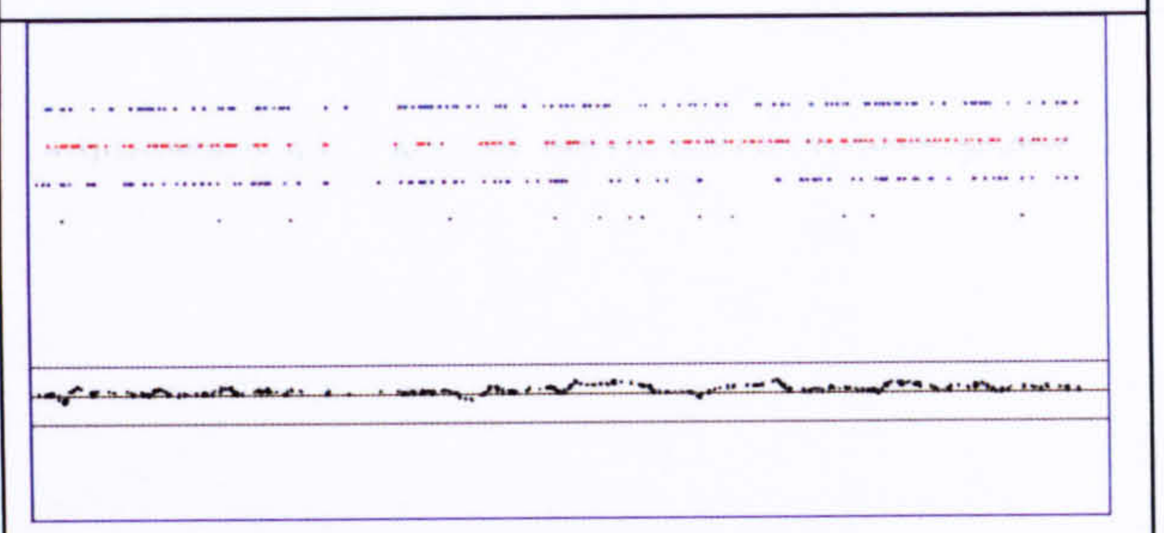
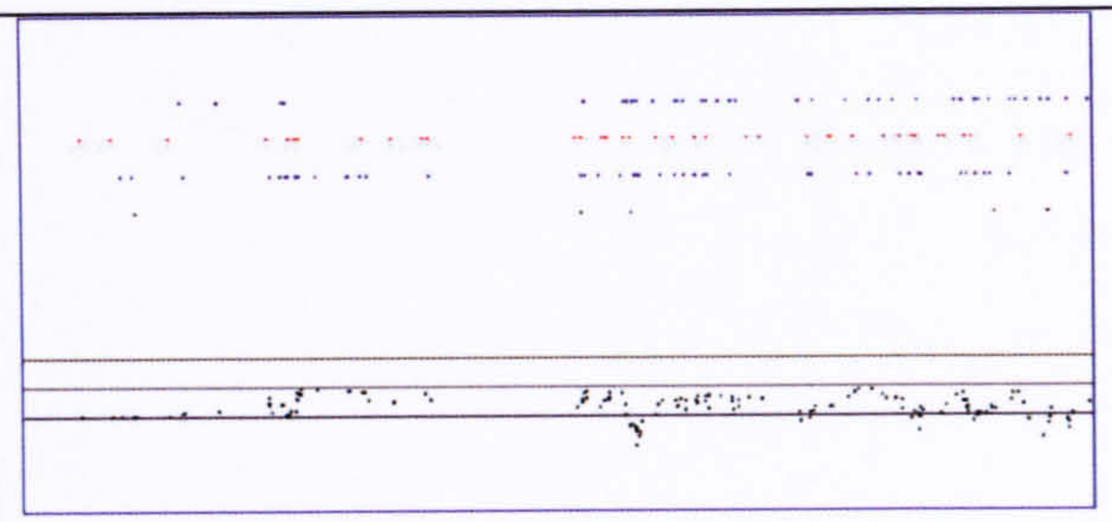
23	46,XX	7q interstitial deletion		Normal DNA
32	47,XY,+8	8 gain		Germ line
39	46,XY	1p gain		Normal DNA

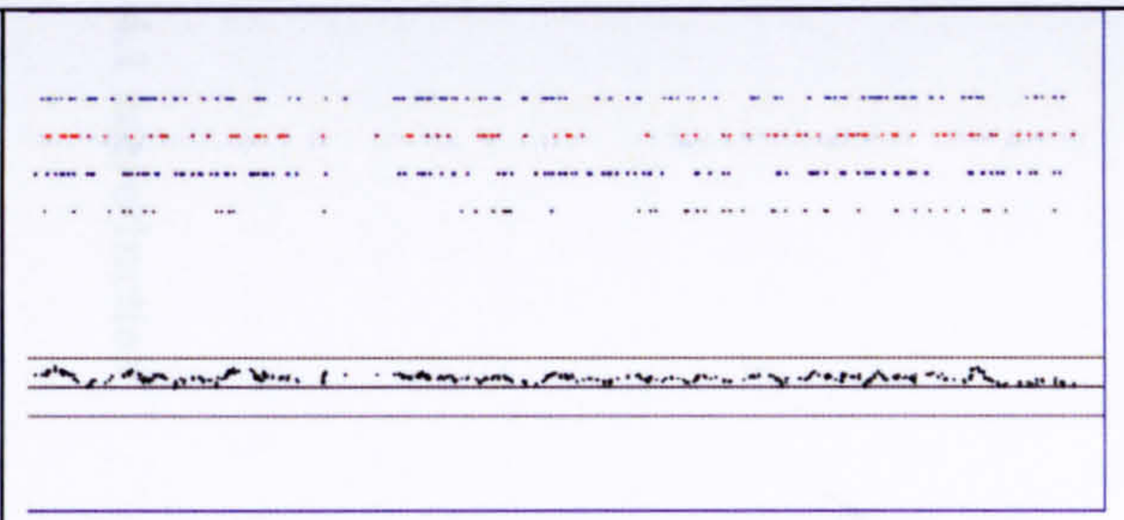
311	46,XY,t(15;17)(q22;q21)[2] /46idem,ider17(q10)t(15;17)(q22;q21)[8]	17p deletion		Germ line
GSM173409	46,XY,(8;9)(q24;q32)[4] /45,Idem,-Y,der(9)add(9)(p13)t(8;9)[11] /45,idem,-Y,der(9)add(9)(p13)t(8;9),add(14)(q32)[3]	9p interstitial deletion		Normal DNA
37	46,XX	1p gain		Normal DNA

58	47,XX,inv(16)(p13q22),+22	22 gain		Germ line
38	46,XY,del(5)(q31)[6]/46,XY	1p gain		Normal DNA
		5q deletion		

38	46,XY,del(5)(q31)[6]/46,XY	17 gain	
		21q gain	
		22 gain	

17	46,XY,t(8,21)	21q gain		Normal DNA
22	47,XY,inv(16)(q13;q22),+22	22 gain		Germ line

GSM173413	49,XY,+5,+8,+19[9]/46,XY[1]	5 gain		Normal DNA
		8 gain		
		19 No gain seen		

GSM173414	47,XY,+8/46,XY	8 gain		Normal DNA

Chapter 4

Acquired uniparental disomy in AML

4.1 Introduction

One of the hallmarks of human cancer is the acquisition of chromosomal abnormalities. The recurrent abnormalities, such as translocations, have been particularly informative in establishing the pathogenesis of haematological malignancies. As discussed in chapter 1, the prognosis of AML correlates closely with the karyotype of the leukaemic cells, but a large proportion of AML has either a normal karyotype (NK) or non-recurrent chromosomal abnormalities.

Expression profiling studies have shown that NK AMLs form a discrete subgroup. These leukaemias have a characteristic upregulation of certain homeobox genes (Debernardi, *et al* 2003). Several single gene mutations have been documented in AML, and some, such as *FLT3* and *NPM1* mutations, have been associated with the presence of a normal karyotype. However, for many NK AML patients, the underlying pathogenesis remains uncertain. Any alternative techniques demonstrating the heterogeneity of this large group presents the opportunity of additional risk stratification, might allow for a more rational selection of therapy, and shed further light on the pathogenesis of the disease.

Because SNP arrays can detect copy number changes as well as LOH, they can identify abnormalities where there is no net change of genetic material, unlike aCGH. Although LOH

often implies deletion, there are large-scale genetic events that would leave the copy number unchanged. Gene conversion is an unequal crossover event, such that one allele becomes the same as the other, and would lead to a region of LOH. Mitotic recombination is an equal cross over event that leads an exchange of material between parental chromosomes (Figure 4.1). At cell division, this subsequently leads to cells that are homozygous for the region distal to the recombination. If one of these cells has a selective advantage, LOH would be detected across the homozygous region.

In this chapter, it is demonstrated that high resolution SNP genotyping can identify large unexpected regions of LOH, with no change in copy number, which results from mitotic recombination leading to acquired partial uniparental disomy.

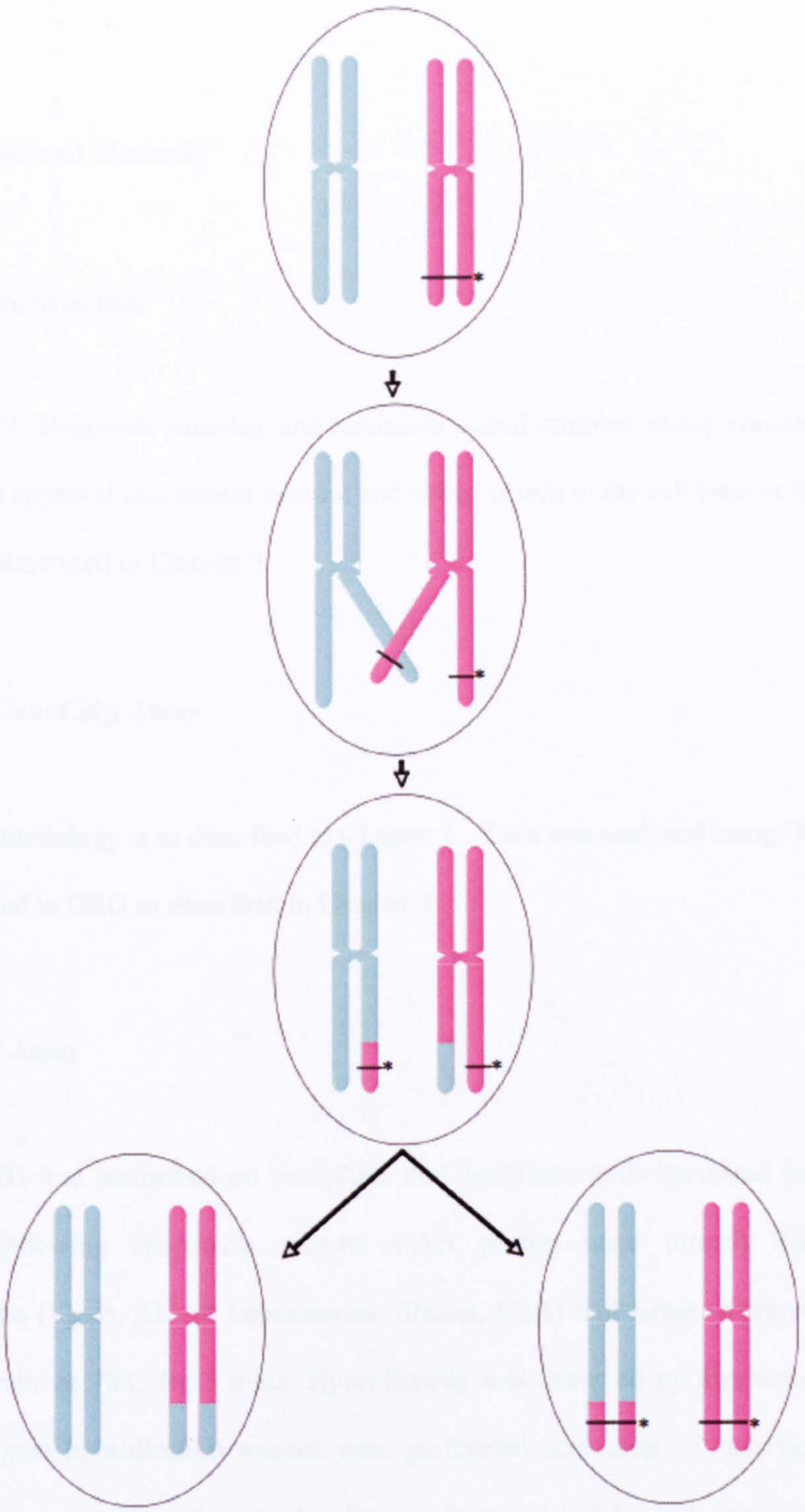


Figure 4.1 Diagrammatic representation of mitotic recombination. The long arm of the chromosome has exchanged material between the two parental chromosomes and on cell division two clones are produced each homozygous for the long arm distal the recombination event. The "" represents a gene that has a selective advantage when homozygous.*

4.2 Materials and Methods

4.2.1 Sample Selection

AML diagnostic samples, and remission paired samples where available were obtained with ethical approval and patient consent and stored frozen in the cell bank at St. Bartholomew's hospital as described in Chapter 3

4.2.2 10K GeneChip Assay

Methodology is as described in Chapter 2 . Data was analysed using GOLF software and deposited in GEO as described in Chapter 3 .

4.2.3 FISH Assay

FISH was performed on metaphase and interphase cells harvested from bone marrow aspirates following short-term culture. FISH probes were directly labelled with the fluorochrome (Vysis, Abbott Laboratories, Illinois, USA) and target slides were co-denatured with the probe at 70°C for 5 mins. Hybridisation was achieved by incubation at 37°C for 24 hours and post hybridization washes were performed according to manufacturer's protocol. Slides were counterstained with 4,6-diamino-2-phenyl-indole and screened using a Leica DMXRA microscope (Leica Imaging Systems, Cambridge, UK) fitted with a COHU CCD camera. MacProbe V4.1.1 CGH software (PSI/Applied Imaging) was used for the analysis.

4.2.4 Bisulphite sequence analysis

The DNA methylation status of the CTCF binding site 6 (CBS6), (GenBank accession no. AF125183; nucleotides 7855-8192) included in the H19 differential methylated region on 11p, was assessed by bisulphite genomic sequencing. The genomic DNA of three samples, two carrying the UPD at the 11p15 site and one control, was amplified by PCR after bisulphite treatment, as previously described (Cui, *et al* 2002) using the primers: *h19-f1*, 5'-GAGTTTGGGGGTTTTTGTATAGTAT-3' and *h19-r1*, 5'-CTTAAATCCCAAACCATAACACTA-3', followed by *h19-f2*,

5'-GTATATGGGTATTTTTGGAGGT-3'and *h19-r2*,

5'-CCATAACACTAAAACCCTCAA-3'. The PCR products were directly sequenced.

4.2.5 Mutation analysis

Primers and precise amplification conditions are as described in Chapter 2 . Mutation screening was done for the entire coding region (*CCND3, CDKN2A, CDKN2B, PU.1*, and *CEBPA*) or specific exons [*WT1* (7-10), *FLT3* (exons 14-15 and 20), *RUNX1* (exons 3-5), and *MLL* (PTD)] as described in Chapter 2 . PCR products were sequenced directly or cloned by TOPO TA cloning (Invitrogen Ltd., Paisley, United Kingdom), plasmid DNA extracted and sequenced by use of an ABI 377 DNA sequencer (PE Applied Biosystems, Foster City, CA). For direct sequencing, unincorporated primer was removed by ultrafiltration using a Centricon YM-100,000 filter device (Millipore). Sequencing data were analyzed using DNASTAR, Inc. (Madison, WI).

4.3 Results

4.3.1 Identification of acquired UPD

A high-resolution genotype analysis was performed on DNA from 64 presentation AML samples for which full karyotype information was available. The range of leukaemias examined was typical of AML and included those with a normal karyotype and with the translocations resulting in chimeric transcription factors such as t(8;21), inv(16), t(15;17) and 11q23. Using the 10K SNP array (Affymetrix, Inc., Santa Clara) (Matsuzaki, *et al* 2004), a mean call rate of 93.3% was obtained resulting in more than 10,000 SNP genotype calls per sample.

Large unexpected regions of homozygosity were observed in 15 of 72 AMLs from a total of 72 (20.8%) (Table 4.1). These regions ranged in size from 16 to 113 million base pairs and would have been visible in the karyotypes had they been due to deletion. The signal values were calculated within the regions of LOH as described in the legend to Table 1 (Bignell, *et al* 2004, Zhao, *et al* 2004). In every example, the SNP signal values were indicative of two copies (Table 4.1). The presence of two copies has been confirmed by fluorescence in situ hybridisation (FISH) in the 4 cases analysed. In order to assess whether these regions of homozygosity were restricted to the leukaemic clones, DNA from five available remission bone marrow samples (for samples paired with AML samples 35, 10, 69, 40 and 49 in Table 4.1) was subjected to SNP genotype analysis. The SNP call data (deposited at the Gene Expression Omnibus under accession number GSE7490) demonstrated clearly that the homozygosity seen in the leukaemic DNA was not present in the respective remission bone marrow DNA. It was also evident that all of the SNP calls in the homozygous regions were concordant with the equivalent calls in the remission bone marrow DNA. A detailed comparison between leukaemic DNA and remission DNA for patient sample 10 is presented in Figure 4.2 for relevant chromosomes. The large region of homozygosity on chromosome 11q is illustrated as a fall in the ratio of heterozygous: homozygous calls (Figure 4.2A). The ratios of leukaemia versus remission signal values were calculated along the length of chromosome 11 as described (Table 4.1). There was no

significant drop in signal ratio values for the region of homozygosity indicating that the leukaemia had a normal copy number for chromosome 11.

Table 4.1 The 12 leukaemias exhibiting UPD, listed with their respective karyotypes.

In patient sample 41, the presence of the PML-RARα gene fusion was confirmed by FISH in the majority of interphase cells. In total, 64 diagnostic acute myeloid leukaemia bone marrow or peripherally taken blood samples from the frozen tissue bank at St Bartholomew's Hospital were analysed using the 10K SNP array that encodes approximately 11,500 SNP loci (Affymetrix, Inc., Santa Clara) (Matsuzaki, et al 2004). For each SNP on the chip there are 20 probe pairs, 10 for allele A and 10 for allele B. A probe pair contains a perfect match (PM) and a mismatch (MM). The signal value is calculated as the average intensity difference between a perfect match and a mismatch for all probe pairs, and is given by 1/20, where PMi is the intensity of perfect match of pair i and MMi is the intensity of mismatch for the same pair. The signal value ratio is the ratio of mean signal within UPD region divided by mean signal outside, for the same chromosome. For patient samples 79 and GSM173396 we have calculated the ratio of mean signal of chromosome 13 to the mean of the mean signal of all other chromosomes. All the statistical analysis is done with R. Abbreviations: ND, not done.

Patient sample	FAB class.	Karyotype	Chromosomal Region of UPD	Signal value ratio	Centromeric heterozygous SNP	Base range	FISH copy no
35	(M1)	46 XY[20]	11p13-11pter	0.92	SNP_A-1509863	1-34938966	ND
37	(M2)	46 XX[20]	11p11-11pter	0.88	SNP_A-1511180	1-52461565	ND
10	(M1)	46,XX,dic(7;22)(q11.2;q10), +8[10]	11q12-11qter	0.85	SNP_A-1512785	65274905-137400000	Failed
20	(M1)	46 XX[20]	11q13-11qter	0.80	SNP_A-1512811	66987057-137400000	11q23(MLLx2) [100]
65	(M4)	46,XX[20]	11p11-11p14	0.85	SNP_A-1508814-SNP_A-1510427	25813006-45885301	ND
69	(M4)	48,XY,+3,+10[10]	6p21-6pter	0.90	SNP_A-1518080	1-44242272	ND
40	(M1)	46,XX[20]	6p11-6pter	1.09	SNP_A-1507946	1-57365515	ND
41	(M3)	46,XY,t(15;17)(q22;q21)[1] /46,XY[19]	9q33-9qter	1.05	SNP_A-1516948	116551991-132400000	9q34(ABLx2) [100]
7	(M5)	46,XY,der(12)t(1;12) (q11;p11.2)[7]/46,XY[15]	9pcen-9pter	0.88	SNP_A-1517078	1- 43264784	9p21(p16x2) [100]
49	(M1)	46,XY[25]	19q12-19qter	0.94	SNP_A-1515117	28798570-60000000	19q13(CEBPAx2) [100]
79	(M1)	46,XX[20]	13	0.80	Whole Chr 13	1-113400000	Failed
76	(M2)	46,XY[20]	21q21.1	1.07	SNP_A-1509329	20420648-44464366	ND
GSM173434	(M4)	46,XY,t(6;9)(q23;q34)[10]	13q-ter	0.94	SNP_A-1511719	18482385-113400000	ND
GSM173423	(M2)	46,XX[25]	13	0.92	Whole Chr 13	1-113400000	ND
GSM173424	(M4)	46,XX	13q-ter	0.88	SNP_A-1511719	19029353-113400000	ND

Since this leukaemia had a rearrangement resulting in a deletion of 7q, dic(7;22), an identical calculation was performed for chromosome 7. As shown in Figure 4.2B there was a large homozygous region on 7q but, in contrast to the results above, the leukaemia to remission signal ratio values fell to approximately 0.5 in the homozygous region. A DNA probe for the *MLL* gene (11q23), which lay within the homozygous region, was used in FISH experiments to both metaphase and interphase cells from this leukaemia. This confirmed the presence of two copies of 11q23 in metaphase and interphase cells (Figure 4.2C). It was therefore concluded that these regions of homozygosity (Table 4.1 and Figure 4.2) represented somatically acquired loss of heterozygosity (LOH), due to the presence of partial uniparental disomy (UPD), which was restricted to the leukaemic clone.

This study demonstrates that the acquisition of partial uniparental disomy is a relatively common event in adult AML and selected examples are illustrated in Figure 4.3. There appears to be a non-random distribution of these events with 4 examples of UPD13q, 3 examples of UPD11p, and 2 examples each of UPD11q and UPD6p.

It was also apparent that the points of recombination are clearly different in all examples. In 12 of the 15 examples, the region of LOH began within a chromosome arm and continued to the telomere implying that this phenomenon is due to mitotic recombination. In two examples, the entire chromosome 13 was homozygous (patient samples 79 and GSM173396) suggesting either a non-disjunction event followed by chromosomal duplication or a recombination close to the centromere. The one example (patient sample 65) of interstitial LOH could be due to a double crossover event affecting 11p. It is interesting to note that there was only one example of UPD in the leukaemias with balanced translocations, although a larger series would be need to be examined to confirm this low frequency.

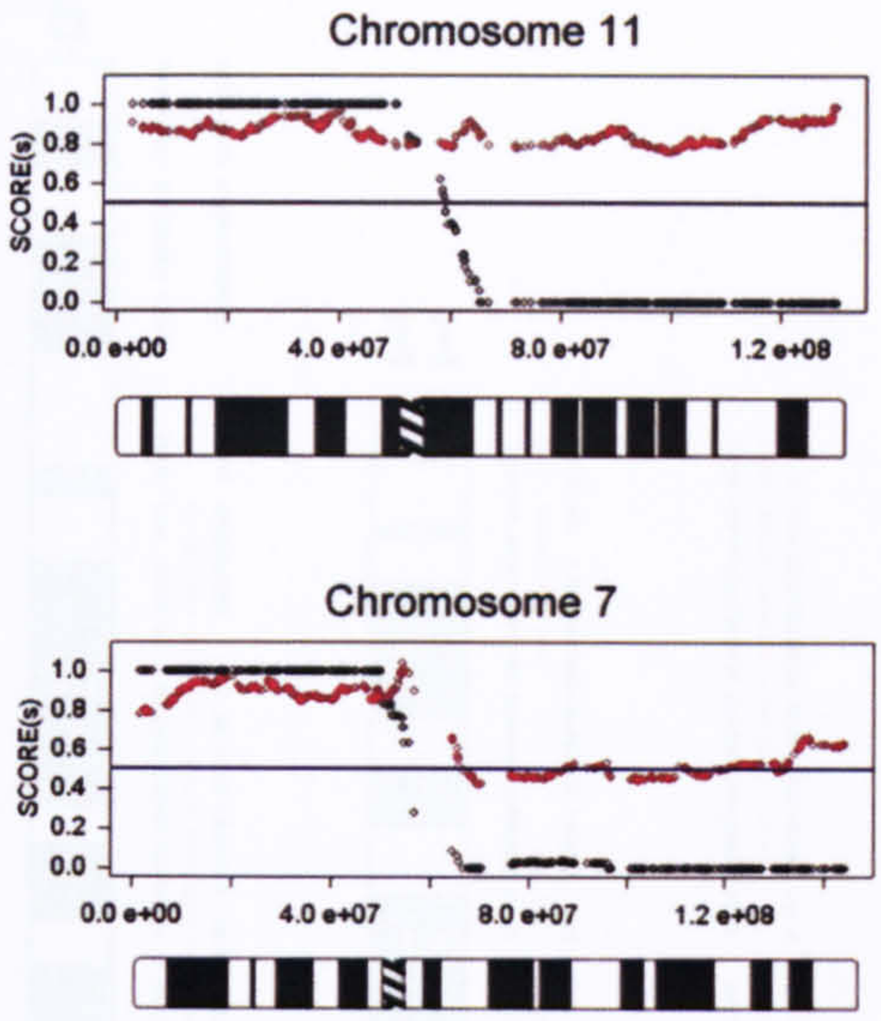


Figure A

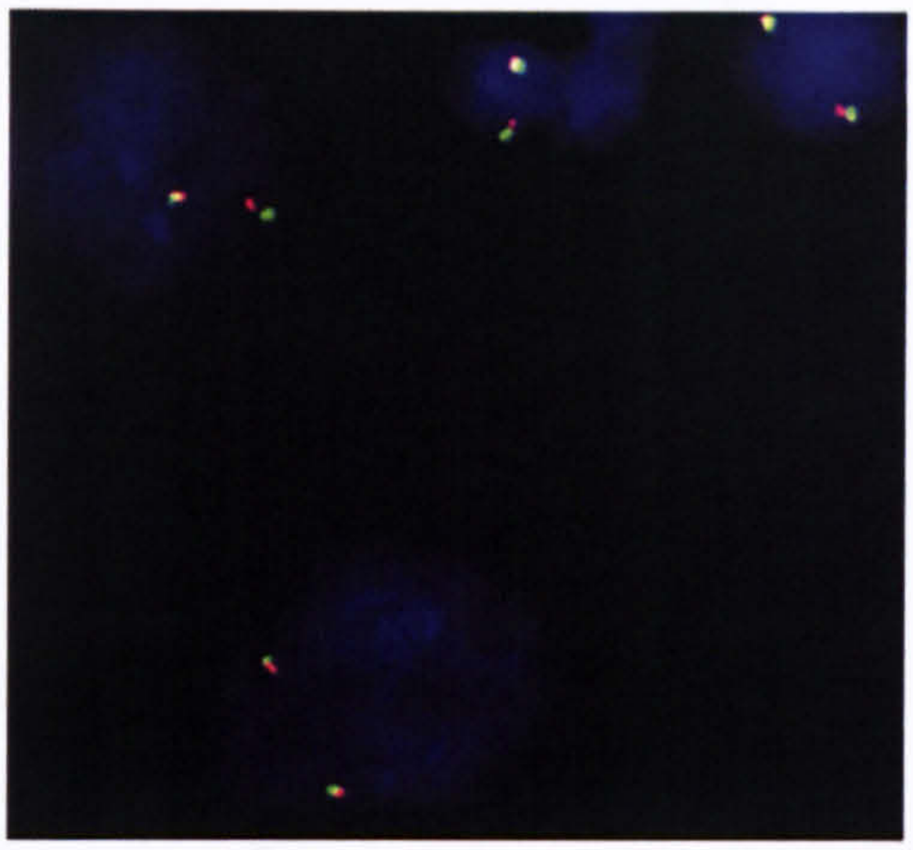


Figure B

Figure 4.2 Analysis of UPD in patient sample 10.

A. The ratio of the number of heterozygous to homozygous calls in a running window of 20Mbp was calculated. The LOH score (red) was calculated by dividing the above ratio in the leukaemia sample by the same ratio in the remission sample. The signal score (black) is calculated as the ratio of the mean signal (Table 4.1), in a running window of 20Mbp, between leukaemia (sample 10) and remission (paired sample 44). Results for analysis of chromosome 11 in upper panel and for analysis of chromosome 7 in lower panel.

B. FISH hybridisation of a two-colour probe for the MLL gene. The hybridisation signals show two copies of the MLL gene in interphase cells from patient sample 10.

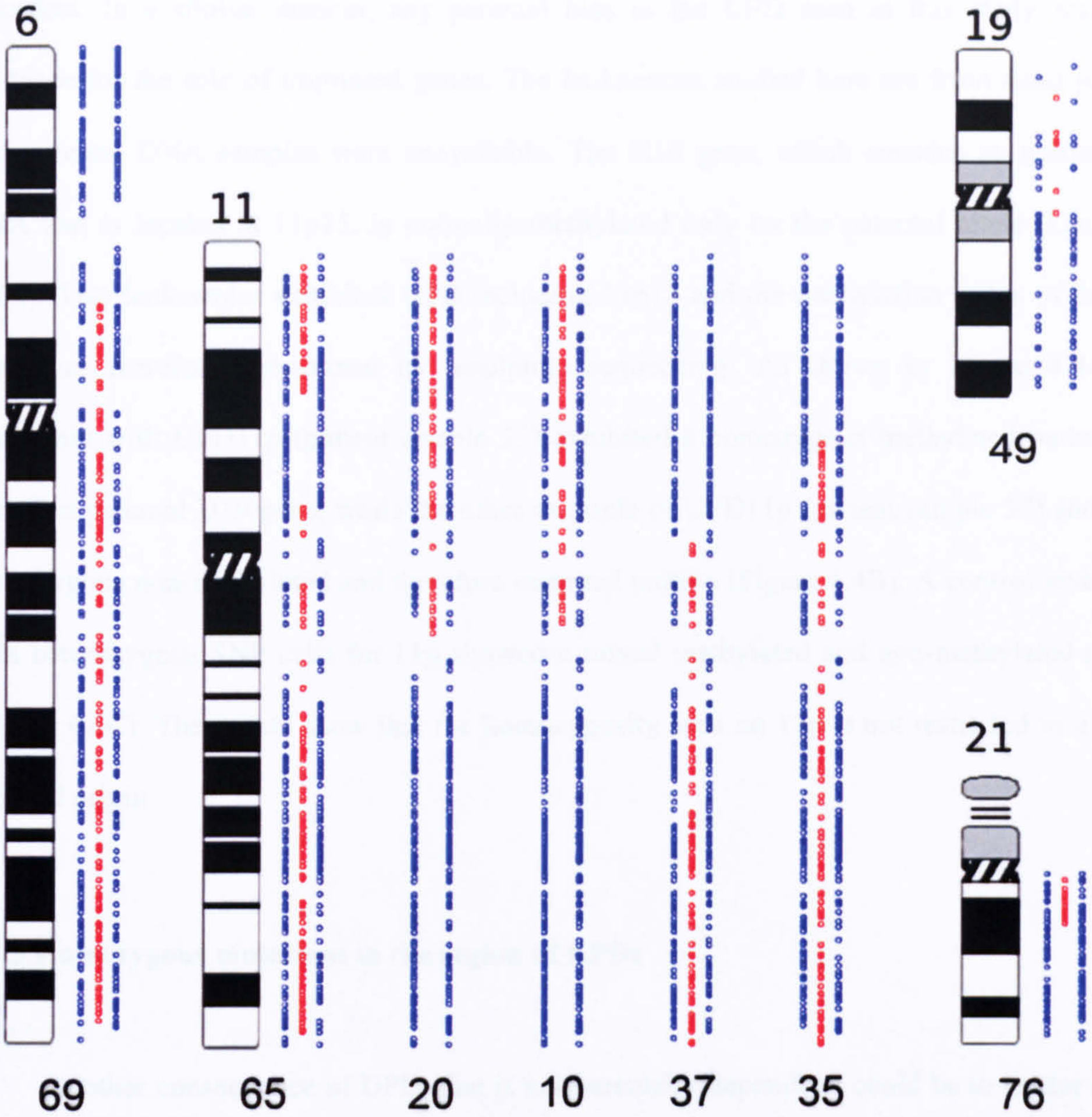


Figure 4.3 Display of SNP calls for chromosomes exhibiting large regions of homozygosity.

The left (blue) column shows A/A calls, the middle column (red) shows A/B calls and the right (blue) column shows B/B calls for each chromosome. The patient numbers (Table 4.1) are indicated below each set of calls. Results are shown for chromosome 6, 11, 19 and 21.

4.3.2 Imprinted genes in the region of UPD

The expression of imprinted genes in the homozygous regions would be expected to be profoundly affected. Congenital disorders associated with UPD have been linked to a single parental origin, e.g. Beckwith-Weidemann syndrome and paternal UPD of 11p15 (Henry, *et al* 1991), suggesting the parental pattern of methylation is important to the pathogenesis of such

disorders. In a similar manner, any parental bias in the UPD seen in this study would be evidence for the role of imprinted genes. The leukaemias studied here are from adult patients and parental DNA samples were unavailable. The H19 gene, which encodes an untranslated RNA and is located at 11p15, is normally methylated only on the paternal allele (Cui, *et al* 2002). Two leukaemias exhibited UPD including 11p15 and the methylation status of the H19 gene was therefore determined by bisulphite sequencing. As shown in Figure 4.4A one leukaemia with UPD11p (patient sample 35) exhibited a homozygous methylated pattern and therefore paternal H19 gene, while the other example of UPD11p (patient sample 37) showed a homozygous non-methylated and therefore maternal pattern (Figure 4.4B). A control leukaemia with heterozygous SNP calls for 11p showed a mixed methylated and non-methylated pattern (Figure 4.4C). These data show that the homozygosity seen on 11p is not restricted to a single parental origin.

4.3.3 Homozygous mutations in the region of UPDs

Another consequence of UPD, that is not parentally dependent, could be to render one of the daughter cells homozygous for a pre-existing mutation. In a previous analysis (Snaddon, *et al* 2003) one of the leukaemias studied here (patient sample 49 in Table 4.1) was shown to be homozygous for a CEBPA mutation and FISH demonstrated 2 copies of the CEBPA gene. This gene is located at 19q13.1, within the area of UPD and we conclude that the mutation occurred prior to the UPD. We can therefore speculate that an important consequence of mitotic recombination could be to render pre-existing mutations homozygous. Eight different chromosomal regions were shown to be affected by UPD in this study and this suggested that there would be at least this number of mutational targets.

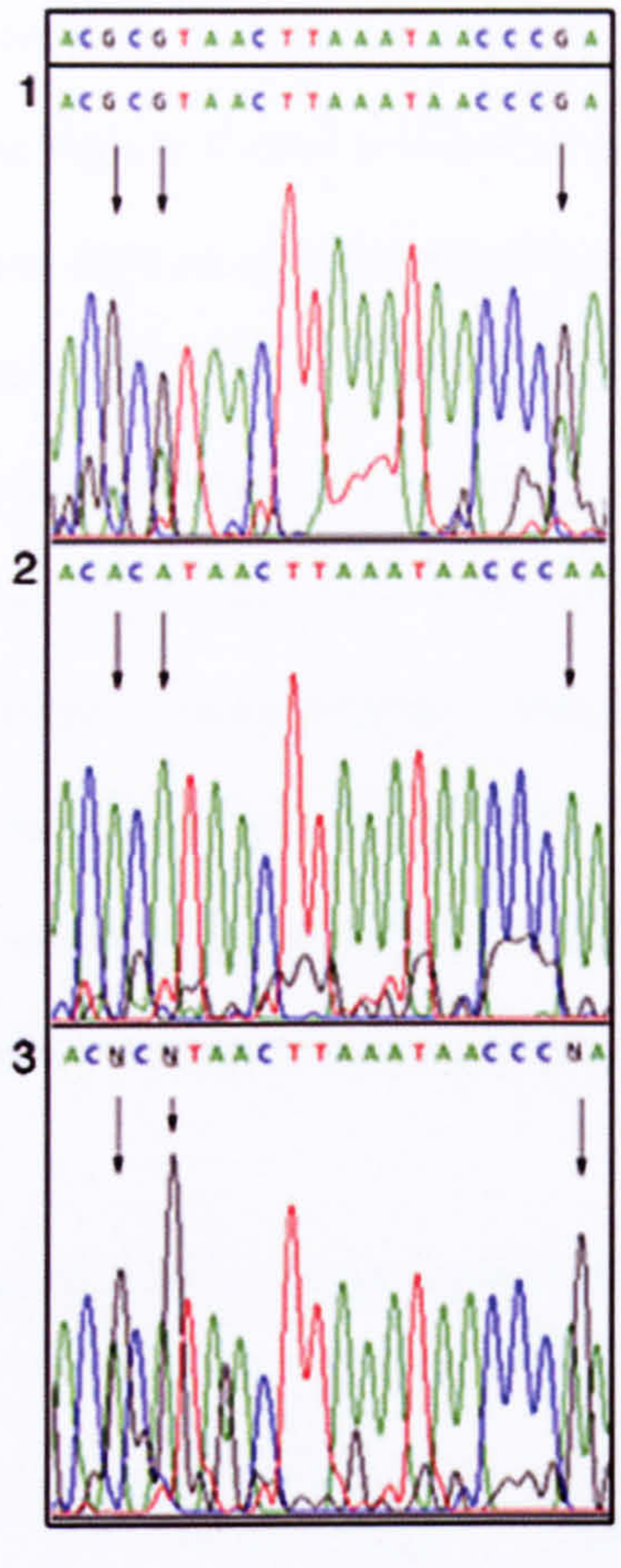


Figure 4.4 DNA methylation of the CTCF binding site 6 (CBS6), included in the H19 DMR.

Part of the sequence obtained with h19r2 oligonucleotide is shown and three CpG sites are indicated by arrows. On top is shown the genomic DNA sequence before bisulphite treatment. 1) In the first sample (patient sample 35, Table 4.1) all the cytosines in the CpG dinucleotides of both alleles remain unchanged, therefore methylated, indicating homozygosity for the paternal allele. 2) In the second sample (patient sample 37, Table 4.1) all the cytosines in the CpG dinucleotides of both alleles are converted in thymine by the treatment (adenine in the chromatogram as the reverse strand is shown), indicating homozygosity for the maternal allele. 3) In the control sample both alleles are present as shown by the two overlapping peaks, indicating a state of heterozygosity.

To investigate this phenomenon further, we selected 13 of the primary AMLs with uniparental disomy encompassing regions known to harbour genes potentially mutated in AML (Table 4.2). Sequence analysis was done on nine genes in the appropriate leukaemias. Mutations were detected at four distinct loci (*WT1, FLT3, CEBPA*, and *RUNX1*) in 7 of the 13 leukaemias examined. In all cases, sequence analysis indicated a homozygous mutation with no evidence of the wild-type sequence. In Figure 4.5, the sequence traces showing a homozygous insertional mutation of the *WT1* gene are illustrated along with the homozygosity map for chromosome 11. Similarly, a homozygous single-base insertion of the *RUNX1* gene is shown with the concurrent homozygosity of 21q. It was notable that uniparental disomy of 13q was associated with homozygous internal tandem duplication mutation of *FLT3* (*FLT3*-ITD) in four separate cases.

Table 4.2 Regions of UPD and associated homozygous gene mutations.

Patient sample	Karyotype	UPD	Gene	Mutation
35	46,XY	11p13-11pter	*WT1, PU.1*	1584 ins16bp *(WT1)*
49	46,XY	19q12-19qter	*CEBPA*	1038 ins57bp *(CEBPA)*
79	46,XY	21q21-21qter	*RUNX1*	970 insA *(RUNX1)*
76	46,XX	13q	*FLT3*	ITD 21bp *(FLT3)*
GSM 173434	46,XY t(6;9)	13q	*FLT3*	ITD 63bp *(FLT3)*
GSM 173396	46,XX	13q	*FLT3*	ITD 36bp/ 39bp* *(FLT3)*
GSM 173397	46,XX	13q	*FLT3*	ITD 108bp *(FLT3)*
37	46,XX	11p11-11pter	*WT1, PU.1*	None
65	46,XX	11p11-11p14	*WT1, PU.1*	None
10	46,XX,dic(7;22)(q11.2;q10),+8	11q12-11qter	*MLL*	None
69	48,XY,+3,+10	6p21-6pter	*CCND3*	None
40	46,XX	6p11-6pter	*CCND3*	None
7	46,XY,der(12)t(1;12)(q11;p11.2)/46,XY	9pcen-9pter	*CDKN2A, CDKN2B*	None

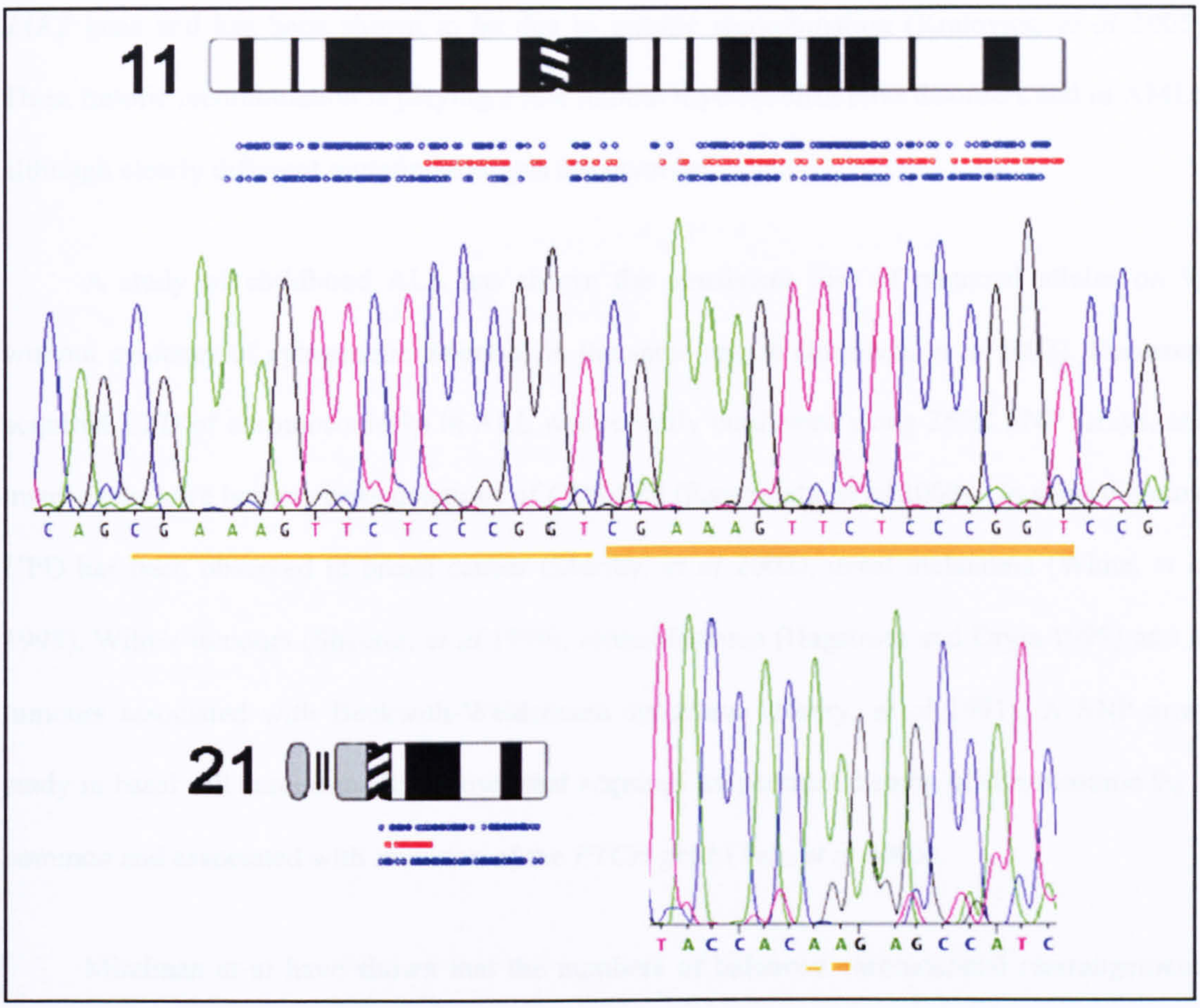


Figure 4.5 Display of UPD on 11p and 21q in two AML samples. Heterozygous calls (red) and homozygous calls (blue). The corresponding sequencing traces showing 1584ins16bp WT1 and 970insA RUNX1 homozygous mutations. WT and RUNX1 duplicated segments (underlined in orange).

4.4 Discussion

There is no *a priori* reason why UPD should be restricted to AML, and indeed the use of microsatellite markers has suggested UPD in other clonal haematological disorders such as polycythaemia vera (Kralovics, *et al* 2002). Several groups have shown that a frequent mutational target in myeloproliferative disorders such as polycythaemia vera is the *JAK2* gene, located within 9p (Baxter, *et al* 2005, James, *et al* 2005, Kralovics, *et al* 2005, Levine, *et al* 2005). Loss of heterozygosity at 9p is frequently observed in association with mutations to the

JAK2 gene and has been shown to be due to mitotic recombination (Kralovics, *et al* 2005). Thus, mitotic recombination is playing a role in both myeloproliferative disorders and in AMLs, although clearly different mutational targets are involved.

A study of childhood ALL has shown the consistent loss of maternal alleles on 9p without evidence of cytogenetic deletions in the same region (Morison, *et al* 2002). Recurrent acquired UPD of chromosome 9p in ALL was recently confirmed using 250K SNP arrays, and many cases have homozygous deletions of *CDKN2A* (Kawamata, *et al* 2008). In solid tumours, UPD has been observed in breast cancer (Murthy, *et al* 2002), uveal melanoma (White, *et al* 1998), Wilm's tumours (Shearer, *et al* 1999), retinoblastoma (Hagstrom and Dryja 1999) and in tumours associated with Beckwith-Weidemann syndrome (Henry, *et al* 1991). A SNP array study in basal cell carcinoma has shown that acquired uniparental disomy of chromosome 9q is common and associated with mutation of the *PTCH* gene (Teh, *et al* 2005).

Mitelman et al have shown that the numbers of balanced chromosomal rearrangements are similar when haematopoietic malignancies are compared to solid tumours (Mitelman, *et al* 2004) suggesting that general pathogenic mechanisms are common to all cancers. In this context it is interesting to note that a recent study of solid tumour cell lines using SNP genotyping has suggested that mitotic recombination may be playing a role since regions of homozygosity were observed which corresponded to normal copy number values (Bignell, *et al* 2004). Because UPD is undetectable by cytogenetic analysis, its frequency in cancer may be much higher than previously thought.

Mitotic recombination, which has been shown to occur at a frequency of 10^{-4}-10^{-5} in normal human and mouse cells (Tischfield 1997), results in a cell population mosaic for the recombined chromosomes. The UPD seen in leukaemias must represent the clonal outgrowth of one of the daughter cells due to a selective advantage. In the general population, inherited UPD is a rare occurrence (Karanjawala, *et al* 2000). Studies using the HLA locus have shown that there is inter-individual variation in the rate of mitotic recombination (Holt, *et al* 1999) and that

it increases with age. Recent studies using a mouse *Trp53*$^{+/-}$ BALB/c strain showed that mitotic recombination played a major role in generating LOH at the *Trp53* locus in a series of mammary tumours (Blackburn, *et al* 2004). Furthermore, the high rate of mitotic recombination was shown to be a dominant trait in crosses with the C57BL/6 strain. It may be speculated that the presence of UPD in a cancer may identify individuals with a higher rate of mitotic recombination and that this could be genetically determined.

The presence of a mutation and the absence of the wild-type sequence has previously been reported in AML for *WT1* (Miyagawa, *et al* 1999), *CEBPA* (Leroy, *et al* 2005), and *RUNX1* (Silva, *et al* 2003, Smith, *et al* 2005). Such a situation could arise by several mechanisms, including the deletion of wild-type allele to yield a hemizygous state, amplification of the mutant allele, or mitotic recombination. Without accurate copy number information, it has been difficult to distinguish between these possibilities. The high degree of association between uniparental disomy and homozygous mutation presented here suggests that uniparental disomy may account for a substantial proportion of cases exhibiting a lack of the wild-type allele. Indeed, it has been previously suggested that homozygous *RUNX1* mutations in leukaemias could be due to mitotic recombination (Silva, *et al* 2003). Because relatively large regions of the genome are involved, there may be more than one potential mutational target in a given region of uniparental disomy. An example of a *WT1* mutation was observed in one of the three examples of UPD11p suggesting other potential targets in this region. Similarly, because no mutations were detected in the *CCND3, CDKN2A, CDKN2B*, and *MLL* genes, our data would suggest that there may be other targets in UPD6p, UPD9p, and UPD11q, respectively.

FLT3-ITD mutation in the heterozygous state is a recognized poor prognostic risk factor for AML. Additionally, the loss of the wild-type *FLT3* allele in leukaemias with a *FLT3* mutation in about 10% of patients with normal karyotype AML is associated with a worse outcome (Thiede, *et al* 2002, Whitman, *et al* 2001). Hence, the association between loss of the wild-type *FLT3* allele and UPD13q could suggest that the presence of uniparental disomy may be linked to clinical outcome. The close proximity of *FLT3* to the centromere and the small

number of SNP markers proximal to *FLT3* on chromosome 13 means that in some cases, we cannot exclude the possibility of mechanisms other than mitotic recombination leading to uniparental disomy of chromosome 13. This could include non-disjunction followed by chromosome duplication. Mitotic recombination is clearly the case for *WT1*, *RUNX1*, and *CEBPA* mutation (Figure 4.5) and suggests that mutation precedes mitotic recombination that acts to remove a normal copy of a gene and replace it with a mutated copy.

The discovery of widespread, somatically acquired, UPD in leukaemias has other potentially important biological and clinical implications. Several recent studies have suggested that there are allelic differences in gene expression from non-imprinted autosomal genes that are heritable (Lo, *et al* 2003, Yan, *et al* 2002). Mitotic recombination leading to UPD may thus affect the expression of certain genes within the homozygous regions. Hence, in addition to rendering pre-existing mutations homozygous, mitotic recombination may also exert subtle but critical effects on gene expression.

In summary, this study has identified hitherto unrecognised abnormalities in 20% of normal karyotype AMLs. The frequency of the occurrence of UPD and the degree of consistency with which it occurs suggests it is a critical event in the development of leukaemia and will be important in the classification of AML. The event gives a marker of disease where none was previously available, and an opportunity to look for molecular targets. Although the prognostic consequences of this phenomenon are currently uncertain, given the importance of other frequent chromosomal events in AML, larger studies to assess the clinical significance of mitotic recombination are required.

Chapter 5

A pilot study to map homozygosity in an unselected group of patients with AML

5.1 Introduction

The discovery of acquired UPD in a group of diagnostic AML samples raises a number of issues. The incidence of acquired UPD was found to be 20%, but a more accurate assessment could be made from a larger prospective study. The sample set from Barts was biased because the patients had previously been analysed for mutations and gene expression, and thus were chosen because of sample availability and by karyotype. Although this gave an assessment of the incidence of uniparental disomy and other copy number changes it is not necessarily accurate. A recent smaller study (Gorletta, *et al* 2005), gave the incidence of acquired UPD as 15%.

Acquired UPD was found in non-random chromosomal regions, some of which were associated with homozygous gene mutations. The other UPDs could have unidentified genes within the homozygous region. However, these are large regions of LOH, making identification of potentially mutated genes very difficult. A map detailing regions of LOH using a large number samples could delineate minimal regions of LOH, and putative oncogenes or TSGs could be identified.

Some of the gene mutations associated with UPD have prognostic implications, e.g the presence of the *FLT3* ITD mutation with a loss of the wild type gene confers a poor prognosis (Thiede, *et al* 2002). With a large enough sample set, and sufficient follow up, it would be

possible to make correlations with prognosis.

As it would be time consuming to prospectively obtain sufficient samples for analysis, the alternative was to use samples from previous national studies on AML. This has the advantage that the AML patients were prospectively entered into the trial, although because the analysis is being done retrospectively, there may still be some bias depending on the availability of samples for analysis. However, the patients have many years of follow up, which gives good survival data. The study samples from MRC AML10 are a well-characterised sample set that has been used in studies investigating cytogenetics, gene mutation status and prognosis.

The study required would be large, and to estimate the number of samples needed to give an answer to questions of prognosis, a pilot study was required. Therefore, the purpose of this study was to show there were a significant proportion of patients with acquired UPD, and to assess how many patients would be required for a larger study.

5.2 Methods

Samples with DNA available at diagnosis were obtained from the Medical Research Council AML10 study. This study recruited from May 1988 to April 1995. In brief, this trial was designed to randomise patients between using daunorubicin and cytarabine with or without etoposide at induction chemotherapy. For good risk patients, there was a further randomisation between completing four or five courses of chemotherapy. For standard risk patients, there was a randomisation between four or five courses of chemotherapy, or using a transplant as the fourth or fifth course. The first 93 patients for whom DNA samples were available were analysed using the Affymetrix 10K 2.0 Mapping Array as described in Chapter 2. Six samples failed to amplify sufficiently during the PCR amplification stage, leaving 87 samples that were successfully hybridised. Comparison was made with cytogenetic data, which was available for 74 patients. The primary SNP microarray data have been deposited in the NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/projects/geo/) and are accessible through GEO series accession number GSE8721.

Remission germ-line DNA was available for only three patients in the pilot group, so only these could be assessed definitively for LOH. Regions of LOH due to deletion could be assessed by correlation of copy number and homozygosity. For UPD, a criterion of a minimum of 30 contiguous homozygous SNPs was used to define a likely region. The mean inter SNP distance on the array is 258kb, giving a minimal size of UPD of 7.7Mb. By decreasing the size of a homozygous region, more regions are detected, but many of these will be germ-line rather than somatically acquired. On analysis of germ-line DNA (blood) from 16 patients with squamous cell carcinoma of the skin (Purdie, *et al* 2007), only one region of homozygosity larger than this size was found, suggesting this to be a reasonable criterion. Additionally, comparison was made with 62 long haplotypes that are described from trios (two parents and off-spring) from the HapMap project (International HapMap Consortium 2005). If any of these haplotypes were inherited from both parents, the child would be homozygous across that region. Only four long haplotypes were discovered that were larger than 7.7Mb; two of these all lay across centromeric regions, which are known to be less polymorphic, and one lay on the X chromosome which has been found to have a greater number of long haplotypes than other chromosomes.

DNA copy number was assessed by comparing the SNP signals of a panel of non-leukaemic DNAs, obtained from the germ line (DNA from venous blood) of a group of patients with squamous and basal cell carcinoma of the skin.

5.3 Results

Allowing for balanced translocations, which cannot be detected by SNP array analysis, there were 44 chromosomal aberrations that were confirmed by cytogenetic analysis. SNP array information gave an additional 48 abnormalities.

5.3.1 Mapping regions of LOH (Figure 5.1)

There were 68 regions of LOH detected in 42 patients. Of these, 27 were identified as deletions, 39 were UPDs and 2 were homozygous chromosomal gains. 9 UPDs extended to the telomere (telomeric UPDs), implying that a single mitotic recombination had occurred. Fifteen of the UPDs were interstitial, implying a double cross-over event. Six interstitial UPDs extended across the centromere of the chromosome, suggesting these may be germline events, i.e. homozygously inherited long haplotypes.

Both of the homozygous gains involved the long arm of chromosome 11. One of the gains affected two regions of 11q, and also involved a telomeric deletion, suggesting that this is actually a complex rearrangement. These homozygous gains may be due to loss of the other parental chromosome, or due to the relative increase in the number of copies of one of the parental chromosomes compared to the other. The relatively high copy number may have affected the MPAM algorithm that calculates the allelotype. Using an alternate algorithm (Huentelman, *et al* 2005), heterozygous calls were detected, suggesting that the alternate parental chromosome was not lost.

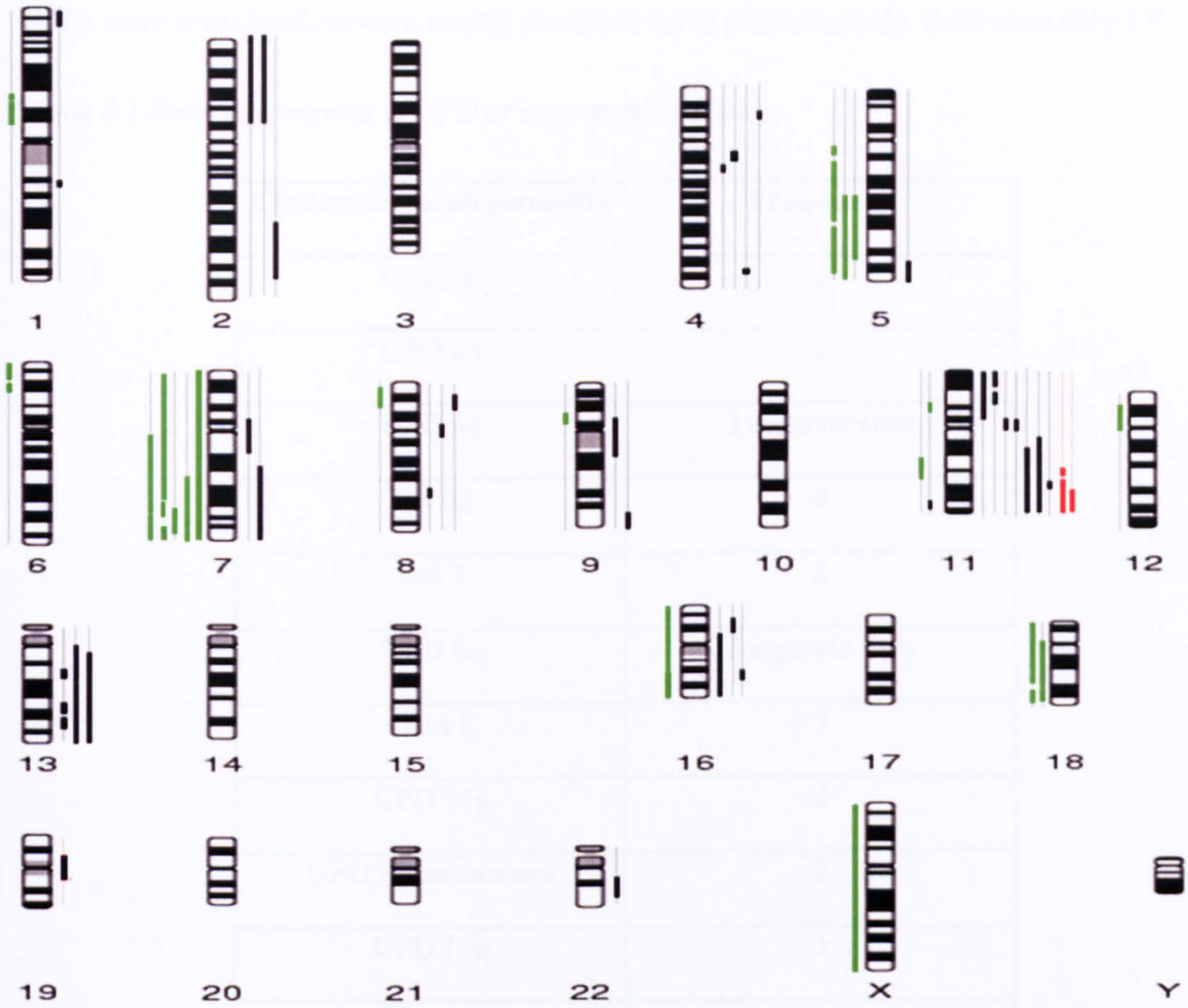


Figure 5.1 Map of homozygous regions greater than 30 consecutive SNPs in length. Green represents deletions, black represents disomy and red represents gains. Details of each abnormality are in Table 5.2.

Recurrent regions of UPD are shown in Figure 5.1. Regions of UPD that had previously been observed in the previous study included chromosome 11, both p and q arms, and chromosome 13. The regions of UPD11p included the *WT1* locus, which had been mutated in other AML samples with UPD11p. Similarly, the regions of UPD13 included the *FLT3* locus, suggesting these samples may have mutations of *FLT3*. New regions identified with recurrent UPDs included chromosome 2p in two patients and 16q in two patients. The location of the point of recombination was close to each other for both the AML samples with UPD2p. The

prevelance of UPD in this group of patients was 32%, however if only patients with telomeric UPDs were considered, as were mostly described in the previous study, there were only 13%.

Table 5.1 Recurrent regions of UPD or copy number change

Chromosomal abnormality	Frequency
Gain 1p	2
UPD 2p	2
UPD 4q	3 separate sites
Del 5q	3
Del 7	5
UPD 8q	3 separate sites
Gain 8	8
UPD 11p	2
UPD 11 centromere	2
UPD 11q	3
Gain 11q	2
Gain 12	2
UPD 13	4
Gain 13	4
UPD 16q	2
Del 18	2
Del 20q	3

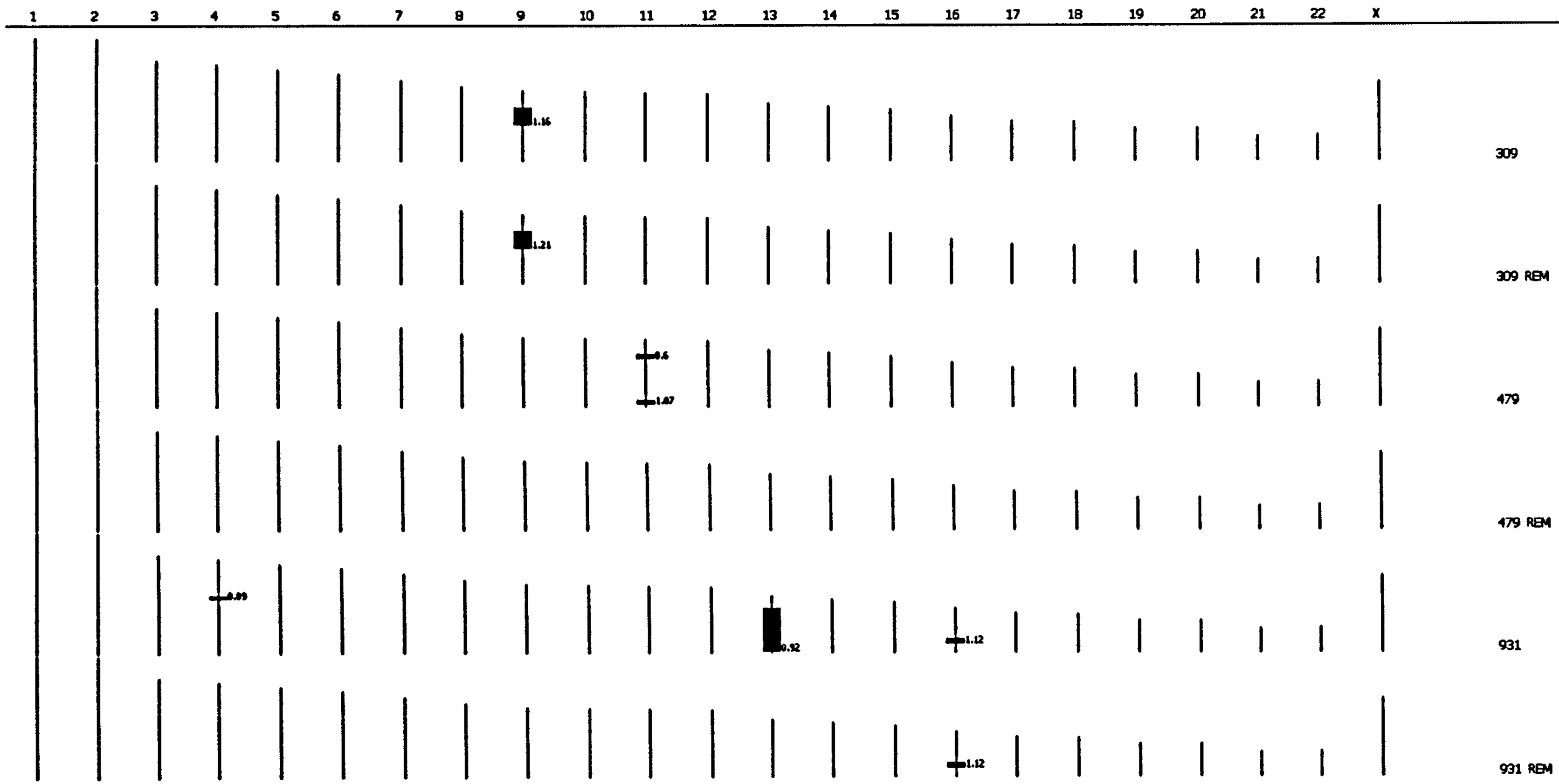


Figure 5.2 Regions of homozygosity in diagnosis versus remission (REM) pairs. Green represents deletion, black represents disomy, with the bars representing regions of homozygosity, each of at least 30 contiguous homozygous SNPs.

Although the regions of interstitial UPD were large, without germline material, it would be difficult to establish if they are somatically acquired. Three examples of paired diagnosis and remission samples are shown in Figure 5.2. Acquired regions of homozygosity are shown in patient 479 on chromosome 11, and in patient 931 on chromosomes 4 and 13. Chromosomes 11 and 13 in the two patients show clear regions of LOH for many informative loci across the regions of homozygosity. Chromosome 13 includes the locus for *FLT3*, and the deletion on chromosome 11 includes the locus for *WT1*. However, for chromosome 4, although the length of the region of homozygosity is more than 30 SNPs, on comparison with the germ line, there is only a single SNP with LOH. With only one SNP that is informative, we cannot rule out noise in the genotype calling, so it is possible that this is not an acquired region of UPD. Chromososome 9 and 16q on 309 and 931 respectively are demonstrated to be germline regions of homozygosity. The former is across the centromere of chromosome 9, which suggests this could be a long haplotype. Patient 931 was found to have a deletion of 16q on cytogenetic analysis at diagnosis. SNP array analysis found no fall in copy number, although there was a region of homozygosity observed. This suggests that the 16q aberration is more complex than a simple deletion. The remission sample does not have the 13q UPD abnormality, but retains the 16q abnormality, which suggests either the 16q aberration is in the germline, or else there is a persisting clonal population in the marrow at remission. Individuals with congenital deletions of this region (which extends from 16q21 to 16q22.3) are known to have microcephaly and learning disabilities amongst other problems (Monaghan, *et al* 1997), but whether this patient had this syndrome is unknown.

5.3.2 Mapping of copy number changes (Figure 5.3)

Deletions smaller than 30 SNPs could be identified by identifying regions with a fall in copy number superimposed on a region of homozygosity. Except for the two regions on chromosome 11 discussed above, no other regions of copy number gain were associated with LOH. Combined with the homozygosity map (Figure 5.1), the map of copy number alterations

reveals regions of deletions commonly found on cytogenetic analysis of AML i.e. del(5)(q), del(7)(q), -7 and del(20)(q). Another recurrent deletion was on chromosome 18, which occurred in two samples. One patient had a deletion of 17p that included the *TP53* locus, which is sometimes observed in AML and MDS.

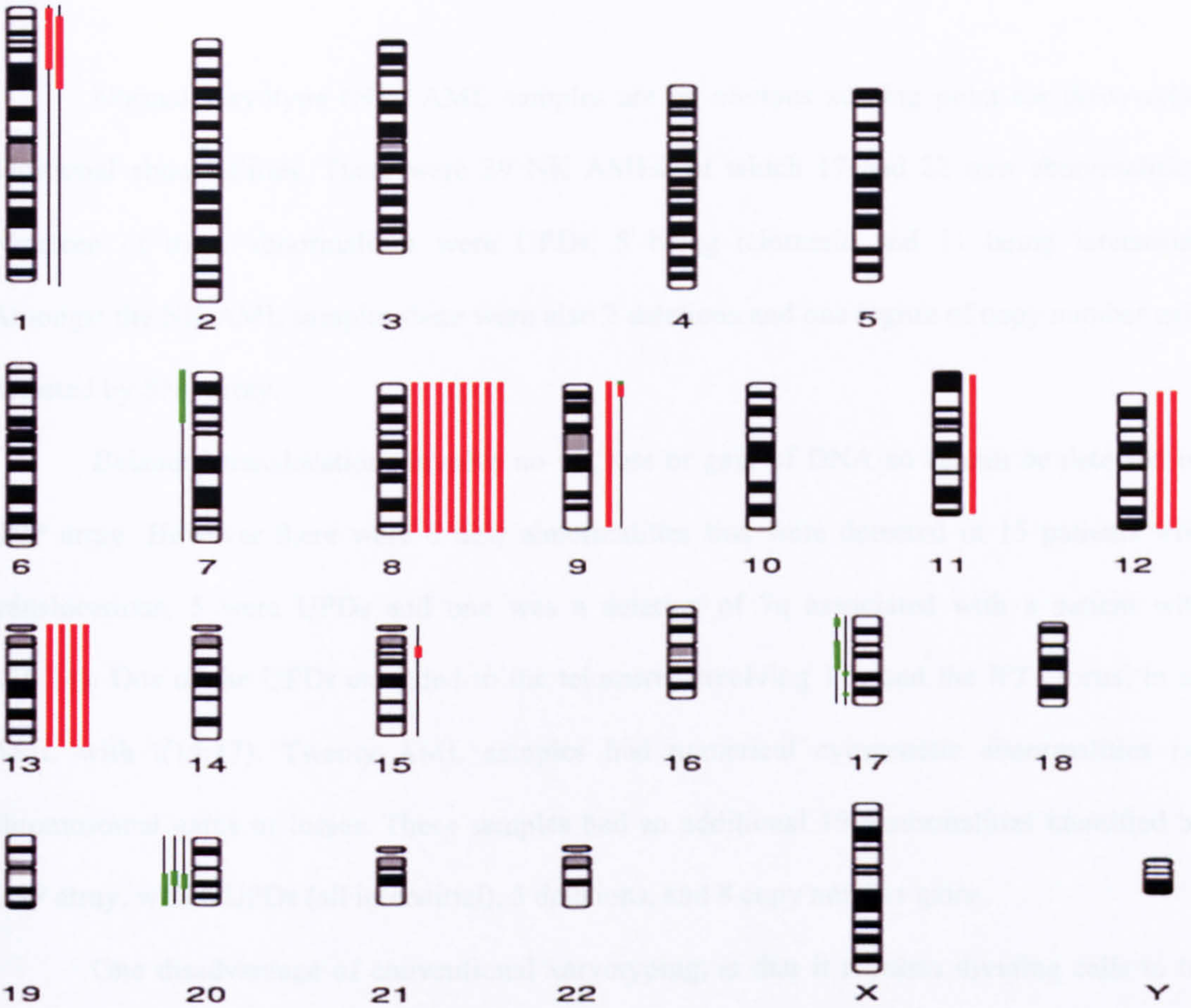


Figure 5.3 Map of DNA copy number by chromosome location.

Green represents deletions, red represent copy number gains. Note the deletions are either smaller than 30 SNPs or include some heterozygous calls, so were not included in figure 2. Details of each abnormality are in Table 5.2.

Recurrent copy number gains occurred on chromosomes 8, 13 and 12. Trisomies of these chromosomes are well recognised in AML (NCI and NCBI 2001). Gain of the short arm

of chromosome 1 was seen in two patients. The cytogenetic abnormality add(1p36) has been recurrently observed in AML (National Cancer Institute 2006) and this locus is included within this region.

5.3.3 Additional abnormalities discovered by SNP array, compared to cytogenetic analysis

Normal karyotype (NK) AML samples are an obvious starting point for discovering additional abnormalities. There were 39 NK AMLs, of which 17 had 22 new abnormalities. Nineteen of these abnormalities were UPDs, 8 being telomeric and 11 being interstitial. Amongst the NK AML samples there were also 2 deletions and one region of copy number gain detected by SNP array.

Balanced translocations involve no net loss or gain of DNA so cannot be detected by SNP array. However there were 6 new abnormalities that were detected in 15 patients with translocations; 5 were UPDs and one was a deletion of 7q associated with a patient with t(10;11). One of the UPDs extended to the telomere, involving 11p and the *WT1* locus, in an AML with t(15;17). Twenty AML samples had numerical cytogenetic abnormalities i.e. chromosomal gains or losses. These samples had an additional 19 abnormalities identified by SNP array, with 6 UPDs (all interstitial), 5 deletions, and 8 copy number gains.

One disadvantage of conventional karyotyping, is that it requires dividing cells to be able to analyse them in metaphase. If the cells die or fail to grow on culture then there will be no result. Thirteen samples had no karyotype information. By SNP array, 7 abnormalities were detected in 6 of these AML samples; 2 were deletions and 5 were UPDs.

A few abnormalities discovered by cytogenetic analysis were not found on SNP array analysis, i.e. patients 513, 551 and 692. In some cases this could be due to rearrangements that involve no copy number change (e.g. der(4)(p) in patient 692), and in others there may have been a rare clone with the identified cytogenetic abnormality. The SNP array can only detect copy number changes in the majority clone, because it analyses DNA altogether.

In summary, the SNP arrays were able to identify more abnormalities than conventional

karyotyping, although cytogenetic analysis complemented SNP array analysis by identifying balanced translocations and minor clones. UPDs were discovered in several different classes of cytogenetic abnormality, but were most often observed in NK AML.

5.4 Discussion

The map produced from this pilot set of data validates the use of SNP arrays to discover new regions with recurrent UPDs. UPDs were identified in 32% of patients (28 patients), which is a little higher figure than the original study. However, only by increasing the number of patients in this study can a more accurate assessment of the prevalence of UPD in AML be made.

The map is validated by the detection of known cytogenetic abnormalities in AML. It can therefore be assumed that new abnormalities are likely to be genuine. There are recurrent regions of UPD that were observed in the previous study, i.e. UPD 13q, 11p and 11q. Other UPDs previously described, 6p, 9p, 19q and 21 were not present suggesting they may be less frequent. However many additional UPDs are seen in this study, particularly the interstitial UPDs. For larger regions such as 2p, 2q, 7q and 16q, it is unlikely that these would be germline abnormalities. However, some of the smaller regions may be homozygously inherited long haplotypes, particularly those at the centromeres. The remission samples show that homozygous regions can be inherited and acquired. The only way to identify recurrent inherited regions of homozygosity would be to analyse data from high-resolution SNP arrays on a population of normal individuals. Some large studies have been performed using SNP arrays to look at haplotype structure (International HapMap Consortium 2005) and structural variation in the human genome (Redon, *et al* 2006), but not yet from the point of view of identifying regions of homozygosity, and not yet with sufficient individuals to identify recurrent regions.

Given the results from the previous study, it seems likely that UPD of chromosome 13 will be associated with a homozygous *FLT3* ITD mutation and UPD of 11p will be associated with a homozygous WT1 mutation. Although the regions of homozygosity are all large, other

regions of UPD may also encode homozygous mutations. With a larger study, it may be possible to identify minimal regions of homozygosity to narrow down the region that contains the gene of interest. This study has now been completed with over 450 AML samples genotyped (Gupta, *et al* 2008).

Because *FLT3* mutations have important prognostic implications, it will be interesting to see if any of the homozygous mutations have prognostic importance. However, the numbers of individual types of UPD is relatively small (only 3.5% for UPD13), therefore a large number of samples will be required to detect a statistically significant difference in survival. Given a 10% difference in survival between *FLT3* ITD positive and negative patients at five years, one might anticipate a smaller difference between heterozygous and homozygous *FLT3* ITD mutations, which will make this difficult to demonstrate.

If the prevalence of heterozygous and homozygous *FLT3* ITD mutations is known, one may be able to predict the overall prevalence of other mutations with UPD. If one assumes the rate of mitotic recombination is constant for all mutations (which it may be if there is a physiological background rate) then by knowing the prevalence of homozygosity, the prevalence of heterozygosity can be calculated. This would not work if there is more than one mechanism leading to homozygosity, e.g. non-disjunction and mitotic recombination. This may be useful where the underlying gene is unknown.

UPDs were seen in all the cytogenetic groupings, although there were a larger number in NK AML patients. With a lack of cytogenetic abnormalities, NK AML patients have only single gene defects, although there are likely to be many as yet unidentified genetic abnormalities e.g. microdeletions or amplifications. Given the association of acquired UPD with single gene mutations, it is perhaps not surprising that NK AML patients would have more UPDs, and mitotic recombination will of course maintain a normal karyotype.

Several genetic changes other than UPD were found that were not observed by cytogenetics. Some of these e.g. loss of chromosome 7 or 20q, would have conferred a poor prognosis on these patients. SNP arrays therefore have the potential to identify prognostically important cytogenetic abnormalities. With higher resolution arrays now available (over 1

million loci across the genome), this opens up the possibility of identifying even smaller somatic changes. The system is robust, using only small amounts of DNA, does not require live cells, unlike expression profiling, and is complementary to cytogenetics in identifying complex rearrangements. It therefore has a potential clinical use, and if UPD is found to be of prognostic value this would increase the need to apply this technology to the diagnostic workup of AML.

5.5 Appendix

Table 5.2 Location of chromosomal abnormalities.

Red indicates abnormalities revealed by SNP array and not by cytogenetic analysis, or vice- versa. c=centromere, tel=telomere.

Patient No.	Gains	Deletions	UPD	Original Karyotype
4			13 whole	46,XX
4			19q 16.6-40.1Mb	46,XX
83		7 whole		N/A
103			16q c-tel	46,XX
117			8p 36.6-53.9Mb	46,XY
117			8q 120.4-126.9	46,XY
137		7q 127.0-150.2Mb		46,XY,t(10;11)
164	13 whole	22q 26.1-43.9Mb		47,XY,+13
169			11p 47.7Mb-tel	N/A
221			8q 101.5-112.2Mb	46,XX
230			13 whole	46,XX
309			9p 31.8Mb-C	46,XX,inv(9)c
314	11q 73.7-94.6Mb, 104.85-131.4Mb	11q 131.4Mb-tel		44,X,-Y,add(1)(q?),-5,del(7)(q22),8p-,add(11)(q?),-16,

314		16 whole		add(17)(p),del(20)(q),-21,+mar1,+mar2/45,idem,+mar3
314		20q 29.5-43.8Mb		44,X,-Y,add(1)(q?),-5,del(7)(q22),8p-,add(11)(q?),-16
314		5q 83.2-95.85Mb		,add(17)(p),del(20)(q),-21,+mar1,+mar2/45,idem,+mar3
314		7q 102.8Mb-tel		44,X,-Y,add(1)(q?),-5,del(7)(q22),8p-,add(11)(q?),-16,
314		8p 30.4M-32.6M		add(17)(p),del(20)(q),-21,+mar1,+mar2/45,idem,+mar3
339		1p 76.3-103.1Mb (79.8-80.75)		44,X,-Y,add(1)(q?),-5,del(7)(q22),8p-,add(11)(q?),-16,
479		11p 29.5-40.4Mb	11q 130Mb-tel	add(17)(p),del(20)(q),-21,+mar1,+mar2/45,idem,+mar3
482			16p 12.7-21Mb	44,X,-Y,add(1)(q?),-5,del(7)(q22),8p-,add(11)(q?),-16,
482			8p 9.0-27.2Mb	add(17)(p),del(20)(q),-21,+mar1,+mar2/45,idem,+mar3
488			4q 181.9-178.7Mb	44,X,-Y,add(1)(q?),-5,del(7)(q22),8p-,add(11)(q?),-16,
490			11p 39.0-tel	add(17)(p),del(20)(q),-21,+mar1,+mar2/45,idem,+mar3
502	8 whole			44,X,-Y,add(1)(q?),-5,del(7)(q22),8p-,add(11)(q?),-16,
506			1p 30.6-tel	add(17)(p),del(20)(q),-21,+mar1,+mar2/45,idem,+mar3
506			1q 156.7-163.3Mb	44,X,-Y,add(1)(q?),-5,del(7)(q22),8p-,add(11)(q?),-16,
510		7q 118.0-134.1Mb		add(17)(p),del(20)(q),-21,+mar1,+mar2/45,idem,+mar3
				46,XY
				N/A
				46,XX,inv(16)
				46,XX,inv(16)
				46,XX
				46,XY,t(15;17)
				46,XX/47,XX,+8
				46,XX
				46,XX
				46,XX,inv(7)(p15q36),inv(16)(p13q22)

513	8 whole			46,XY/47,XY,+8,-11,+der(11),t(11;?)(q23;?)
544			3p 46.1-56.9Mb	N/A
549	8 whole			46,XX/47,XX,+8
551		20q c-40.3Mb		46,XX,21s+c/34-37,XX,-1,-3,-6, -8,-11,-12,-13,-15,
581	8 whole			46,XY,t(15;17)/47,XY,+8,t(15;17)
587			11q 62.8Mb-tel	46,XX
587			2p t-81.5Mb	46,XX
590	11p 45.8-57.7			46,XY
622		-7p c-tel		46,XY/46,XY,-7,+mar/45,XY,-7
664			4p 25.8-34.7Mb	45,X,-X,t(8;21)
664		-X whole		45,X,-X,t(8;21)
690			9q 118.1-132.8	46,XX
692	8, +9, +13 whole	-7 whole		50,XY,+8,+9,+13,+mar/50,XY,der(4)(q?), +8,+9,+13,+mar
706	21 16.7-tel	17p 3.90Mb-12.5Mb		46,XY/46,XY,del(5)(q?),-7,+8,-17,-18,del(20)(q?),+mar, +mar
706	8, +11, +12 whole	-18p c-tel		46,XY/46,XY,del(5)(q?),-7,+8,-17,- 18,del(20)(q?),+mar,+mar
706		-20q c-tel		46,XY/46,XY,del(5)(q?),-7,+8,-17,-18,del(20)(q?),+mar, +mar

706		-5 99.6Mb-tel		46,XY/46,XY,del(5)(q?),-7,+8,-17,-18,del(20)(q?),+mar, +mar
706		-7q c-tel		46,XY/46,XY,del(5)(q?),-7,+8,-17,-18,del(20)(q?),+mar, +mar
717			2p 81.9Mb-tel	46,XY
729			5q 159.0Mb-tel	46,XY
768	13 whole		16p 12.8-28.0Mb	46,XX
770			7q 90.4Mb-tel	46,XY
775	8, +12, +13 Whole			47,XY,+8
777	11q 91.7Mb-tel	-18 whole		43,with structural abnormalities
777	13, +21 whole	-5q 67.9Mb-tel		43,with structural abnormalities
777	1p 46.7Mb-c	12p t-36.5Mb		43,with structural abnormalities
777	2, +3, +4 whole	6p t-16.2Mb		43,with structural abnormalities
777	9p 3.64-18.6Mb	8p t-32.2Mb		43,with structural abnormalities
777		9p 27.5-38.3Mb		43,with structural abnormalities
810			11q 104.5-112.4Mb	46,XX,t(8;21)
854			11q 74.4Mb-tel	N/A
867			4q 61.9-71.5Mb	
878	15p t-37.0Mb		17q 41.4-50.0Mb	46,XX,del(16)(q?)
878	1p t-56.0Mb			46,XX,del(16)(q?)

878				46,XX,del(16)(q?)
884			15q 76.8-84.0Mb	46,XX,t(8;21)(q22;q22)
884			11p 45.6-c-57.5Mb	46,XX,t(8;21)(q22;q22)
900			2q 76.8-224.8Mb	46,XY
900			7p 45.3-77.8Mb	46,XY
927	1p t-73.6Mb			46,XY/46,XY,ins(2)(p?)del(11)(q?)
927		11q 84.8-105.9Mb		46,XY/46,XY,ins(2)(p?)del(11)(q?)
931			16q 61.0-72.2Mb	46,XX,del(16)(q?)
931			13q 25.3Mb-tel	46,XX,del(16)(q?)
931			4q 74.1-83.2Mb	46,XX,del(16)(q?)

Chapter 6

Segmental UPD is a commonly acquired genetic event in relapsed AML

6.1 Introduction

AML is an aggressive hematological malignancy characterized by a proliferation of immature myeloid cells. Dramatic advances have been made over the last 40 years in the treatment of AML such that approximately 30% of those presenting under the age of 60 years can be cured with initial therapy. Complete remission, determined clinically, morphologically, immunologically and sometimes at the molecular level is reliably reported to be achieved in up to 80% of patients in this age group (Hann, *et al* 1997), but despite this success, re-emergence of leukemia is the rule rather than the exception. The 5 year relapse rate is 50% for patients less than 60 years, and increases to 80% in the over 60s (Goldstone, *et al* 2001, Grimwade, *et al* 1998). Registry data from our institution similarly demonstrates that recurrence occurs in two-thirds of those achieving first complete remission. The prognosis of relapsed AML is dismal, with only 20% of patients surviving at 4 years (Milligan, *et al* 2006).

Relapsed leukemic cells are characteristically resistant to chemotherapy. AML classified as poor risk by cytogenetic profile often has *in vitro* resistance to one of the most important chemotherapeutic drugs, cytosine arabinoside, which may explain the high risk of relapse (Zwaan, *et al* 2000). However, the underlying pathological mechanisms leading to relapse are poorly understood. The main pointers towards reasons for relapse are the numerous clinical and biological prognostic risk factors. Clinical features such as increasing age, high white cell count

at diagnosis and an initial failure to respond to chemotherapy are established poor risk factors. The amount of residual disease after chemotherapy also increases the risk of relapse.

Cytogenetic abnormalities are the most important biological factor in predicting risk of relapse, as detailed in chapter 1, but only three case series have reviewed cytogenetic changes at diagnosis and relapse (Estey, *et al* 1995, Garson, *et al* 1989, Kern, *et al* 2002). Frequently, karyotypic abnormalities present at diagnosis were the same at recurrence, and no consistent cytogenetic abnormalities were acquired on the return of the AML. Therefore, it is not clear what role acquired karyotypic abnormalities play in the pathogenesis of relapse.

Single gene mutations can also predict for risk of relapse, with *FLT3* internal tandem duplication (ITD) being an important predictor of poor prognosis (Kottaridis, *et al* 2001). *FLT3* ITD, observed at diagnosis, can be increased in the level of DNA to above 50% in some cases, suggesting the mutation may have become homozygous (Kottaridis, *et al* 2002). Homozygous single gene mutations do occur in AML and these are associated with mitotic recombination or possibly a non-disjunction event leading to acquired uniparental disomy (UPD) (Fitzgibbon, *et al* 2005). This suggests a mechanism for relapse in AML in which a heterozygous mutation is followed by a recombination event leading to homozygosity and clonal evolution. To evaluate this possibility, this study used 27 paired diagnosis and relapse AML samples, and array-based genome-wide mapping to identify regions of segmental UPD, with further mutation analysis.

6.2 Methods

6.2.1 Sample Selection

Sufficient diagnostic and relapse blood and bone marrow samples for analysis were available from 27 patients with AML. Ethical approval was obtained and samples stored as detailed in Chapter 2. Because samples were chosen based on their availability, there was a bias towards samples with relatively high white blood cell counts (WBC). Patients presented between April 1983 and September 2004. Patient demographics are shown in *Table 6.1*. Median age was 51

(range 35 to 68) years; 18 were male and 9 female. The median number of days to relapse was 179 (range 88-758) days. The median overall survival of this group of patients was 1 year, which was inferior to those of all patients that had relapsed presenting to St. Bartholomew's Hospital during this period (1.5 years, p=0.001). Standard cytogenetic analysis was performed using bone marrow or peripheral blood. Metaphase preparations were harvested, and Giemsa banding was performed (Czepulkowski, *et al* 1992). Patient karyotypes were described according to the International System for Human Cytogenetic Nomenclature (Mitelman 1995). By MRC criteria (Grimwade, *et al* 1998), at diagnosis two patients had a favorable risk karyotype, twenty had intermediate risk karyotypes, one had a poor risk karyotype and two were unknown.

Table 6.1 Demographic data

Patient	Sample number		Sex	Age at diagnosis	FAB type	Presentation WBC, ×109/l
	Diagnosis	Remission				
1	GSM173388	GSM173415	F	25	M5	89.0
2	GSM173389	GSM173416	M	47	M1	182.0
3	GSM173390	GSM173417	F	60	M2	10.4
4	GSM173391	GSM173418	M	27	M1	65.9
5	GSM173392	GSM173419	M	35	M4	79.7
6	GSM173393	GSM173420	F	67	M5	127.2
7	GSM173394	GSM173421	M	61	M4	11.3
8	GSM173395	GSM173422	F	47	M2	53.1
9	GSM173396	GSM173423	F	47	M4	88.1
10	GSM173397	GSM173424	F	41	M1	9.0
11	GSM173398	GSM173425	M	47	M1	90.0
12	GSM173399	GSM173426	M	68	M2	19.8
13	GSM173400	GSM173427	M	59	M1	88.4
14	GSM173401	GSM173428	M	64	M4	31.2
15	GSM173402	GSM173429	M	48	M5	33.2
16	GSM173403	GSM173430	F	36	M4	30.7
17	GSM173404	GSM173431	F	65	M1	4.4
18	GSM173405	GSM173432	M	52	M4	5.5
19	GSM173406	GSM173433	M	52	M2	11.3
20	GSM173407	GSM173434	M	46	M4	127.2
21	GSM173408	GSM173435	M	43	M4	11.4
22	GSM173409	GSM173436	M	61	M2	29.1
23	GSM173410	GSM173437	M	50	M4	5.7
24	GSM173411	GSM173438	F	65	M1	6.2
25	GSM173412	GSM173439	M	25	M5	96.8
26	GSM173413	GSM173440	M	21	M5	51.8
27	GSM173414	GSM173441	M	43	M5	43.6

6.2.2 10K GeneChip Assay

This was performed as previously described in Chapter 2. The data discussed have been deposited in the National Cancer Biotechnology Information (NCBI) Gene Expression Omnibus (GEO)(2005) and are accessible through GEO Series accession number GSE7210. Analysis of the call and copy number data was performed using GOLF. Regions of LOH were determined by comparing diagnostic and relapse samples. SNP copy number analysis identified these regions as being deleted, gained or as being disomic (Bignell, *et al* 2004). SNP and gene annotatations used NCBI genome build 35.

6.2.3 Mutation analysis

Mutation screening was done for the entire coding region of *CEBPA* (Snaddon, *et al* 2003) and specific exons of *FLT3* (exons 14-15 and 20) (Abu-Duhier, *et al* 2001, Kiyoi, *et al* 1999) as described in Chapter 2 .

PCR products were sequenced directly by use of an ABI 377 DNA sequencer (PE Applied Biosystems, Foster City, CA). Unincorporated primer was removed by ultrafiltration using a Centricon YM-100,000 filter device (Millipore). Sequencing data were analyzed using 4Peaks (Griekspoor and Groothuis 2005).

Fragment analysis of *FLT3* exon 14-15 PCR products was performed using the same conditions but using QIAGEN Multiplex PCR kit, and measured by capillary electrophoresis using an ABI Prism 3700 Genetic Analyser (PE Applied Biosystems). Fluorescence signals were analysed using Genotyper 3.7 software (PE Applied Biosystems).

6.3 Results

6.3.1 Clonal evolution of segmental UPD at relapse

Clonal evolution resulted in segmental UPD and a homozygous mutation at relapse but not present in the paired diagnostic sample in about 40% (11/27) of cases. The most common abnormality detected was segmental UPD of chromosome 13 in 6 of 27 patients (22%) (Table 5.1). The size of UPD varied from about 88Mb including the telomere, to the whole chromosome (114Mb). Copy number analysis revealed two copies of chromosome 13 across the region of LOH, excluding the possibility of a hemizygous deletion and included the locus for *FLT3*. Sequencing of *FLT3* showed an ITD mutation, which was heterozygous at diagnosis and homozygous at relapse in all six cases (Table 6.3 and Figure 6.1). The mutations were identical at diagnosis and relapse in each case, demonstrating that the second clone had evolved from the initial leukemia.

Patient 16 acquired segmental UPD of 19q at relapse, and was associated with a mutation of *CEBPA* located within the region of homozygosity. This mutation was also heterozygous at diagnosis and homozygous at relapse. The mutation, 207 C→T (Genbank accession Y11525), created a stop codon which results in a truncated protein (Snaddon, *et al* 2003). A further segmental UPD was acquired at 4q in patient 21. There is likely to be an associated homozygous mutation in the region of UPD, but in view of the large region involved (104Mb), this was not investigated further.

Chromosome 13

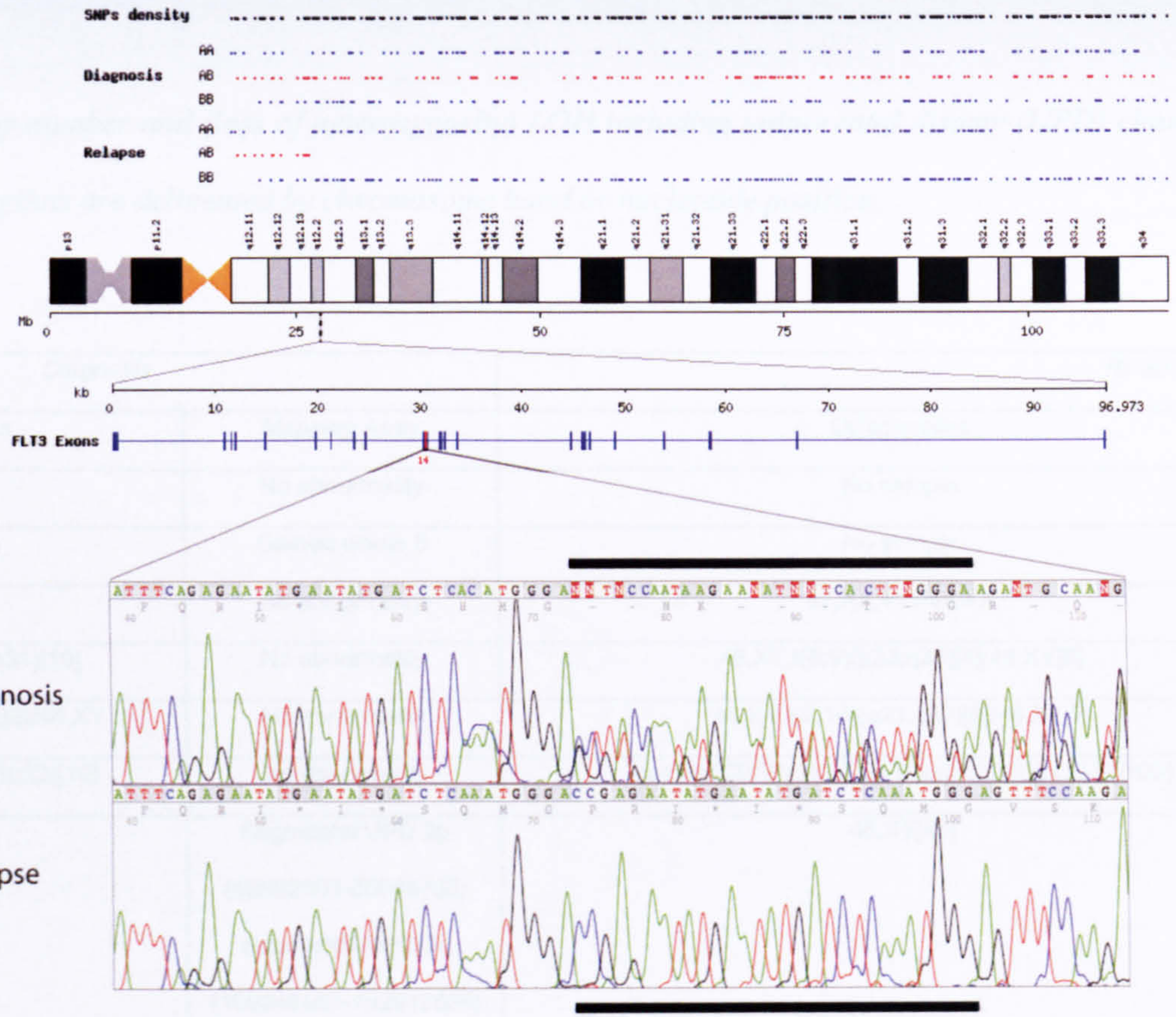


Figure 6.1 Chromosome 13q segmental UPD acquired at relapse in patient 4. LOH is shown by a change from heterozygosity (alleles AB in red) to homozygosity (alleles AA or BB in blue). An identical mutated sequence of FLT3 ITD is shown in the region of UPD at q12.2 at diagnosis and relapse, with a change from homozygosity to heterozygosity.

Table 6.2 Chromosomal, copy number and (loss of heterozygosity) LOH including uniparental disomy (UPD) changes between diagnosis and relapse. Each change is marked in bold. Regions are delineated by chromosome band or nucleotide position.

Patient	*Diagnosis*		*Relapse*	
	Cytogenetics	*Mapping Array*	*Cytogenetics*	*Mapping Array*
1	46,XX	No abnormality	No sample	**LOH whole 13**
2	No sample	Gained whole 6	No sample	Gained whole 6
3	46,XX	No abnormality	47,XX,**+11**/46,XX	**Gained whole 11**
4	46,XY,t(6;9)(p23;q34)[10]	No abnormality	46,XY,t(6;9)(p23;q34)[4]/46,XY[6]	**Segmental UPD13q (262590909-term)**
5	46,XY,t(2;14)(q21;q32)/46,XY	No abnormality	46,XY,t(2;14)(q21;q32)[8]/46,XY[2]	No abnormality
6	46,XX,inv(16)(p13.1q22)[10]	No abnormality	nuc ish(CBFBx2)(5'CBFB sep 3'CBFBx1)[82/100]	**Gained whole 22**
7	46,XY	Segmental UPD 2p (45892803-66988762) Segmental UPD 5q (109245567-132912666)	46,XY[47]	Segmental UPD 2p (45892803-66988762) Segmental UPD 5q (109245567-132912666) **LOH whole 13**
8	46,XX,del(16)(q12.1)/46,XX	Segmental UPD 13q (19029353-term) Del 16q (29142199-85631136)	46,XX,**del(11)(q13q22)**,del(16)(q12.1)[2]/46,idem,**t(4;6)(q35;q15)**[8]	**Del 4q (67619803-93329691) Del 11q (85287757-115209411)** Segmental UPD 13q (19029353-term) Del 16q (29142199-85631136)

9	46,XX[25]	No abnormality	46,XX[25]	**Segmental UPD whole 13**
10	46,XX	No abnormality	46,XX	**Segmental UPD 13q (19029353-term)**
11	46,XY	No abnormality	46,XY	No abnormality
12	45,X,-Y,t(8;21)(q22;q22)[10]	Segmental UPD 11 (44160735-62518998)	45,X,-Y,t(8;21)(q22;q22)	**Del 2q (198963363-240642812)** Segmental UPD 11 (44160735-62518998)
13	46,XY,del(7)(q11q21)/46,XY	No abnormality	No sample	No abnormality
14	46,XY	No abnormality	46,XY	**LOH whole 13**
15	46,XY[20]	No abnormality	47,XX,+8,add(9)(q34)[8]/46,XY[2]	No abnormality
16	46,XX [25]	No abnormality	No sample	**Segmental UPD 19q (20533429-term) Gained whole 21**
17	47,XX,+8[10]	No abnormality	47,XX,del(6)(q13q21),der(7)t(7;17)(q11.1;q11.2),+8,add(17)(q11.2), del(20)(q13.1q13.3 or q11.2q13.1)[5]/47,idem,del(2)(q13q2?)[5]	**Del 6q (64252743-101001308) Del 7q (69512516-term) Segmental UPD whole 13 Del 20q (40318090-110881790)**
18	46,XY[20]	No abnormality	46,XY[20]	No abnormality
19	46,XY,t(6;9)(p23;q34)[6]/46,XY[55]	No abnormality	46,XY,t(6;9)(p23;q34)[10]	**Segmental UPD 13q (18482385-term)**
20	46,XY[20]	Gained 1p (term-59198211)	46,XY[20]	No abnormality
21	46,XY	No abnormality	46,XY	**Segmental UPD 4q (87522823-term)**
22	46,XY,t(8;9)(q24;q32)[4]/ 45,sl,-Y,der(9)add(9)(p13)t(8;9)[11]/ 45,sdl1,add(14)(q32)[3]	Del 9p (19871864-25253918)	No sample	Del 9p (19871864-25253918)

23	No sample	No abnormality	No sample	No abnormality
24	46,XX[20]	No abnormality	46,XX,t(1;5)(q4;q11.2),?t(11;17)(q13;q11.2),del(12)(p11.2p13)[3]/46,XX[15]	No abnormality
25	46,XY	No abnormality	46,XY	Segmental UPD 13q (20860773-term)
26	49,XY,+5,+8,+19[9]/46,XY[1]	Gained whole 5 Gained whole 8	49,XY,+5,+8,+19[1]/50,idem+mar[2]/46,Y,der(X),t(X;13)(q26;q12),add(7)(p11.2),-13,+mar[4]	Del 1p (5708940-38792696) Gained 3p (125054444-term) Del 7p (term-61521113)
27	47,XY,+8/46,XY	Gained whole 8	No sample	Gained whole 8

***Table 6.3** Associated mutation and acquired segmental uniparental disomy (UPD) or loss of call (ITD, internal tandem duplication).*

Patient	Segmental UPD acquired at relapse	Associated Mutation Status	
		Diagnosis	Relapse
1	Loss of call whole 13	*FLT3* ITD heterozygous	*FLT3* ITD heterozygous
4	13q	*FLT3* ITD heterozygous	*FLT3* ITD homozygous
7	Loss of call whole 13	D835Y heterozygous	D835Y homozygous
9	13q	*FLT3* ITD heterozygous	*FLT3* ITD homozygous
10	13q	*FLT3* ITD heterozygous	*FLT3* ITD homozygous
14	Loss of call whole 13	D835Y heterozygous	*FLT3* wild type
16	19q	*CEBPA* 207 C→T heterozygous	*CEBPA* 207 C→T homozygous
17	Whole 13	*FLT3* ITD heterozygous	*FLT3* ITD homozygous
19	13q	*FLT3* ITD heterozygous	*FLT3* ITD homozygous
21	4q	Unknown	Unknown
25	13q	*FLT3* ITD heterozygous	*FLT3* ITD homozygous

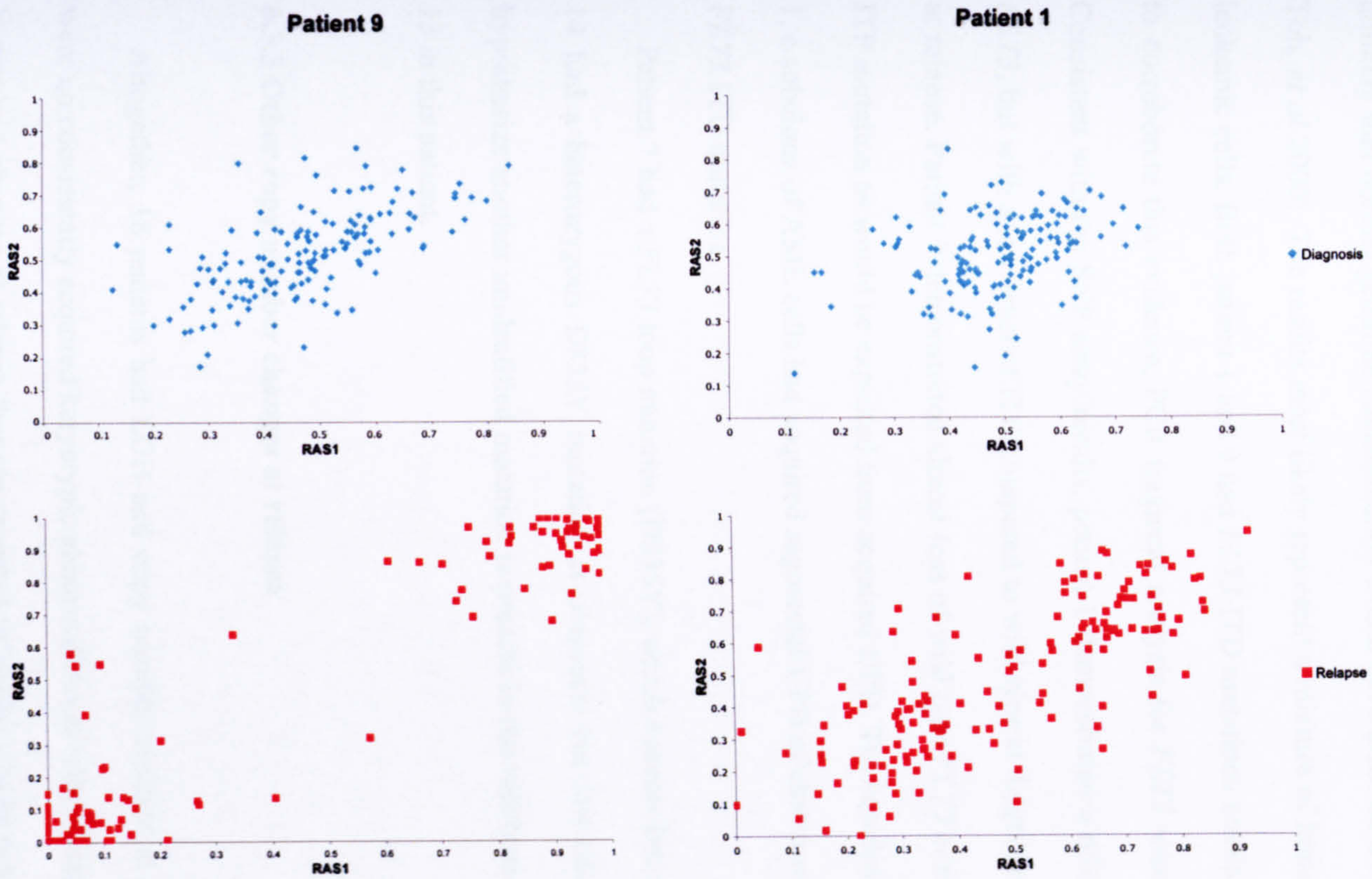
6.3.2 Subclones with segmental UPD

Three of the eleven patients (1, 7 and 14) acquired UPD of chromosome 13 at relapse as defined by LOH with no change in copy number. However, they were distinguished from the other six cases with acquired UPD of chromosome 13 because their heterozygous calls did not become homozygous, but instead became no-calls. To understand how this may occur, a rudimentary understanding of the SNP calling algorithm is required. The Affymetrix 10K SNP

array uses an algorithm that compares the relative allele signal (RAS) from each allele on the forward and reverse strands (RAS1 and RAS2 respectively) to determine heterozygosity or homozygosity for individual SNPs (Liu, *et al* 2003). No-calls are made when the RAS values do not determine the allele. If a proportion of cells within a sample are heterozygous and a proportion is homozygous at a given SNP locus, then the algorithm will be unable to assess homozygosity or heterozygosity and report a no call. Patients 1, 7 and 14 each had a high frequency of no-calls along chromosome 13 at relapse.

Figure 6.2 A. Change in relative allele signal (RAS) values between diagnosis and relapse, for SNPs heterozygous at diagnosis in the region of LOH on chromosome 13 for patient 1 and 9. B. Electropherogram demonstrating PCR fragment quantitation of FLT3 wild type, at 329 bp and the larger ITD fragment in patients 1 and 9.

A.


Patient 9
Patient 1
RAS2
RAS1
Diagnosis
Relapse

B.

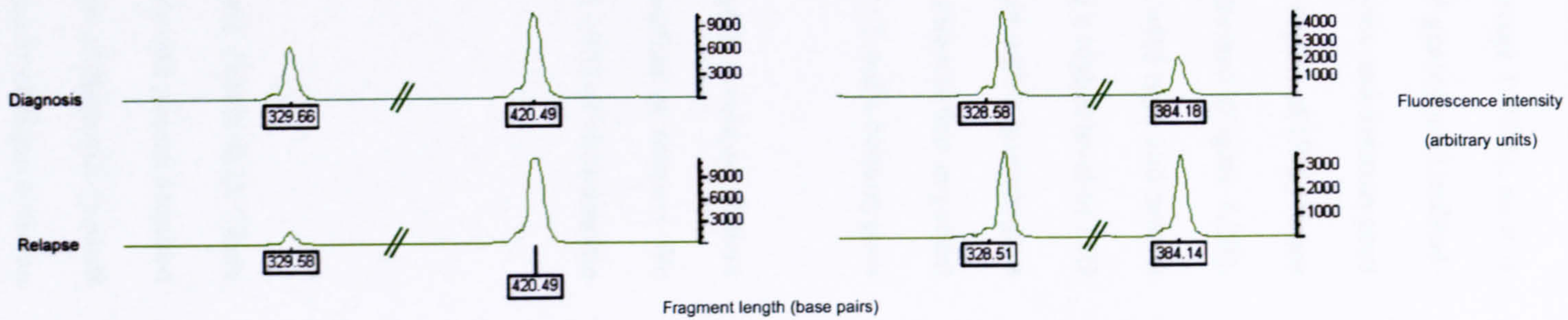

Diagnosis
Relapse
329.66
420.49
329.58
420.49
328.58
384.18
328.51
384.14
Fluorescence intensity
(arbitrary units)
Fragment length (base pairs)

Figure 6.2A displays the RAS values for patient 9 who has conventional LOH due to UPD, compared to patient 1. For patient 1, the RAS values are observed to move towards the values for homozygosity, but insufficiently, and therefore the SNPs are reported as no-calls. This demonstrates that the no-calls are due to a change in the proportion of alleles. The allele call rates for all other regions were greater than 90%, confirming that the no-calls were not due to a failure of hybridization. These AML samples all had a high blast percentage (>70%), so it is unlikely that normal germline contamination would have affected the SNP genotypes(Lindblad-Toh, *et al* 2000). The results most likely represent a mixture of homozygous and heterozygous leukemic cells. Both patient 1 and 9 had *FLT3* ITD mutations within the region of UPD, hence to corroborate this evidence, PCR fragment analysis for *FLT3* was performed (Figure 6.2B). Consistent with the SNP array results, patient 1 demonstrated a mixture wild type and mutant *FLT3*, but with a low level of ITD compared to wild type at diagnosis and a higher level of ITD at relapse. Patient 9 demonstrated almost loss of wild type *FLT3* consistent with a homozygous ITP mutation as would be expected from acquired UPD. The conclusion drawn is that in patient 1, a subclone of AML cells had acquired segmental UPD of chromosome 13 and a homozygous *FLT3* ITD mutation.

Patient 7 had a *FLT3* loop mutation (D835Y), which became homozygous at relapse. Patient 14 had a heterozygous D835Y mutation at diagnosis but lost this mutation at relapse. We hypothesize another unidentified mutation is present in the subclone with UPD of chromosome 13 in this patient.

6.3.3 Other copy number changes at relapse

Altogether, 18 patients had LOH and copy number changes at relapse (Table 6.2). There were no consistently acquired karyotypic abnormalities at relapse, and only one patient acquired karyotypic changes at relapse that are reported to confer a poor prognosis at diagnosis (patient 17, deletions of chromosome 7q and 20q). Seven patients had no detectable changes between diagnosis and relapse. Patient 20 had a gain of chromosome 1p at diagnosis that was lost at

relapse.

Patient 24 had a rare clone with a deletion of 12p that would go undetected by SNP array. Patient 15 also had cytogenetic changes undetected by SNP array, but the cytogenetic abnormalities were found in the dominant clone. We hypothesize that the dominant leukemic clone had a normal karyotype, but did not grow on short-term culture for karyotypic analysis.

6.3.4 Chromosomal abnormalities at diagnosis (Table 6.2)

Most of the regions of deletion and gain were confirmed by cytogenetic analysis. Those not confirmed were deletions that were either too small to be seen on cytogenetic analysis, or were regions of uniparental disomy (UPD). Regions of whole chromosome gain were also seen that corresponded with trisomy in the karyotype. Patients 8, 14, 17 and 24 had balanced translocations that cannot be detected by SNP array.

There were four regions of UPD at diagnosis, one that was confirmed to be acquired from genotype analysis of the remission sample (patient 13), and another had homozygosity for *FLT3* ITD, indicating acquired UPD of 13q (patient 8). The other UPDs were much larger and not reported as large haplotypes blocks in the HapMap (International HapMap Consortium 2005), suggesting they are due to mitotic recombination, although homozygosity across the centromere of chromosome 11 (patient 12) is probably germline because of a lack of meiotic recombination near centromeric regions.

6.4 Discussion

This study has revealed the importance of acquired UPD in relapsed AML. Acquired segmental UPD resulted in a homozygous mutation, most commonly involving chromosome 13 with *FLT3* ITD. Interestingly, previous studies have shown that loss of wild type *FLT3* in the context of *FLT3* ITD is a poor prognostic marker at diagnosis (Thiede, *et al* 2002, Whitman, *et al* 2001). At presentation, a higher level of *FLT3* ITD on PCR fragment analysis of genomic

DNA also confers an increased relapse risk (Gale, *et al* 2005), and segmental UPD of chromosome 13 provides a mechanism for this increase. Acquired segmental UPD is the most consistent genetic change in relapse, therefore it is important step in disease progression.

The FLT3 receptor is a homodimer that is activated by FLT3 ligand at the cell surface. The *FLT3* ITD mutation constitutively activates the receptor leading to disregulation of intracellular cell signaling pathways (Griffith, *et al* 2004). One explanation for the advantage gained by a homozygous *FLT3* mutation may be that a homodimer with two mutated proteins has more activity than if one part is wild type. Homozygosity of the *FLT3* D835Y loop mutation at relapse lends further support to the hypothesis that loss of wild type leads to relapse. Loss of a loop mutation without an ITD mutation was also observed associated with acquisition of UPD of 13q at relapse. This suggests there may be mutations of *FLT3* outside of the conventional regions, or alternatively, other mutated genes on 13q. Homozygous *FLT3* mutations would be ideal candidates for tyrosine kinase inhibitors. Following acquisition of UPD of chromosome 13, inhibition of *FLT3* may be an effective method of treating relapse in conjunction with other therapy, or a method of preventing relapse if used as maintenance after consolidation chemotherapy.

Biallelic mutations of *CEBPA* are well recognized in AML, and are occasionally homozygous (Lin, *et al* 2005). Although a biallelic mutation is not a prerequisite for AML, it is important in its development, as demonstrated in familial cases of AML in which a germ line *CEBPA* mutation is followed after a long latency by a second *CEBPA* mutation and subsequently develops AML (Smith, *et al* 2004b). AML patients with mutations of *CEBPA* have a relatively good prognosis (Leroy, *et al* 2005). In this study, evolution of a homozygous *CEBPA* mutation was associated with relapsed AML, but further studies will be required to discover if homozygous *CEBPA* mutations have any prognostic implications.

Mitotic recombination or non-disjunction leads to segmental or whole chromosomal UPD respectively. It is unknown whether these events occur before, during or after treatment. A rare subclone may be present in the initial population at diagnosis, but this would be extremely difficult to detect given the small proportional change in alleles that would result. In favor of

this theory, there are *in vitro* studies on mouse fibroblasts and normal human lymphocytes that have shown a physiological background rate of mitotic recombination of about 1 in 100 000 cells (de Nooij-van Dalen, *et al* 1997). The alternative is that mitotic recombination occurs after exposure to cytotoxic chemotherapy, due to an increased rate of DNA double strand breaks. The relapse clone would then arise during or shortly after treatment, leading to a short time to relapse. With this small study it was not possible to make statistical comparisons in the time to relapse between those acquiring UPD at relapse and the rest of the AML patients.

When segmental UPD clonally evolves at relapse, it infers that there is a proliferative advantage gained by the cell from the homozygous mutation. One model for clonal selection suggests that these cells are already present at diagnosis, and exhibit drug resistance, surviving through chemotherapy. The alternate model is that a leukemic stem cell, which is intrinsically more resistant to chemotherapy than differentiated cells, escapes chemotherapy induced death and subsequently undergoes mitotic recombination.

The evolution of subclones with segmental UPD, seen here in three cases, is consistent with the clonal heterogeneity often seen with cytogenetic abnormalities in AML. Indeed, clonal heterogeneity is a feature of many cancers (Merlo, *et al* 2006). It raises the possibility that segmental UPD is under detected, since a low proportion of homozygous cells cannot be observed using SNP array technology, and as discussed above, low levels of segmental UPD at diagnosis would have the potential to lead to relapse.

The frequency of acquired segmental UPD at diagnosis is 15-20% (Gorletta, *et al* 2005, Raghavan, *et al* 2005) and this study appears to demonstrate an increased frequency of UPD (40%). The study is limited by being a retrospective sample set and does not represent all subtypes of AML. However, it is still clear that mitotic recombination is an important means of disease progression. The phenomenon has been observed in one case of childhood AML involving *FLT3* ITD (Bungaro, *et al* 2006), and in a further case of a WT1 mutation (Nyvold, *et al* 2006). Several other cancers have demonstrated acquired UPD at diagnosis (Andersen, *et al* 2007, Teh, *et al* 2005), and these may also progress by mitotic recombination. Although segmental UPD of 13q is the commonest abnormality described here, further studies into

acquired segmental UPDs will reveal other targets in the treatment of AML and other cancers.

Chapter 7

Discussion

This study has used a new technology, microarray based genotyping, to demonstrate the importance of an old concept, mitotic recombination. The observation that acquired UPD is a frequent finding in AML, and its association with homozygous mutations, particularly at disease recurrence, gives an insight into the natural history of the progression of AML.

7.1 Genotyping arrays, cytogenetic analysis and copy number

Unlike many studies using aCGH or SNP arrays looking at copy number in solid tumours, in AML it is fortunate that a karyotype is available. Comparison can then be made with the expected copy number changes. Interpretation of deletions and gains could then be made. For example, patient 69 had trisomy 3 and 10 and the copy number ratio graph shows an increase compared to the line representing a copy number ratio of one (Table 3.3).

These interpretations are particularly useful where there is more variability or noise. Firstly, SNP genotyping tends to be noisier in interpreting copy number at any individual SNP compared with aCGH because the size of the oligomers is so much smaller (25 nucleotides compared with several kilobases) leading to variation in hybridisation. Secondly, where a sample was processed with its germ-line remission, the kinetics of the PCR are similar, leading to a more comparable ratio. Thirdly, comparing like with like (i.e. AML to germ-line) is preferable to having to compare against a bank of normal DNAs. The kinetics of the PCR

between germ line and leukaemia are much more similar for a single individual; as others have noted, there are differences in the amplification of genomic regions, because of differences in the GC content of each region (Nannya, *et al* 2005). If there was no available remission sample, then a bank of normal DNAs was required, to reduce the effects of copy number polymorphisms, although these have more implications for higher resolution genotyping arrays i.e. 250K than 10K arrays.

Although G-banding can identify gross chromosomal abnormalities, small abnormalities cannot be detected. Such small deletions are seen in patient 78 at 5p, in patient 53 at 1p, and in patient 23 at 7q. The median inter-SNP distance on the 10K array is 113kb (Affymetrix 2003), so LOH of even a few SNPs can cover a megabase, and involve several genes. Recent studies using higher resolution arrays in ALL have demonstrated recurrent microdeletions involving genes of B cell development e.g. *PAX5*, amongst others (Mullighan, *et al* 2007). It is likely many more of these microdeletions will be discovered in AML when using higher resolution arrays, and the pattern of genes involved may reveal much about the pathogenesis of AML.

Other abnormalities are coincident with described karyotypic abnormalities, but suggest more complexity in the rearrangement. For example, GSM173409 had a derivative chromosome 9, with additional material and a translocation with chromosome 8. However, the array shows a deletion of part of the short arm of chromosome 9, giving more information as to the nature of the rearrangement. Similarly, patient 34 has a deletion of the whole of chromosome 7 by cytogenetic analysis, but the array shows part of the long arm of the chromosome adjacent to the centromere is retained. The remaining part was found by FISH on a marker chromosome. Small chromosomal aberrations such as these are difficult to see by conventional cytogenetic analysis, but the SNP array genotyping makes a useful tool to identify complex chromosomal rearrangements.

7.2 UPD and homozygosity

Mitotic recombination is a well recognised concept; one of the initial descriptions was

of its association with twin spots occurring in the fruit fly, *Drosophila melanogaster* (Stern 1936). The fruit flies were heterozygous for genotypes that could cause phenotypic changes in skin colour and hair type, but had mosaic spots whose phenotype was of the recessive genotype. Stern showed that a mitotic cross-over event had led to homozygosity for that genotype. A similar phenomenon is seen in some vascular skin naevi (Koopman 1999), where each adjacent spot is homozygous for a different gene that is on the same chromosome arm.

In cancer, the recombination leads to outgrowth of one of the clones because of a selective advantage from an underlying homozygous abnormality. The homozygously mutated genes associated with acquired UPD in this study include *CEBPA*, *WT1*, *FLT3* and *RUNX1* mutations. All of these were previously known to be mutated in AML (King-Underwood, *et al* 1996, Smith, *et al* 2005). This still leaves several examples of acquired UPD without a known associated mutation e.g. UPD 6p, 4q, 9p and 11q. It should not be overlooked that even if there is no mutated gene associated with these regions, it is still possible that there is an associated gene with altered transcription due to a homozygous epigenetic effect, e.g. promoter methylation.

Identifying genes associated with UPD could be a difficult task. The regions of homozygosity are tens or hundreds of megabases long, so there are hundreds of potential target genes. However, it is notable that for the regions where the associated gene is known, the point of recombination is frequently less than 10-15Mb away from the gene, despite the region of LOH being several times larger than this (an example of this is seen in Figure 6.1). However, this is still a large region from which to identify a candidate gene. Another method for gene discovery would be to identify highly or lowly expressed genes in the region of interest, compared with a leukaemia without acquired UPD. Array based gene expression would make the analysis of a large number of genes in a region of LOH a practical proposition. However, it may be neither of these methods can identify a gene, and it may be left to future developments in high throughput sequencing to identify mutations in a region that is several megabases in length. Despite the difficulties, there are likely to be novel genes involved in the pathogenesis of AML in regions of acquired UPD. The completion and publication of a map of UPD in AML

samples from the national MRC trials should reveal common recurrent regions of UPD and regions that are more rare. One use for the map could be to prompt groups who have genes of interest within regions of UPD to investigate these further.

7.3 Homozygous versus heterozygous mutations

In this study, the main consequence of acquisition of UPD in AML was to homozygously inactivate a TSG or increase the gene dosage of an oncogene. Conventionally, regions of LOH have been thought to harbour TSGs, but in this study oncogenes, *FLT3*, and TSGs, *CEBPA*, *RUNX1* and *WT1* have been found in regions of UPD.

Both the ITD and TKD mutations of *FLT3* are activating mutations that cause constitutive phosphorylation of the receptor. Homozygosity increases the gene dosage in AML cells, suggesting that increased transcription of the mutant *FLT3* causes more activation of the downstream cell-signalling pathway. However, the FLT3 ITD mutant receptor activates wild type FLT3 receptor as a heterodimer in AML cell lines (Kiyoi, *et al* 1998), so it would appear to be unnecessary to have both alleles mutated. One could speculate that there are degrees of activation of the FLT3 receptor pathway, and a homozygous mutation may increase the pathway activation.

CEBPA is a transcription factor that has two types of mutation in AML, both of which inactivate the protein. N terminal mutations cause deficient transactivation and are dominant negative in vitro, while C terminal mutations affect the zinc finger DNA binding domain (Pabst, *et al* 2001). Frequently the mutations are biallelic, with one N- and one C-terminal mutation on each allele. These mutations can be seen in the context of *CEBPA* being a TSG. However, monoallelic heterozygous mutations are well recorded, as was seen at diagnosis in the case that acquired UPD19q at relapse. This suggests that there are varying degrees of inactivation of *CEBPA* and that there is a selective advantage for biallelic and homozygous mutations.

RUNX1, another transcription factor, has also been described as a TSG. Mutations are found in its DNA binding domain, indicating loss of function. In familial platelet disorder, it

causes haploinsufficiency, as there is only a single allele affected (Song, *et al* 1999), but in *de novo* AML, especially FAB type M0, there are frequent biallelic mutations (Roumier, *et al* 2003, Roumier, *et al* 2006, Silva, *et al* 2003) (although the example in this study had FAB M2). Again, although a monoallelic mutation is sufficient to lead to AML, in some circumstances, a biallelic mutation is selected for.

WT1 mutations were first described in association with Wilms' tumours. It is another transcription factor in which most mutations associated with AML are in its zinc finger domain, suggesting they would cause loss of function and reduce the expression of downstream transcription targets. Most mutations are heterozygous, implying the mutations cause haploinsufficiency (King-Underwood, *et al* 1996). Congenital heterozygous mutations of WT1 in the same region cause renal diseases such as Denys Drash syndrome, also indicating that a haploid mutation is sufficient to cause disease (Gao, *et al* 2004). However, a few homozygous mutations have been described in AML (Nyvold, *et al* 2006), as have been noted associated with acquired UPD 11p in this study.

What is common amongst these mutations is that there is a difference in the effect of monoallelic and biallelic mutations on the leukaemic cell, so that at least in some AML samples, gene dosage confers a selective advantage. The implication is that for oncogenes and TSGs alike, the genes are not simply "on" or "off", but there are gradations, probably reflecting more or fewer downstream targets of the genes being affected.

A graduated evolution of genetic abnormalities in a single pathway suggests an increasing dependence on that pathway by the leukaemic cell, a concept that has been termed oncogene addiction (Weinstein and Joe 2006). *In vitro* evidence for oncogene addiction comes from mouse models and cell lines in which over expression of an oncogene such as MYC induced T cell lymphomas and AML, but which regressed with MYC inhibition (Felsher and Bishop 1999). Further clinical evidence has been cited from CML, where targeting a single oncogenic protein, BCR/ABL, with the tyrosine kinase inhibitor imatinib has a marked clinical response (Kantarjian, *et al* 2002). Furthermore, accelerated phase and blast crisis tend to be associated with imatinib resistance due to mutations in *BCR/ABL* (Talpaz, *et al* 2006),

demonstrating the importance of retaining this oncogenic pathway.

However, for most forms of malignancy, including AML, the concept of addiction to a single pathway seems counter-intuitive. Combination chemotherapy affects more than one pathway and that has led to the improvement in treatment for AML and other malignancies. However, the evidence from this study suggests that in patients who acquire UPD at relapse, there is an increased dependence on the pathway associated with the homozygous mutation, implying targeting the pathway would be treat the recurrent AML. For one of these pathways, downstream of the FLT3 receptor, inhibitors are being used in clinical trials. Small phase 1/2 trials of FLT3 receptor inhibitors have been tried with multiply treated, relapsed or refractory AML patients with *FLT3* mutations (Smith, *et al* 2004a, Stone, *et al* 2005). Some of the patients have had a haematological response, but only a few obtain a complete remission. It is just as likely that other pathways are also involved in relapse, and discovering these will be critical to improve the targeted treatment of relapsed AML.

7.3.1 Transgenic animal models of MR and homozygous mutations

Further investigation of the functional difference between AML with homozygous and heterozygous mutations may require animal models. Conventional methods of insertional mutagenesis involve inserting several copies of a mutation into a cell, either of a cell line, or of an embryonic stem cell in order to create a transgenic animal. The wild type copy is retained and the mutant alleles could be inserted anywhere into the genome. This is more analogous to gene amplification of one allele, rather than the situation that occurs with single gene mutation with or without MR.

A transgenic animal model of MR has been described using mouse embryonic stem cells (Liu, *et al* 2002). In this experiment, two complementary halves of the *HPRT* gene (3'and 5'), each in a cassette between two *loxP* sites, were targeted to the same allelic position of complementary chromosomes. Each part of the *HPRT* gene was non-functional without the other half. On addition of Cre recombinase, some cells would undergo mitotic recombination,

fusing the two halves of *HPRT* together. The cell colonies were selected for by growing them in HAT (hypoxanthine/aminopterin/thymine) medium, as cells with a complete *HPRT* gene would be resistant to HAT medium. The rate of mitotic recombination observed was between $7x10^{-3}$ and $3.5x10^{-4}$, depending on the chromosomal location targeted.

A similar system using mouse embryonic stem cells heterozygous for a *TP53* mutation and using a constituitively active FLP recombinase has been used study the development of tumours in transgenic mice produced from the embryonic stem cells (Wang, *et al* 2007). The tumours produced in these transgenic mice were homozygous for mutant *TP53*, demonstrating the originating cells had undergone mitotic recombination. In addition, an increased number of epithelial tumours were observed compared with transgenic mice that were heterozygous for *TP53*. Hypothetically, if a heterozygous mutant mouse model was created for one of the genes known in AML instead of *TP53*, i.e. *FLT3*, *RUNX1*, *WT1* or *CEBPA*, then the phenotypic effect of mitotic recombination on these genes could be observed. Although a transgenic *FLT3* ITD mouse exists (Lee, *et al* 2005), this involved random insertion of the mutant gene, which would not be practical for this type of experiment.

7.4 Clonal evolution

The generally accepted model of clonal progression of neoplasia is that there is a sequence of genetic or epigenetic changes each leading to a selective (growth) advantage to the neoplastic clone (Nowell 1976). Evidence for this model comes from colorectal tumours where a sequence of genetic abnormalities was seen in adenomas and adjacent carcinomas from patients with familial adenomatous polyposis (Vogelstein, *et al* 1988). In this study, with increasing histological grade of adenoma and carcinoma, an increasing percentage of tumours with *RAS* mutations, chromosome 18q deletions and chromosome 17p deletions were observed.

This model predicts that it is the dominant clone, with the greatest selective advantage, that makes up the tumour. However, the development of oesophageal cancer shows much greater genetic heterogeneity than would be predicted by this model alone (Maley, *et al* 2006,

Maley, *et al* 2004). Patients with Barrett's oesophagus, a premalignant condition, have an increased risk of developing the malignant form of disease, squamous carcinoma of the oesophagus. The patients were monitored by sequential biopsies of the squamous epithelium of the oesophagus. Increased clonal diversity in these biopsies, as defined by more regions with LOH, *TP53* mutations or *CDKN2A* mutations, lead to an increased risk of developing carcinoma. This implies development of a malignant clone comes about on the background of clonal heterogeneity. It explains why more than one clone with slightly different karyotypes may be observed on cytogenetic analysis. It also explains why acquired UPD was observed as a subclone in some cases. Some of these genetic aberrations may be carrier or hitchhiker mutations, as has also been observed in Barrett's oesophagus (Maley, *et al* 2004), however, recurrent aberrations are likely to be important in malignant progression e.g. acquired UPD 13.

Given the heterogeneous genetics of tumour development, there are likely to be additional mutations or UPDs in rare clones during leukaemogenesis. These may be selected for by chemotherapy e.g. there is some evidence that *FLT3* ITD mutations confer resistance to chemotherapy (Seedhouse, *et al* 2006). One implication of this hypothesis is that if rare clones with acquired UPD are identified at diagnosis, this would predict for the subsequent relapse of these patients.

An important aspect to acquired UPD is that its development occurs on the background of a prior gene mutation. Mitotic recombination (or non-disjunction) is therefore a stepwise progression in the development of the leukaemia. The current model of leukaemogenesis, in which there is a mutation preventing differentiation (class 1), followed by a mutation increasing proliferation (class 2) (Kelly and Gilliland 2002), although useful in classifying types of mutation, is likely to be simplistic in terms of the progression of the leukaemia. Mitotic recombination could be just one of a number of progressive genetic events leading to leukaemogenesis. Mouse models in which haematopoietic progenitors are transduced with CBFbeta-SMMHC (a class 1 mutation) and *FLT3* ITD (a class 2 mutation) produce AML 3-5 months after transplantation, suggesting further mutations are required during the latent period (Kim, *et al* 2007).

7.5 Mitotic recombination versus non-disjunction

Most of the examples of acquired UPD are partial, involving a chromosome arm, so the mechanism involved is mitotic recombination. Some appear to involve the whole chromosome e.g. chromosome 13, suggesting not recombination, but non-disjunction, i.e. a failure of the chromosomes to separate equally at mitosis, as the mechanism leading to UPD. However, the most telomeric SNP on the arrays is still several megabases from the telomere, so a recombination point close to the telomere cannot be ruled out. Meiotic recombination is thought to be more common close to the telomeres, because linkage disequilibrium is low in these regions (International HapMap Consortium 2005), but mitotic recombination may not necessarily occur in the same regions as meiotic recombination. AML cells are known to undergo non-disjunction, resulting in trisomic chromosomes. It is worth noting that the SNP arrays show an increased copy number but no LOH for these trisomies, implying non-disjunction, with both parental chromosomes retained. There were no examples where one of the parental chromosomes was reduplicated. This has also been noted in hyperdiploid ALL where several chromosomes are trisomic (Paulsson, *et al* 2005, Paulsson, *et al* 2003). It suggests the importance of retaining both parental chromosomes i.e. avoiding loss of imprinting (LOI), however, acquired UPD does result in LOI. LOI of genes, such as *IGF2*, has been associated with tumorigenesis in other malignancies such as colorectal cancer (Sakatani, *et al* 2005). Although imprinting was not important for UPD of 11p, its role in other regions of UPD cannot be ruled out.

AML also undergoes the opposite event i.e. monosomy, particularly for chromosomes 5 and 7. Loss of chromosome 5 or 5q is associated with reduced expression of *CTNNA1* by methylation or histone deacetylation (Liu, *et al* 2007). In acquired UPD, both chromosomes appear to be identical. The only apparent difference between UPD and monosomy is copy number, implying an advantage in retaining a diploid karyotype. For genes such as *CTNNA1*,

whose methylation may be important in the pathogenesis of AML, mitotic recombination would an alternative to chromosome loss. There are likely to be many other genes whose epigenetic suppression is important in AML, and this approach has led to the use of demethylating agents, e.g. decitabine (Kantarjian, *et al* 2007), and histone deacetylase inhibitors, eg. valproic acid (Kuendgen, *et al* 2005), in MDS/AML.

Some epigenetically suppressed genes will be in regions of acquired UPD, therefore depending on which allele becomes homozygous, their expression would be altered. Hypothetically, some of these genes may contribute to leukaemogenesis. However, one main advantage of acquired UPD over monosomy is the maintenance of a diploid genome, allowing a physiological expression level, because gene dosage affects expression level (Schoch, *et al* 2006). Genes that do not maintain a normal expression level in the region of UPD could be important in the pathogenesis of AML.

7.6 The mechanism of mitotic recombination

A crossover event during mitosis is not a physiological event, unlike during meiosis. It implies a DNA double strand break (DSB) has taken place and then an incorrect repair. There are two major processes in double strand break repair (DSBR), non-homologous end joining (NHEJ) and homologous recombination (HR) (O'Driscoll and Jeggo 2006). NHEJ is the predominant method of repair in mammalian cells, but is error prone because it involves the ligation of the ends without a template. However, it is quick and can occur at all stages of the cell cycle (Sengupta and Harris 2005). HR is much more accurate because it uses a template with a matching sequence for repair. There are two models for homologous repair; strand displacement annealing (SDSA) does not involve a crossover event, whereas the DSBR model has the possibility crossing over or not. Two 4-way DNA strand junctions are formed, known as Holliday junctions (HJs). Depending on how the HJs are resolved depends on whether a crossover event occurs. Crossovers are not unusual in eukaryotic systems, which are the most studied (Prado, *et al* 2003). In mammalian cells, if a double strand break is induced, it can be

repaired using the homologous chromosome by allelic recombination or gene conversion, causing LOH (Moynahan and Jasin 1997). Short and long tract gene conversion is seen, with the former involving only the gene, and the latter extending much farther downstream. This is a potential mechanism for the MR seen in AML. However, in mammalian cells DSBR by HR usually occurs by sister chromatid exchange (Sung and Klein 2006, Takata, *et al* 1998), which is at the G2/S phase of the cell cycle, rather than with the homologous chromosome. It has also been seen that although heterologous chromosomes with similar sequence can be used to repair DSBs, crossovers are markedly suppressed, to prevent translocations (Richardson, *et al* 1998). This suggests there are normally strong mechanisms to prevent chromosomal aberrations due to crossovers.

Helicases are enzymes that can separate the two strands of a DNA double helix, and some are involved in regulating crossover in HR (Sung and Klein 2006). For example, BLM helicase is involved in resolving double Holliday junctions formed as an intermediary during DSBR by HR. Other helicases involved in HR have been described in yeast, e.g. sgs2, but their orthologues have not been found in mammalian cells. Congenital mutation of *BLM* leads to Bloom syndrome, an autosomal recessive disease. The children have a short stature, a predisposition to malignancy and photosensitivity (German 1997). Bloom syndrome is associated with an unstable karyotype, and the cells of these individuals have an increase in sister chromatid exchange as well as exchange between homologous chromosomes (Chaganti, *et al* 1974). A mouse model for Bloom syndrome also shows increased tumourigenesis and mitotic recombination (Luo, *et al* 2000). One may hypothesise that AML patients with UPD have defects in BLM or other RecQ helicases, leading to a mitotic recombination event. However, one would expect cells with defects in HR to have very disordered karyotypes, and many of the AML samples had a normal karyotype. A polymorphism in the gene may confer a subtler defect, increasing the risk of DNA DSBs whilst maintaining karyotype stability. There is some evidence that polymorphisms in *RAD51* (RAD51-G135C) and *XRCC3 (*XRCC3-241Met) predispose to *de novo* AML and therapy related AML (Seedhouse, *et al* 2004). Both of these genes are important in HR and DSBR. RAD51 is of particular importance as it assembles onto

single stranded DNA to form a presynaptic filament, prior to searching for homologous regions in double stranded DNA (Sung and Klein 2006). It would be interesting to observe whether patients with these polymorphisms who develop AML are more likely to have acquired UPD.

On top of inherited defects and polymorphisms in DNA repair, there are also acquired DNA repair defects that are prevalent in cancer cells. It has long been assumed that it is a prerequisite for cancer cells to have defects in DNA repair, which leads to the genetic instability observed in cancer. It has been proposed that the exposure to specific mutagens selects for cells with specific types of genetic instability (Breivik and Gaudernack 2004). Evidence for this comes from genetically stable colorectal cancer cell lines (Bardelli, *et al* 2001). Exposure of these cells to bulky-adduct-forming agents selects for cells with chromosomal instability i.e. gains and losses of chromosomes in subsequent generations. Conversely, cells from the same cell line selected after exposure to an alkylating agent develop microsatellite instability rather than chromosomal instability. The cells that survive in the presence of the mutagen no longer repair these DNA defects because it is time and energy consuming to do so. Therefore, they are able to proliferate, and although they accumulate mutations, as long as they are not deleterious, they will continue to proliferate.

AML cells prior to acquiring UPD have a heterozygous mutation that probably confers a proliferative advantage. If the marrow environment additionally confers an advantage to those cells that acquire a defect in HR, then this will increase the chance of mitotic recombination. However, because there is a background rate of mitotic recombination, at least in vitro, it may not be necessary for any defect in HR to be present. One way to determine whether there is a difference in HR between AML with UPD and without would be to observe the frequency of sister chromatid exchanges (SCE) as a surrogate for HR. As with Bloom syndrome this would be increased if there were a defect in HR. However, previous studies have shown no difference in SCE between normal individuals and AML cells (Jones, *et al* 1992). The variation in SCEs in AML is small (0.4-1.8 per cycle) compared with the high number in Bloom syndrome (45-162) (Chaganti, *et al* 1974), therefore it is unlikely this method would find a difference, and again suggests if there is any variation in HR that it is subtle.

The implication from experiments on colorectal cell lines is that the environment of the bone marrow is important in the development of genetic and cytogenetic abnormalities in AML. AML patients with chromosomal instability, i.e. gains and losses will have different DNA repair defects from those with translocations, single gene mutations and UPD. Examples of this are seen in therapy related AML where patients exposed to alkylating agents are at risk of AML with del(7q) and have other karyotypic abnormalities associated with chromosome instability, and those exposed to topoisomerase II inhibitors are at risk of AML with *MLL* and other translocations, which are defects associated with DSBR. The characteristics of the environment leading to *de novo* AML are unknown, but the nature of their genetic abnormalities and the nature of any DNA repair defects may give clues.

7.6.1 Breakpoints of mitotic recombination

The sequence at the breakpoints of mitotic recombination may give a clue as to the nature of the DNA break involved. This has been particularly important for secondary AMLs where the patient has been exposed to cytotoxic agents that can cause DSBs. An example of this has been described in therapy related APML, where a 8 base pair hotspot in *PML* was found to be the point of translocation, after exposure to mitoxantrone (Mistry, *et al* 2005).

Fragile sites throughout the genome have been documented by them forming breaks on addition of aphidocolin, a DNA polymerase inhibitor. There may be characteristics of DNA sequence that cause regions of the genome to be fragile and lead to a tendency for DSBs (Glover, *et al* 2005). Rare, Mendelian inherited fragile sites are associated with repeat sequences, but the only consistent feature of common fragile sites is an increase in AT rich sequences. Some (but not all) of the breakpoints of UPD do lie in fragile sites, e.g. the breakpoints on 11p associated with *WT1* mutations lie within the common fragile site FRA11E at 11p13.

7.7 Conclusions

Much of the investigation of the pathogenesis of AML has, not surprisingly focussed on the primary mutations in leukaemogenesis. However, the acquisition of UPD is a secondary step in the development of AML. Cancer has been considered a multi-step genetic disease, and the investigation of acquired UPD demonstrates not only which mutations are important, but also how they progress at the presentation of AML and at recurrence. It can be used as a model for the multi-step pathogenesis of cancer, and therefore further investigation of UPD, particularly at relapse, may have implications for our understanding of how cancers and leukaemia in particular develop.

However, mitotic recombination is only a mechanism in the development of the leukaemia, and it is the underlying homozygous mutation that is pathogenetic. The mutations are potential drug targets and markers for disease progression, and the discovery of novel regions of acquired UPD may reveal new targets and markers.

Appendix: Barts diagnostic AML sample details

Patient sample	Source	FAB type	Cytogenetics
AML sample 10 diagnosis paired with sample 44	AML blood/marrow cells diagnosis	M5	46,XX,dic(7;22),+8
AML sample 104 remission paired with sample 71	AML marrow cells remission	NA	46,XY
AML sample 105 remission paired with sample 14	AML marrow cells remission	NA	46,XX
AML sample 106 remission paired with sample 49	AML marrow cells remission	NA	46,XY
AML sample 107 remission paired with sample 35	AML marrow cells remission	NA	46,XY
AML sample 11 diagnosis paired with sample 12	AML marrow cells remission	NA	not analysed
AML sample 12 diagnosis paired with sample 11	AML blood/marrow cells diagnosis	M2	46,XX
AML sample 13 diagnosis paired with sample 90	AML blood/marrow cells diagnosis	M1	46,XY,t(8,21)[8]/46,XX
AML sample 131 remission paired with sample 72	AML marrow cells remission	NA	46,XY
AML sample 132 remission paired with GSM173397	AML marrow cells remission	NA	46,XX
AML sample 14 diagnosis paired with sample 105	AML blood/marrow cells diagnosis	M2	46,XX,t(8,21)(q22;q22)

AML sample 15 diagnosis	AML blood/marrow cells diagnosis	M2	46,XY,t(8;21)
AML sample 16 diagnosis paired with sample 89	AML blood/marrow cells diagnosis	M4	46,X,-X,t(8,21)(q22;q22)
AML sample 17 diagnosis	AML blood/marrow cells diagnosis	M2	46,XY,t(15;17)
AML sample 18 diagnosis	AML blood/marrow cells diagnosis	M4	46,XY,inv(16)(p13q22)
AML sample 19 diagnosis	AML blood/marrow cells diagnosis	M4	46,XY,dir ins(6;11)(q27;q13q23)
AML sample 20 diagnosis	AML blood/marrow cells diagnosis	M1	46,XX
AML sample 21 diagnosis paired with sample 22	AML blood/marrow cells diagnosis	M1	46,XY
AML sample 21 remission paired with sample 22	AML marrow cells remission	NA	46,XY
AML sample 23 diagnosis	AML blood/marrow cells diagnosis	M2	46,XX
AML sample 24 remission paired with sample 45	AML marrow cells remission	NA	46,XY
AML sample 26 remission paired with GSM173401	AML marrow cells remission	NA	46,XY
AML sample 27 diagnosis paired with sample 28	AML blood/marrow cells diagnosis	M6	46,XY
AML sample 28 remission paired with sample 27	AML marrow cells remission	NA	not analysed
AML sample 29 diagnosis paired with sample 30	AML blood/marrow cells diagnosis	M2	46,XX

AML sample 30 remission paired with sample 29	AML marrow cells remission	NA	46,XX
AML sample 300 diagnosis	AML blood/marrow cells diagnosis	M4	46,XY
AML sample 31 diagnosis	AML blood/marrow cells diagnosis	M1	46,XY,-9,+17
AML sample 311 diagnosis paired with sample 81	AML blood/marrow cells diagnosis	M3	46,XY,t(15;17)(q22;q21)[2]/46idem,ider17(q10)t(15;17)(q22;q21)[8]
AML sample 32 diagnosis paired with sample 96	AML blood/marrow cells diagnosis	M1	46,XY,+8
AML sample 34 diagnosis	AML blood/marrow cells diagnosis	M1	45,XY,-7,+mar. ish ider(7)add(7)(wcp7+,CEP7+,D75486-)
AML sample 35 diagnosis paired with sample 107	AML blood/marrow cells diagnosis	M1	46,XY
AML sample 36 diagnosis	AML blood/marrow cells diagnosis	M1	46,XX
AML sample 37 diagnosis	AML blood/marrow cells diagnosis	M2	46,XX
AML sample 38 diagnosis	AML blood/marrow cells diagnosis	M2	46,del(5)(q31)XY
AML sample 39 diagnosis	AML blood/marrow cells diagnosis	M1	46,XY
AML sample 40 diagnosis paired with sample 41	AML blood/marrow cells diagnosis	M2	46,XX
AML sample 41 remission paired with sample 40	AML marrow cells remission	NA	46,XX
AML sample 42 remission paired with sample 43	AML marrow cells remission	NA	not analysed

AML sample 43 diagnosis paired with sample 42	AML blood/marrow cells diagnosis	M1	46,XX
AML sample 44 remission paired with sample 10	AML marrow cells remission	NA	46,XX
AML sample 45 diagnosis paired with sample 24	AML blood/marrow cells diagnosis	M2	46,XY
AML sample 48 diagnosis paired with sample 93	AML blood/marrow cells diagnosis	M1	46,XX
AML sample 49 diagnosis paired with sample 106	AML blood/marrow cells diagnosis	M1	46,XY
AML sample 50 diagnosis	AML blood/marrow cells diagnosis	M1	46,XX
AML sample 51 diagnosis	AML blood/marrow cells diagnosis	M2	46,XY
AML sample 52 diagnosis	AML blood/marrow cells diagnosis	M1	46,XY
AML sample 53 diagnosis paired with sample 88	AML blood/marrow cells diagnosis	M0	47,XY,+19 and del 1x *ABL* homologue
AML sample 54 diagnosis	AML blood/marrow cells diagnosis	M4	46,XY
AML sample 56 diagnosis	AML blood/marrow cells diagnosis	M1	46,XY
AML sample 57 diagnosis	AML blood/marrow cells diagnosis	M1	46,XX
AML sample 58 diagnosis paired with sample 97	AML blood/marrow cells diagnosis	M4	47,XX, inv(16)(p13q22),+22
AML sample 59 diagnosis paired with sample 92	AML blood/marrow cells diagnosis	M4	47,XY,inv(16)(p13q22),+22

AML sample 60 diagnosis	AML blood/marrow cells diagnosis	M5	46,XX
AML sample 61 diagnosis	AML blood/marrow cells diagnosis	M3	46,XX,t(1;10)(p32;p11.2),t(15;17)(q22;q21)
AML sample 62 diagnosis	AML blood/marrow cells diagnosis	M5	46,XX
AML sample 63 diagnosis	AML blood/marrow cells diagnosis	M1	46,XX
AML sample 64 diagnosis	AML blood/marrow cells diagnosis	M3	46,XY,t(15;17)(q22;q21)[1]/46,XY[19]
AML sample 65 diagnosis	AML blood/marrow cells diagnosis	M4	46,XX
AML sample 66 diagnosis	AML blood/marrow cells diagnosis	M4	46,XX
AML sample 68 diagnosis	AML blood/marrow cells diagnosis	M2	48,XY,+11,+13
AML sample 69 diagnosis paired with sample 94	AML blood/marrow cells diagnosis	M4	46,XY,+3,+10
AML sample 7 diagnosis	AML blood/marrow cells diagnosis	M5	46,XY,der(12)t(1;12)(q11;p11.2)[7]/46,XY[15]
AML sample 70 diagnosis	AML blood/marrow cells diagnosis	M1	45,XX,t(6,9)(q22.3;q34)/45,XX,t(6,9)(q22.3;q34),del(7)(q35)
AML sample 71 diagnosis paired with sample 104	AML blood/marrow cells diagnosis	M3	46,XY,t(15;17)(q22;q21)
AML sample 72 diagnosis paired with sample 131	AML blood/marrow cells diagnosis	M5	47,XY,+der(1;19)(q10;p10),t(9;11)(q22;q23)
AML sample 74 diagnosis	AML blood/marrow cells diagnosis	M1	46,XX,t(15;17),idem17

AML sample 75 diagnosis	AML blood/marrow cells diagnosis	M5	46,XX
AML sample 76 diagnosis	AML blood/marrow cells diagnosis	M1	46,XX
AML sample 77 diagnosis	AML blood/marrow cells diagnosis	M5	46,XY
AML sample 78 diagnosis	AML blood/marrow cells diagnosis	M4	46,XX
AML sample 79 diagnosis	AML blood/marrow cells diagnosis	M2	46,XY
AML sample 80 diagnosis	AML blood/marrow cells diagnosis	M5	46,XX
AML sample 81 remission paired with sample 311	AML marrow cells remission	NA	46,XY
AML sample 82 diagnosis	AML blood/marrow cells diagnosis	M1	45,X-Y
AML sample 88 remission paired with sample 53	AML marrow cells remission	NA	46,XY
AML sample 89 remission paired with sample 16	AML marrow cells remission	NA	46,XX
AML sample 90 remission paired with sample 13	AML marrow cells remission	NA	46,XY
AML sample 91 remission paired with GSM173411	AML marrow cells remission	NA	not analysed
AML sample 92 remission paired with sample 59	AML marrow cells remission	NA	not analysed
AML sample 93 remission paired with sample 48	AML marrow cells remission	NA	not analysed

AML sample 94 remission paired with sample 69	AML marrow cells remission	NA	46,XY
AML sample 95 remission paired with GSM173396	AML marrow cells remission	NA	46,XX
AML sample 96 remission paired with sample 32	AML marrow cells remission	NA	46,XY
AML sample 97 remission paired with sample 58	AML marrow cells remission	NA	46,XX
GSM173396	AML blood/marrow cells relapse	M4	46,XX
GSM173397	AML blood/marrow cells relapse	M1	46,XX
GSM173401	AML blood/marrow cells relapse	M4	46,XY
GSM173405	AML blood/marrow cells relapse	M4	46,XY
GSM173411	AML blood/marrow cells relapse	M1	46,XX

References

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

Affymetrix (2003) Datasheet Human Mapping 10K Array and Assay Kit.

Ahmadi, K.R., Weale, M.E., Xue, Z.Y., Soranzo, N., Yarnall, D.P., Briley, J.D., Maruyama, Y., Kobayashi, M., Wood, N.W., Spurr, N.K., Burns, D.K., Roses, A.D., Saunders, A.M. & Goldstein, D.B. (2005) A single-nucleotide polymorphism tagging set for human drug metabolism and transport. *Nat Genet,* **37,** 84-89.

Ahmed, M., Sternberg, A., Hall, G., Thomas, A., Smith, O., O'Marcaigh, A., Wynn, R., Stevens, R., Addison, M., King, D., Stewart, B., Gibson, B., Roberts, I. & Vyas, P. (2004) Natural history of GATA1 mutations in Down syndrome. *Blood,* **103,** 2480-2489.

Aitman, T.J., Dong, R., Vyse, T.J., Norsworthy, P.J., Johnson, M.D., Smith, J., Mangion, J., Roberton-Lowe, C., Marshall, A.J., Petretto, E., Hodges, M.D., Bhangal, G., Patel, S.G., Sheehan-Rooney, K., Duda, M., Cook, P.R., Evans, D.J., Domin, J., Flint, J., Boyle, J.J., Pusey, C.D. & Cook, H.T. (2006) Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature,* **439,** 851-855.

Alcalay, M., Tiacci, E., Bergomas, R., Bigerna, B., Venturini, E., Minardi, S.P., Meani, N., Diverio, D., Bernard, L., Tizzoni, L., Volorio, S., Luzi, L., Colombo, E., Lo Coco, F., Mecucci, C., Falini, B., Pelicci, P.G. & for the Gruppo Italiano Malattie Ematologiche Maligne dell'Adulto Acute Leukemia Working, P. (2005) Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc+ AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance. *Blood,* **106,** 899-902.

Andersen, C.L., Wiuf, C., Kruhoffer, M., Korsgaard, M., Laurberg, S. & Orntoft, T.F. (2007)

Frequent occurrence of uniparental disomy in colorectal cancer. *Carcinogenesis,* **28,** 38-48.

Ayton, P.M. & Cleary, M.L. (2001) Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. *Oncogene,* **20,** 5695-5707.

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

Bain, B.J. (1999) Cytology, Cytochemistry, Classification. In: *Leukaemia Diagnosis*, p. 24. Blackwell Science, Oxford.

Bain, B.J., Barnett, D., Linch, D., Matutes, E. & Reilly, J.T. (2002) Revised guideline on immunophenotyping in acute leukaemias and chronic lymphoproliferative disorders. *Clin Lab Haematol,* **24,** 1-13.

Baldus, C.D., Thiede, C., Soucek, S., Bloomfield, C.D., Thiel, E. & Ehninger, G. (2006) BAALC Expression and FLT3 Internal Tandem Duplication Mutations in Acute Myeloid Leukemia Patients With Normal Cytogenetics: Prognostic Implications. *J Clin Oncol,* **24,** 790-797.

Bardelli, A., Cahill, D.P., Lederer, G., Speicher, M.R., Kinzler, K.W., Vogelstein, B. & Lengauer, C. (2001) Carcinogen-specific induction of genetic instability. *Proceedings of the National Academy of Sciences,* **98,** 5770-5775.

Barjesteh van Waalwijk van Doorn-Khosrovani, S., Erpelinck, C., van Putten, W.L.J., Valk, P.J.M., van der Poel-van de Luytgaarde, S., Hack, R., Slater, R., Smit, E.M.E., Beverloo, H.B., Verhoef, G., Verdonck, L.F., Ossenkoppele, G.J., Sonneveld, P., de Greef, G.E., Lowenberg, B. & Delwel, R. (2003) High EVI1 expression predicts poor survival in acute myeloid leukemia: a study of 319 de novo AML patients. *Blood,* **101,** 837-845.

Basecke, J., Whelan, J.T., Griesinger, F. & Bertrand, F.E. (2006) The MLL partial tandem duplication in acute myeloid leukaemia. *British Journal of Haematology,* **135,** 438-449.

Baxter, E.J., Scott, L.M., Campbell, P.J., East, C., Fourouclas, N., Swanton, S., Vassiliou, G.S., Bench, A.J., Boyd, E.M., Curtin, N., Scott, M.A., Erber, W.N. & Green, A.R. (2005) Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet,* **365,** 1054-1061.

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

Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R. & Sultan, C. (1976) Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol,* **33,** 451-458.

Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R. & Sultan, C. (1982) Proposals for the classification of the myelodysplastic syndromes. *Br J Haematol,* **51,** 189-199.

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

Bergmann, L., Miething, C., Maurer, U., Brieger, J., Karakas, T., Weidmann, E. & Hoelzer, D. (1997) High levels of Wilms' tumor gene (wt1) mRNA in acute myeloid leukemias are associated with a worse long-term outcome. *Blood,* **90,** 1217-1225.

Bignell, G.R., Huang, J., Greshock, J., Watt, S., Butler, A., West, S., Grigorova, M., Jones, K.W., Wei, W., Stratton, M.R., Futreal, P.A., Weber, B., Shapero, M.H. & Wooster, R. (2004) High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res,* **14,** 287-295.

Blackburn, A.C., McLary, S.C., Naeem, R., Luszcz, J., Stockton, D.W., Donehower, L.A., Mohammed, M., Mailhes, J.B., Soferr, T., Naber, S.P., Otis, C.N. & Jerry, D.J. (2004) Loss of heterozygosity occurs via mitotic recombination in Trp53+/- mice and associates with mammary tumor susceptibility of the BALB/c strain. *Cancer Res,* **64,** 5140-5147.

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

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

Bradstock, K.F., Matthews, J.P., Lowenthal, R.M., Baxter, H., Catalano, J., Brighton, T., Gill, D., Eliadis, P., Joshua, D., Cannell, P., Schwarer, A.P., Durrant, S., Gillett, A., Koutts, J., Taylor, K., Bashford, J., Arthur, C., Enno, A., Dunlop, L., Szer, J., Leahy, M., Juneja, S., Young, G.A.R. & for the Australasian Leukaemia and Lymphoma, G. (2005) A randomized trial of high-versus conventional-dose cytarabine in consolidation chemotherapy for adult de novo acute myeloid leukemia in first remission after induction therapy containing high-dose cytarabine. *Blood,* **105,** 481-488.

Breivik, J. & Gaudernack, G. (2004) Resolving the evolutionary paradox of genetic instability: a cost-benefit analysis of DNA repair in changing environments. *FEBS Lett,* **563,** 7-12.

Brunning, R.D., Matutes, E., Harris, N.L., Flandrin, G., Vardiman, J.W., Bennett, J. & Head, D. (2001) Acute Myeloid Leukaemia. In: *Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues* (ed. by E.S. Jaffe, N.L. Harris, H. Stein & J.W. Vardiman), pp. 75-107. IARC Press, Lyon.

Bungaro, S., Raghavan, M., Dell'Oro, M.G., Paolucci, P., Young, B.D., Biondi, A. & Cazzaniga, G. (2006) Assessment of submicroscopic genetic lesions by single nucleotide polymorphism arrays in a child with acute myeloid leukemia and FLT3-internal tandem duplication. *Haematologica,* **91,** 998-1000.

Burchenal, J.H., Murphy, M.L., Ellison, R.R., Sykes, M.P., Tan, T.C., Leone, L.A., Karnof-Sky, D.A., Craver, L.F., Dargeon, H.W. & Rhoads, C.P. (1953) Clinical Evaluation of a New Antimetabolite, 6-Mercaptopurine, in the Treatment of Leukemia and Allied Diseases. *Blood,* **8,** 965-999.

Burnett, A.K., Goldstone, A.H., Stevens, R.M., Hann, I.M., Rees, J.K., Gray, R.G. & Wheatley, K. (1998) Randomised comparison of addition of autologous bone-marrow transplantation to intensive chemotherapy for acute myeloid leukaemia in first remission: results of MRC AML 10 trial. UK Medical Research Council Adult and Children's Leukaemia Working Parties. *Lancet,* **351,** 700-708.

Burnett, A.K., Milligan, D., Prentice, A.G., Goldstone, A.H., McMullin, M.F., Hills, R.K., Wheatley, K. & National Cancer Research Institute Haematological Oncology Study Group

Adult Leukemia Working Party (2007) A comparison of low-dose cytarabine and hydroxyurea with or without all-trans retinoic acid for acute myeloid leukemia and high-risk myelodysplastic syndrome in patients not considered fit for intensive treatment. *Cancer,* **109,** 1114-1124.

Burnett, A.K., Wheatley, K., Goldstone, A.H., Stevens, R.F., Hann, I.M., Rees, J.H. & Harrison, G. (2002) The value of allogeneic bone marrow transplant in patients with acute myeloid leukaemia at differing risk of relapse: results of the UK MRC AML 10 trial. *Br J Haematol,* **118,** 385-400.

Byrd, J.C., Mrozek, K., Dodge, R.K., Carroll, A.J., Edwards, C.G., Arthur, D.C., Pettenati, M.J., Patil, S.R., Rao, K.W., Watson, M.S., Koduru, P.R.K., Moore, J.O., Stone, R.M., Mayer, R.J., Feldman, E.J., Davey, F.R., Schiffer, C.A., Larson, R.A. & Bloomfield, C.D. (2002) Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Blood,* **100,** 4325-4336.

Calin, G.A. & Croce, C.M. (2006) MicroRNA signatures in human cancers. *Nat Rev Cancer,* **6,** 857-866.

Cervantes, F., Barosi, G., Demory, J.-L., Reilly, J., Guarnone, R., Dupriez, B., Pereira, A. & Montserrat, E. (1998) Myelofibrosis with myeloid metaplasia in young indidviduals: disease characteristics, prognostic factors and identification of risk groups. *British Journal of Haematology,* **102,** 684-690.

Chaganti, R.S.K., Schonberg, S. & German, J. (1974) A Manyfold Increase in Sister Chromatid Exchanges in Bloom's Syndrome Lymphocytes. *Proceedings of the National Academy of Sciences,* **71,** 4508-4512.

Chapiro, E., Delabesse, E., Asnafi, V., Millien, C., Davi, F., Nugent, E., Beldjord, K., Haferlach, T., Grimwade, D. & Macintyre, E.A. (2006) Expression of T-lineage affiliated transcripts and TCR rearrangements in acute promyelocytic leukemia: implications for the cellular target of the t(15;17). *Blood.*

Chen, J.D. & Evans, R.M. (1995) A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature,* **377,** 454-457.

Christiansen, D.H., Andersen, M.K. & Pedersen-Bjergaard, J. (2003) Methylation of p15INK4B is common, is associated with deletion of genes on chromosome arm 7q and predicts a poor prognosis in therapy-related myelodysplasia and acute myeloid leukemia. *Leukemia,* **17,** 1813-1819.

Colita, A., Belhabri, A., Chelghoum, Y., Charrin, C., Fiere, D. & Thomas, X. (2001) Prognostic factors and treatment effects on survival in acute myeloid leukemia of M6 subtype: A retrospective study of 54 cases. *Ann Oncol,* **12,** 451-455.

Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E. & Pritchard, J.K. (2006) A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet,* **38,** 75-81.

Cornelissen, J.J., van Putten, W.L.J., Verdonck, L.F., Theobald, M., Jacky, E., Daenen, S.M.G., van Marwijk Kooy, M., Wijermans, P., Schouten, H., Huijgens, P.C., van der Lelie, H., Fey, M., Ferrant, A., Maertens, J., Gratwohl, A. & Lowenberg, B. (2007) Results of a HOVON/SAKK donor versus no-donor analysis of myeloablative HLA-identical sibling stem cell transplantation in first remission acute myeloid leukemia in young and middle-aged adults: benefits for whom? *Blood,* **109,** 3658-3666.

Court-Brown, W.M. & Doll, R. (1957) Leukaemia and aplastic anaemia in patients irradiated for ankylosing spondylitis. *Spec Rep Ser Med Res Counc (G B),* 1-135.

Craigie, D. (1845) Case of disease and enlargement of the spleen in which death took place from the presence of purulent matter in the blood. *Edinburgh Medical and Surgical Journal,* **64,** 400-413.

Crowther, D., Bateman, C.J., Vartan, C.P., Whitehouse, J.M., Malpas, J.S., Fairley, G.H. & Scott, R.B. (1970) Combination chemotherapy using L-asparaginase, daunorubicin, and cytosine arabinoside in adults with acute myelogenous leukaemia. *Br Med J,* **4,** 513-517.

Cuenco, G.M., Nucifora, G. & Ren, R. (2000) Human AML1/MDS1/EVI1 fusion protein induces an acute myelogenous leukemia (AML) in mice: a model for human AML. *Proc Natl Acad Sci U S A,* **97,** 1760-1765.

Cui, H., Onyango, P., Brandenburg, S., Wu, Y., Hsieh, C.L. & Feinberg, A.P. (2002) Loss of imprinting in colorectal cancer linked to hypomethylation of H19 and IGF2. *Cancer Res,* **62,**

6442-6446.

Czepulkowski, B.H., Bhatt, B. & Rooney, D.E. (1992) Basic techniques for the analysis of chromosomes from bone marrow and leukaemic blood. In: *Human cytogenetics: A practical approach, Malignancy and acquired abnormalities.* (ed. by D. Rickwood & B.D. Hames), Vol. 2, pp. 1-15. Oxford University Press, Oxford.

Dale, D.C., Cottle, T.E., Fier, C.J., Bolyard, A.A., Bonilla, M.A., Boxer, L.A., Cham, B., Freedman, M.H., Kannourakis, G., Kinsey, S.E., Davis, R., Scarlata, D., Schwinzer, B., Zeidler, C. & Welte, K. (2003) Severe chronic neutropenia: treatment and follow-up of patients in the Severe Chronic Neutropenia International Registry. *Am J Hematol,* **72,** 82-93.

Davies, J.K., Taussig, D.C., Oakervee, H., Davies, A.J., Agrawal, S.G., Gribben, J.G., Lister, T.A. & Cavenagh, J.D. (2006) Long-Term Follow-Up After Reduced-Intensity Conditioning Allogeneic Transplantation for Acute Myeloid Leukemia/Myelodysplastic Syndrome: Late CNS Relapses Despite Graft-Versus-Host Disease. *J Clin Oncol,* **24,** e23-25.

de Nooij-van Dalen, A.G., van Buuren-van Seggelen, V.H., Mulder, A., Gelsthorpe, K., Cole, J., Lohman, P.H. & Giphart-Gassler, M. (1997) Isolation and molecular characterization of spontaneous mutants of lymphoblastoid cells with extended loss of heterozygosity. *Mutat Res,* **374,** 51-62.

de The, H., Chomienne, C., Lanotte, M., Degos, L. & Dejean, A. (1990) The t(15;17) translocation of acute promyelocytic leukaemia fuses the retinoic acid receptor [alpha] gene to a novel transcribed locus. *Nature,* **347,** 558-561.

Debernardi, S., Lillington, D.M., Chaplin, T., Tomlinson, S., Amess, J., Rohatiner, A., Lister, T.A. & Young, B.D. (2003) Genome-wide analysis of acute myeloid leukemia with normal karyotype reveals a unique pattern of homeobox gene expression distinct from those with translocation-mediated fusion events. *Genes Chromosomes Cancer,* **37,** 149-158.

Debernardi, S., Skoulakis, S., Molloy, G., Chaplin, T., Dixon-McIver, A. & Young, B.D. (2007) MicroRNA miR-181a correlates with morphological sub-class of acute myeloid leukaemia and the expression of its target genes in global genome-wide analysis. *Leukemia,* **21,** 912-916.

DiMartino, J.F. & Cleary, M.L. (1999) MLL rearrangements in haematological malignancies:

lessons from clinical and biological studies. *British Journal of Haematology,* **106,** 614-626.

Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B.D. & Evans, G.A. (1992) A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. *Nat Genet,* **2,** 113-118.

Dorrance, A.M., Liu, S., Yuan, W., Becknell, B., Arnoczky, K.J., Guimond, M., Strout, M.P., Feng, L., Nakamura, T., Yu, L., Rush, L.J., Weinstein, M., Leone, G., Wu, L., Ferketich, A., Whitman, S.P., Marcucci, G. & Caligiuri, M.A. (2006) Mll partial tandem duplication induces aberrant Hox expression in vivo via specific epigenetic alterations. *J Clin Invest,* **116,** 2707-2716.

Druker, B.J., Guilhot, F., O'Brien, S.G., Gathmann, I., Kantarjian, H., Gattermann, N., Deininger, M.W., Silver, R.T., Goldman, J.M., Stone, R.M., Cervantes, F., Hochhaus, A., Powell, B.L., Gabrilove, J.L., Rousselot, P., Reiffers, J., Cornelissen, J.J., Hughes, T., Agis, H., Fischer, T., Verhoef, G., Shepherd, J., Saglio, G., Gratwohl, A., Nielsen, J.L., Radich, J.P., Simonsson, B., Taylor, K., Baccarani, M., So, C., Letvak, L. & Larson, R.A. (2006) Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N Engl J Med,* **355,** 2408-2417.

Dupret, C., Asnafi, V., Leboeuf, D., Millien, C., Ben Abdelali, R., Preudhomme, C., Beldjord, K., Delabesse, E. & Macintyre, E. (2005) IgH/TCR rearrangements are common in MLL translocated adult AML and suggest an early T/myeloid or B/myeloid maturation arrest, which correlates with the MLL partner. *Leukemia,* **19,** 2337-2338.

Easton, D.F. & Pooley, K.A. & Dunning, A.M. & Pharoah, P.D. & Thompson, D. & Ballinger, D.G. & Struewing, J.P. & Morrison, J. & Field, H. & Luben, R. & Wareham, N. & Ahmed, S. & Healey, C.S. & Bowman, R. & Meyer, K.B. & Haiman, C.A. & Kolonel, L.K. & Henderson, B.E. & Le Marchand, L. & Brennan, P. & Sangrajrang, S. & Gaborieau, V. & Odefrey, F. & Shen, C.Y. & Wu, P.E. & Wang, H.C. & Eccles, D. & Evans, D.G. & Peto, J. & Fletcher, O. & Johnson, N. & Seal, S. & Stratton, M.R. & Rahman, N. & Chenevix-Trench, G. & Bojesen, S.E. & Nordestgaard, B.G. & Axelsson, C.K. & Garcia-Closas, M. & Brinton, L. & Chanock, S. & Lissowska, J. & Peplonska, B. & Nevanlinna, H. & Fagerholm, R. & Eerola, H. & Kang, D. &

Yoo, K.Y. & Noh, D.Y. & Ahn, S.H. & Hunter, D.J. & Hankinson, S.E. & Cox, D.G. & Hall, P. & Wedren, S. & Liu, J. & Low, Y.L. & Bogdanova, N. & Schurmann, P. & Dork, T. & Tollenaar, R.A. & Jacobi, C.E. & Devilee, P. & Klijn, J.G. & Sigurdson, A.J. & Doody, M.M. & Alexander, B.H. & Zhang, J. & Cox, A. & Brock, I.W. & MacPherson, G. & Reed, M.W. & Couch, F.J. & Goode, E.L. & Olson, J.E. & Meijers-Heijboer, H. & van den Ouweland, A. & Uitterlinden, A. & Rivadeneira, F. & Milne, R.L. & Ribas, G. & Gonzalez-Neira, A. & Benitez, J. & Hopper, J.L. & McCredie, M. & Southey, M. & Giles, G.G. & Schroen, C. & Justenhoven, C. & Brauch, H. & Hamann, U. & Ko, Y.D. & Spurdle, A.B. & Beesley, J. & Chen, X. & Mannermaa, A. & Kosma, V.M. & Kataja, V. & Hartikainen, J. & Day, N.E. & Cox, D.R. & Ponder, B.A. (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature,* **447,** 1087-1093.

Estey, E., Keating, M.J., Pierce, S. & Stass, S. (1995) Change in karyotype between diagnosis and first relapse in acute myelogenous leukemia. *Leukemia,* **9,** 972-976.

Falini, B., Mecucci, C., Tiacci, E., Alcalay, M., Rosati, R., Pasqualucci, L., La Starza, R., Diverio, D., Colombo, E., Santucci, A., Bigerna, B., Pacini, R., Pucciarini, A., Liso, A., Vignetti, M., Fazi, P., Meani, N., Pettirossi, V., Saglio, G., Mandelli, F., Lo-Coco, F., Pelicci, P.G. & Martelli, M.F. (2005) Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N Engl J Med,* **352,** 254-266.

Felsher, D.W. & Bishop, J.M. (1999) Reversible Tumorigenesis by MYC in Hematopoietic Lineages. *Molecular Cell,* **4,** 199-207.

Fialkow, P.J., Gartler, S.M. & Yoshida, A. (1967) Clonal origin of chronic myelocytic leukemia in man. *Proc Natl Acad Sci U S A,* **58,** 1468-1471.

Fitzgibbon, J., Iqbal, S., Davies, A., O'Shea, D., Carlotti, E., Chaplin, T., Matthews, J., Raghavan, M., Norton, A., Lister, T.A. & Young, B.D. (2007) Genome-wide detection of recurring sites of uniparental disomy in follicular and transformed follicular lymphoma. *Leukemia,* **21,** 1514-1520.

Fitzgibbon, J., Smith, L.L., Raghavan, M., Smith, M.L., Debernardi, S., Skoulakis, S., Lillington, D., Lister, T.A. & Young, B.D. (2005) Association between acquired uniparental

disomy and homozygous gene mutation in acute myeloid leukemias. *Cancer Res,* **65,** 9152-9154.

Flotho, C., Steinemann, D., Mullighan, C.G., Neale, G., Mayer, K., Kratz, C.P., Schlegelberger, B., Downing, J.R. & Niemeyer, C.M. (2007) Genome-wide single-nucleotide polymorphism analysis in juvenile myelomonocytic leukemia identifies uniparental disomy surrounding the NF1 locus in cases associated with neurofibromatosis but not in cases with mutant RAS or PTPN11. *Oncogene.*

Folley, J.H., Borges, W. & Yamawaki, T. (1952) Incidence of leukemia in survivors of the atomic bomb in Hiroshima and Nagasaki, Japan. *Am J Med,* **13,** 311-321.

Freedman, M.H. (2000) Diamond-Blackfan anaemia. *Best Practice & Research Clinical Haematology,* **13,** 391-406.

Gale, R.E., Hills, R., Kottaridis, P.D., Srirangan, S., Wheatley, K., Burnett, A.K. & Linch, D.C. (2005) No evidence that FLT3 status should be considered as an indicator for transplantation in acute myeloid leukemia (AML): an analysis of 1135 patients, excluding acute promyelocytic leukemia, from the UK MRC AML10 and 12 trials. *Blood,* **106,** 3658-3665.

Gale, R.E., Wheadon, H., Goldstone, A.H., Burnett, A.K. & Linch, D.C. (1993) Frequency of clonal remission in acute myeloid leukaemia. *Lancet,* **341,** 138-142.

Gao, F., Maiti, S., Sun, G., Ordonez, N.G., Udtha, M., Deng, J.M., Behringer, R.R. & Huff, V. (2004) The Wt1+/R394W Mouse Displays Glomerulosclerosis and Early-Onset Renal Failure Characteristic of Human Denys-Drash Syndrome. *Mol. Cell. Biol.,* **24,** 9899-9910.

Gao, Q., Horwitz, M., Roulston, D., Hagos, F., Zhao, N., Freireich, E.J., Golomb, H.M. & Olopade, O.I. (2000) Susceptibility gene for familial acute myeloid leukemia associated with loss of 5q and/or 7q is not localized on the commonly deleted portion of 5q. *Genes Chromosomes Cancer,* **28,** 164-172.

Garson, O.M., Hagemeijer, A., Sakurai, M., Reeves, B.R., Swansbury, G.J., Williams, G.J., Alimena, G., Arthur, D.C., Berger, R., de la Chapelle, A. & et al. (1989) Cytogenetic studies of 103 patients with acute myelogenous leukemia in relapse. *Cancer Genet Cytogenet,* **40,** 187-202.

Gelmetti, V., Zhang, J., Fanelli, M., Minucci, S., Pelicci, P.G. & Lazar, M.A. (1998) Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO. *Mol Cell Biol,* **18,** 7185-7191.

German, J. (1997) Bloom's syndrome. XX. The first 100 cancers. *Cancer Genetics and Cytogenetics,* **93,** 100-106.

Gilliland, D.G., Jordan, C.T. & Felix, C.A. (2004) The molecular basis of leukemia. *Hematology Am Soc Hematol Educ Program,* 80-97.

Glover, T.W., Arlt, M.F., Casper, A.M. & Durkin, S.G. (2005) Mechanisms of common fragile site instability. *Hum Mol Genet,* **14 Spec No. 2,** R197-205.

Gluzman, D., Imamura, N., Sklyarenko, L., Nadgornaya, V., Zavelevich, M. & Machilo, V. (2006) Patterns of hematological malignancies in Chernobyl clean-up workers (1996-2005). *Exp Oncol,* **28,** 60-63.

Goddard, A.D., Borrow, J., Freemont, P.S. & Solomon, E. (1991) Characterization of a zinc finger gene disrupted by the t(15;17) in acute promyelocytic leukemia. *Science,* **254,** 1371-1374.

Goldstone, A.H., Burnett, A.K., Wheatley, K., Smith, A.G., Hutchinson, R.M. & Clark, R.E. (2001) Attempts to improve treatment outcomes in acute myeloid leukemia (AML) in older patients: the results of the United Kingdom Medical Research Council AML11 trial. *Blood,* **98,** 1302-1311.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D. & Lander, E.S. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science,* **286,** 531-537.

Gorletta, T.A., Gasparini, P., D'Elios, M.M., Trubia, M., Pelicci, P.G. & Di Fiore, P.P. (2005) Frequent loss of heterozygosity without loss of genetic material in acute myeloid leukemia with a normal karyotype. *Genes Chromosomes Cancer,* **44,** 334-337.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., Edkins, S., O'Meara, S., Vastrik, I., Schmidt, E.E., Avis, T.,

Barthorpe, S., Bhamra, G., Buck, G., Choudhury, B., Clements, J., Cole, J., Dicks, E., Forbes, S., Gray, K., Halliday, K., Harrison, R., Hills, K., Hinton, J., Jenkinson, A., Jones, D., Menzies, A., Mironenko, T., Perry, J., Raine, K., Richardson, D., Shepherd, R., Small, A., Tofts, C., Varian, J., Webb, T., West, S., Widaa, S., Yates, A., Cahill, D.P., Louis, D.N., Goldstraw, P., Nicholson, A.G., Brasseur, F., Looijenga, L., Weber, B.L., Chiew, Y.-E., deFazio, A., Greaves, M.F., Green, A.R., Campbell, P., Birney, E., Easton, D.F., Chenevix-Trench, G., Tan, M.-H., Khoo, S.K., Teh, B.T., Yuen, S.T., Leung, S.Y., Wooster, R., Futreal, P.A. & Stratton, M.R. (2007) Patterns of somatic mutation in human cancer genomes. *Nature,* **446,** 153-158.

Griekspoor, A. & Groothuis, T. (2005) *4Peaks.*

Griffith, J., Black, J., Faerman, C., Swenson, L., Wynn, M., Lu, F., Lippke, J. & Saxena, K. (2004) The structural basis for autoinhibition of FLT3 by the juxtamembrane domain. *Mol Cell,* **13,** 169-178.

Grignani, F., De Matteis, S., Nervi, C., Tomassoni, L., Gelmetti, V., Cioce, M., Fanelli, M., Ruthardt, M., Ferrara, F.F., Zamir, I., Seiser, C., Grignani, F., Lazar, M.A., Minucci, S. & Pelicci, P.G. (1998) Fusion proteins of the retinoic acid receptor-[alpha] recruit histone deacetylase in promyelocytic leukaemia. *Nature,* **391,** 815-818.

Grignani, F., Ferrucci, P.F., Testa, U., Talamo, G., Fagioli, M., Alcalay, M., Mencarelli, A., Grignani, F., Peschle, C., Nicoletti, I. & Pelicci, P.G. (1993) The acute promyelocytic leukemia-specific PML-RAR[alpha] fusion protein inhibits differentiation and promotes survival of myeloid precursor cells. *Cell,* **74,** 423-431.

Grimwade, D. (2002) The significance of minimal residual disease in patients with t(15;17). *Best Practice & Research Clinical Haematology,* **15,** 137-158.

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

Grimwade, D., Walker, H., Oliver, F., Wheatley, K., Harrison, C., Harrison, G., Rees, J., Hann,

I., Stevens, R., Burnett, A. & Goldstone, A. (1998) The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood,* **92,** 2322-2333.

Grimwade, D.J., Stephenson, J., De Silva, C., Dalton, R.G. & Mufti, G.J. (1993) Familial MDS with 5q- abnormality. *Br J Haematol,* **84,** 536-538.

Grimwood, J., Gordon, L.A., Olsen, A., Terry, A., Schmutz, J., Lamerdin, J., Hellsten, U., Goodstein, D., Couronne, O., Tran-Gyamfi, M., Aerts, A., Altherr, M., Ashworth, L., Bajorek, E., Black, S., Branscomb, E., Caenepeel, S., Carrano, A., Caoile, C., Chan, Y.M., Christensen, M., Cleland, C.A., Copeland, A., Dalin, E., Dehal, P., Denys, M., Detter, J.C., Escobar, J., Flowers, D., Fotopulos, D., Garcia, C., Georgescu, A.M., Glavina, T., Gomez, M., Gonzales, E., Groza, M., Hammon, N., Hawkins, T., Haydu, L., Ho, I., Huang, W., Israni, S., Jett, J., Kadner, K., Kimball, H., Kobayashi, A., Larionov, V., Leem, S.H., Lopez, F., Lou, Y., Lowry, S., Malfatti, S., Martinez, D., McCready, P., Medina, C., Morgan, J., Nelson, K., Nolan, M., Ovcharenko, I., Pitluck, S., Pollard, M., Popkie, A.P., Predki, P., Quan, G., Ramirez, L., Rash, S., Retterer, J., Rodriguez, A., Rogers, S., Salamov, A., Salazar, A., She, X., Smith, D., Slezak, T., Solovyev, V., Thayer, N., Tice, H., Tsai, M., Ustaszewska, A., Vo, N., Wagner, M., Wheeler, J., Wu, K., Xie, G., Yang, J., Dubchak, I., Furey, T.S., DeJong, P., Dickson, M., Gordon, D., Eichler, E.E., Pennacchio, L.A., Richardson, P., Stubbs, L., Rokhsar, D.S., Myers, R.M., Rubin, E.M. & Lucas, S.M. (2004) The DNA sequence and biology of human chromosome 19. *Nature,* **428,** 529-535.

Grisendi, S., Mecucci, C., Falini, B. & Pandolfi, P.P. (2006) Nucleophosmin and cancer. *Nat Rev Cancer,* **6,** 493-505.

Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L.T., Gudbjartsson, D., Helgason, A., Rafnar, T., Bergthorsson, J.T., Agnarsson, B.A., Baker, A., Sigurdsson, A., Benediktsdottir, K.R., Jakobsdottir, M., Xu, J., Blondal, T., Kostic, J., Sun, J., Ghosh, S., Stacey, S.N., Mouy, M., Saemundsdottir, J., Backman, V.M., Kristjansson, K., Tres, A., Partin, A.W., Albers-Akkers, M.T., Godino-Ivan Marcos, J., Walsh, P.C., Swinkels, D.W., Navarrete, S., Isaacs,

S.D., Aben, K.K., Graif, T., Cashy, J., Ruiz-Echarri, M., Wiley, K.E., Suarez, B.K., Witjes, J.A., Frigge, M., Ober, C., Jonsson, E., Einarsson, G.V., Mayordomo, J.I., Kiemeney, L.A., Isaacs, W.B., Catalona, W.J., Barkardottir, R.B., Gulcher, J.R., Thorsteinsdottir, U., Kong, A. & Stefansson, K. (2007) Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet,* **39,** 631-637.

Gupta, M., Raghavan, M., Gale, R.E., Chelala, C., Allen, C., Molloy, G., Chaplin, T., Linch, D.C., Cazier, J.-B. & Young, B.D. (2008) Novel regions of acquired uniparental disomy discovered in acute myeloid leukemia. *Genes, Chromosomes and Cancer,* **47,** 729-818.

Hagstrom, S.A. & Dryja, T.P. (1999) Mitotic recombination map of 13cen-13q14 derived from an investigation of loss of heterozygosity in retinoblastomas. *Proc Natl Acad Sci U S A,* **96,** 2952-2957.

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

Harrison, C.N. (2005) Essential thrombocythaemia: challenges and evidence-based management. *Br J Haematol,* **130,** 153-165.

Hayes, R.B., Yin, S.N., Dosemeci, M., Li, G.L., Wacholder, S., Travis, L.B., Li, C.Y., Rothman, N., Hoover, R.N. & Linet, M.S. (1997) Benzene and the dose-related incidence of hematologic neoplasms in China. Chinese Academy of Preventive Medicine--National Cancer Institute Benzene Study Group. *J Natl Cancer Inst,* **89,** 1065-1071.

Hehlmann, R., Heimpel, H., Hasford, J., Kolb, H.J., Pralle, H., Hossfeld, D.K., Queisser, W., Loffler, H., Heinze, B. & Georgii, A. (1993) Randomized comparison of busulfan and hydroxyurea in chronic myelogenous leukemia: prolongation of survival by hydroxyurea. The German CML Study Group. *Blood,* **82,** 398-407.

Heinzel, T., Lavinsky, R.M., Mullen, T.M., Soderstrom, M., Laherty, C.D., Torchia, J., Yang, W.M., Brard, G., Ngo, S.D., Davie, J.R., Seto, E., Eisenman, R.N., Rose, D.W., Glass, C.K. &

Rosenfeld, M.G. (1997) A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. *Nature,* **387,** 43-48.

Henry, I., Bonaiti-Pellie, C., Chehensse, V., Beldjord, C., Schwartz, C., Utermann, G. & Junien, C. (1991) Uniparental paternal disomy in a genetic cancer-predisposing syndrome. *Nature,* **351,** 665-667.

Hess, J.L. (2001) Familial Monosomy 7 Syndrome. Atlas Genet Cytogenet Oncol Haematol, http://AtlasGeneticsOncology.org/Kprones/FamilMono7ID10059.html

Hess, J.L., Yu, B.D., Li, B., Hanson, R. & Korsmeyer, S.J. (1997) Defects in Yolk Sac Hematopoiesis in Mll-Null Embryos. *Blood,* **90,** 1799-1806.

Heuser, M., Beutel, G., Krauter, J., Dohner, K., von Neuhoff, N., Schlegelberger, B. & Ganser, A. (2006) High meningioma 1 (MN1) expression as a predictor for poor outcome in acute myeloid leukemia with normal cytogenetics. *Blood,* **108,** 3898-3905.

Holt, D., Dreimanis, M., Pfeiffer, M., Firgaira, F., Morley, A. & Turner, D. (1999) Interindividual variation in mitotic recombination. *Am J Hum Genet,* **65,** 1423-1427.

Hoque, M.O., Lee, J., Begum, S., Yamashita, K., Engles, J.M., Schoenberg, M., Westra, W.H. & Sidransky, D. (2003) High-throughput molecular analysis of urine sediment for the detection of bladder cancer by high-density single-nucleotide polymorphism array. *Cancer Res,* **63,** 5723-5726.

Huang, J., Wei, W., Zhang, J., Liu, G., Bignell, G.R., Stratton, M.R., Futreal, P.A., Wooster, R., Jones, K.W. & Shapero, M.H. (2004) Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum Genomics,* **1,** 287-299.

Huentelman, M.J., Craig, D.W., Shieh, A.D., Corneveaux, J.J., Hu-Lince, D., Pearson, J.V. & Stephan, D.A. (2005) SNiPer: improved SNP genotype calling for Affymetrix 10K GeneChip microarray data. *BMC Genomics,* **6,** 149.

Human Genome Structural Variation Working Group (2007) Completing the map of human genetic variation. *Nature,* **447,** 161-165.

Inoue, K., Sugiyama, H., Ogawa, H., Nakagawa, M., Yamagami, T., Miwa, H., Kita, K., Hiraoka, A., Masaoka, T., Nasu, K. & et al. (1994) WT1 as a new prognostic factor and a new

marker for the detection of minimal residual disease in acute leukemia. *Blood,* **84,** 3071-3079.

International HapMap Consortium (2005) A haplotype map of the human genome. *Nature,* **437,** 1299-1320.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature,* **409,** 860-921.

James, C., Ugo, V., Le Couedic, J.P., Staerk, J., Delhommeau, F., Lacout, C., Garcon, L., Raslova, H., Berger, R., Bennaceur-Griscelli, A., Villeval, J.L., Constantinescu, S.N., Casadevall, N. & Vainchenker, W. (2005) A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature,* **434,** 1144-1148.

Janne, P.A., Li, C., Zhao, X., Girard, L., Chen, T.H., Minna, J., Christiani, D.C., Johnson, B.E. & Meyerson, M. (2004) High-resolution single-nucleotide polymorphism array and clustering analysis of loss of heterozygosity in human lung cancer cell lines. *Oncogene,* **23,** 2716-2726.

Jawad, M., Seedhouse, C.H., Russell, N. & Plumb, M. (2006) Polymorphisms in human homeobox HLX1 and DNA repair RAD51 genes increase the risk of therapy-related acute myeloid leukemia. *Blood.*

Jones, B.M., White, A.D., Culligan, D.J. & Jacobs, A. (1992) Cell-cycle progression rates and sister chromatid exchange frequencies in the bone marrow of patients with myelodysplastic syndrome and acute myeloid leukemia. *Cancer Genetics and Cytogenetics,* **62,** 66-69.

Jourdan, E., Boiron, J.-M., Dastugue, N., Vey, N., Marit, G., Rigal-Huguet, F., Molina, L., Fegueux, N., Pigneux, A., Recher, C., Rossi, J.-F., Attal, M., Sotto, J.-J., Maraninchi, D., Reiffers, J., Bardou, V.-J., Esterni, B. & Blaise, D. (2005) Early Allogeneic Stem-Cell Transplantation for Young Adults With Acute Myeloblastic Leukemia in First Complete Remission: An Intent-to-Treat Long-Term Analysis of the BGMT Experience. *J Clin Oncol,* **23,** 7676-7684.

Kallioniemi, O.P., Kallioniemi, A., Piper, J., Isola, J., Waldman, F.M., Gray, J.W. & Pinkel, D. (1994) Optimizing comparative genomic hybridization for analysis of DNA sequence copy number changes in solid tumors. *Genes Chromosomes Cancer,* **10,** 231-243.

Kantarjian, H., Oki, Y., Garcia-Manero, G., Huang, X., O'Brien, S., Cortes, J., Faderl, S.,

Bueso-Ramos, C., Ravandi, F., Estrov, Z., Ferrajoli, A., Wierda, W., Shan, J., Davis, J., Giles, F., Saba, H.I. & Issa, J.-P.J. (2007) Results of a randomized study of 3 schedules of low-dose decitabine in higher-risk myelodysplastic syndrome and chronic myelomonocytic leukemia. *Blood,* **109,** 52-57.

Kantarjian, H., Sawyers, C., Hochhaus, A., Guilhot, F., Schiffer, C., Gambacorti-Passerini, C., Niederwieser, D., Resta, D., Capdeville, R., Zoellner, U., Talpaz, M., Druker, B., Goldman, J., O'Brien, S.G., Russell, N., Fischer, T., Ottmann, O., Cony-Makhoul, P., Facon, T., Stone, R., Miller, C., Tallman, M., Brown, R., Schuster, M., Loughran, T., Gratwohl, A., Mandelli, F., Saglio, G., Lazzarino, M., Russo, D., Baccarani, M. & Morra, E. (2002) Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. *N Engl J Med,* **346,** 645-652.

Karanjawala, Z.E., Kaariainen, H., Ghosh, S., Tannenbaum, J., Martin, C., Ally, D., Tuomilehto, J., Valle, T. & Collins, F.S. (2000) Complete maternal isodisomy of chromosome 8 in an individual with an early-onset ileal carcinoid tumor. *Am J Med Genet,* **93,** 207-210.

Kawamata, N., Ogawa, S., Zimmermann, M., Kato, M., Sanada, M., Hemminki, K., Yamatomo, G., Nannya, Y., Koehler, R., Flohr, T., Miller, C.W., Harbott, J., Ludwig, W.-D., Stanulla, M., Schrappe, M., Bartram, C.R. & Koeffler, H.P. (2008) Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high-resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood,* **111,** 776-784.

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

Kelsell, D.P., Norgett, E.E., Unsworth, H., Teh, M.T., Cullup, T., Mein, C.A., Dopping-Hepenstal, P.J., Dale, B.A., Tadini, G., Fleckman, P., Stephens, K.G., Sybert, V.P., Mallory, S.B., North, B.V., Witt, D.R., Sprecher, E., Taylor, A.E., Ilchyshyn, A., Kennedy, C.T., Goodyear, H., Moss, C., Paige, D., Harper, J.I., Young, B.D., Leigh, I.M., Eady, R.A. & O'Toole, E.A. (2005) Mutations in ABCA12 underlie the severe congenital skin disease harlequin ichthyosis. *Am J Hum Genet,* **76,** 794-803.

Kern, W., Haferlach, T., Schnittger, S., Ludwig, W.D., Hiddemann, W. & Schoch, C. (2002)

Karyotype instability between diagnosis and relapse in 117 patients with acute myeloid leukemia: implications for resistance against therapy. *Leukemia,* **16,** 2084-2091.

Kern, W., Kohlmann, A., Wuchter, C., Schnittger, S., Schoch, C., Mergenthaler, S., Ratei, R., Ludwig, W.D., Hiddemann, W. & Haferlach, T. (2003) Correlation of protein expression and gene expression in acute leukemia. *Cytometry B Clin Cytom,* **55,** 29-36.

Kim, H.-G., Kojima, K., Swindle, C.S., Cotta, C.V., Huo, Y., Reddy, V. & Klug, C.A. (2007) FLT3-ITD cooperates with inv(16) to promote progression to acute myeloid leukemia. *Blood,* blood-2006-2006-030312.

Kimchi-Sarfaty, C., Oh, J.M., Kim, I.-W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V. & Gottesman, M.M. (2007) A "Silent" Polymorphism in the MDR1 Gene Changes Substrate Specificity. *Science,* **315,** 525-528.

King-Underwood, L., Renshaw, J. & Pritchard-Jones, K. (1996) Mutations in the Wilms' tumor gene WT1 in leukemias. *Blood,* **87,** 2171-2179.

Kinzler, K.W. & Vogelstein, B. (2002) Introduction. In: *The Genetic Basis of Human Cancer* (ed. by K.W. Kinzler & B. Vogelstein), p. 6. McGraw-Hill, New York.

Kitamura, Y. & Hirotab, S. (2004) Oncogenic protein tyrosine kinases. *Cellular and Molecular Life Sciences (CMLS),* **61,** 2924-2931.

Kiyoi, H., Naoe, T., Nakano, Y., Yokota, S., Minami, S., Miyawaki, S., Asou, N., Kuriyama, K., Jinnai, I., Shimazaki, C., Akiyama, H., Saito, K., Oh, H., Motoji, T., Omoto, E., Saito, H., Ohno, R. & Ueda, R. (1999) Prognostic implication of FLT3 and N-RAS gene mutations in acute myeloid leukemia. *Blood,* **93,** 3074-3080.

Kiyoi, H., Towatari, M., Yokota, S., Hamaguchi, M., Ohno, R., Saito, H. & Naoe, T. (1998) Internal tandem duplication of the FLT3 gene is a novel modality of elongation mutation which causes constitutive activation of the product. *Leukemia,* **12,** 1333-1337.

Knapper, S., Burnett, A.K., Littlewood, T., Kell, W.J., Agrawal, S., Chopra, R., Clark, R., Levis, M.J. & Small, D. (2006) A phase 2 trial of the FLT3 inhibitor lestaurtinib (CEP701) as first-line treatment for older patients with acute myeloid leukemia not considered fit for intensive chemotherapy. *Blood,* **108,** 3262-3270.

Knudson, A.G., Jr. (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A,* **68,** 820-823.

Koopman, R.J. (1999) Concept of twin spotting. *Am J Med Genet,* **85,** 355-358.

Kottaridis, P.D., Gale, R.E., Frew, M.E., Harrison, G., Langabeer, S.E., Belton, A.A., Walker, H., Wheatley, K., Bowen, D.T., Burnett, A.K., Goldstone, A.H. & Linch, D.C. (2001) The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. *Blood,* **98,** 1752-1759.

Kottaridis, P.D., Gale, R.E., Langabeer, S.E., Frew, M.E., Bowen, D.T. & Linch, D.C. (2002) Studies of FLT3 mutations in paired presentation and relapse samples from patients with acute myeloid leukemia: implications for the role of FLT3 mutations in leukemogenesis, minimal residual disease detection, and possible therapy with FLT3 inhibitors. *Blood,* **100,** 2393-2398.

Kralovics, R., Guan, Y. & Prchal, J.T. (2002) Acquired uniparental disomy of chromosome 9p is a frequent stem cell defect in polycythemia vera. *Experimental Hematology,* **30,** 229-236.

Kralovics, R., Passamonti, F., Buser, A.S., Teo, S.S., Tiedt, R., Passweg, J.R., Tichelli, A., Cazzola, M. & Skoda, R.C. (2005) A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med,* **352,** 1779-1790.

Kuendgen, A., Knipp, S., Fox, F., Strupp, C., Hildebrandt, B., Steidl, C., Germing, U., Haas, R. & Gattermann, N. (2005) Results of a phase 2 study of valproic acid alone or in combination with all- trans retinoic acid in 75 patients with myelodysplastic syndrome and relapsed or refractory acute myeloid leukemia. *Annals of Hematology,* **84,** 61-66.

Kundu, M., Chen, A., Anderson, S., Kirby, M., Xu, L., Castilla, L.H., Bodine, D. & Liu, P.P. (2002) Role of Cbfb in hematopoiesis and perturbations resulting from expression of the leukemogenic fusion gene Cbfb-MYH11. *Blood,* **100,** 2449-2456.

Kuptsova, N., Kopecky, K.J., Godwin, J., Anderson, J., Hoque, A., Willman, C.L., Slovak, M.L. & Ambrosone, C.B. (2007) Polymorphisms in DNA repair genes and therapeutic outcomes of AML patients from SWOG clinical trials. *Blood,* **109,** 3936-3944.

Laslo, P., Spooner, C.J., Warmflash, A., Lancki, D.W., Lee, H.-J., Sciammas, R., Gantner, B.N., Dinner, A.R. & Singh, H. (2006) Multilineage Transcriptional Priming and Determination of Alternate Hematopoietic Cell Fates. *Cell,* **126,** 755-766.

Lauchle, J.O., Braun, B.S., Loh, M.L. & Shannon, K. (2006) Inherited predispositions and hyperactive Ras in myeloid leukemogenesis. *Pediatric Blood & Cancer,* **46,** 579-585.

Lee, B.H., Williams, I.R., Anastasiadou, E., Boulton, C.L., Joseph, S.W., Amaral, S.M., Curley, D.P., Duclos, N., Huntly, B.J., Fabbro, D., Griffin, J.D. & Gilliland, D.G. (2005) FLT3 internal tandem duplication mutations induce myeloproliferative or lymphoid disease in a transgenic mouse model. *Oncogene,* **24,** 7882-7892.

Leroy, H., Roumier, C., Huyghe, P., Biggio, V., Fenaux, P. & Preudhomme, C. (2005) CEBPA point mutations in hematological malignancies. *Leukemia,* **19,** 329-334.

Levine, R.L., Wadleigh, M., Cools, J., Ebert, B.L., Wernig, G., Huntly, B.J., Boggon, T.J., Wlodarska, I., Clark, J.J., Moore, S., Adelsperger, J., Koo, S., Lee, J.C., Gabriel, S., Mercher, T., D'Andrea, A., Frohling, S., Dohner, K., Marynen, P., Vandenberghe, P., Mesa, R.A., Tefferi, A., Griffin, J.D., Eck, M.J., Sellers, W.R., Meyerson, M., Golub, T.R., Lee, S.J. & Gilliland, D.G. (2005) Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell,* **7,** 387-397.

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

Lindblad-Toh, K., Tanenbaum, D.M., Daly, M.J., Winchester, E., Lui, W.O., Villapakkam, A., Stanton, S.E., Larsson, C., Hudson, T.J., Johnson, B.E., Lander, E.S. & Meyerson, M. (2000) Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. *Nat Biotechnol,* **18,** 1001-1005.

Linggi, B., Muller-Tidow, C., van de Locht, L., Hu, M., Nip, J., Serve, H., Berdel, W.E., van der Reijden, B., Quelle, D.E., Rowley, J.D., Cleveland, J., Jansen, J.H., Pandolfi, P.P. & Hiebert, S.W. (2002) The t(8;21) fusion protein, AML1-ETO, specifically represses the

transcription of the p14ARF tumor suppressor in acute myeloid leukemia. *Nat Med,* **8,** 743-750.

Liu, P., Jenkins, N.A. & Copeland, N.G. (2002) Efficient Cre-loxP-induced mitotic recombination in mouse embryonic stem cells. *Nat Genet,* **30,** 66-72.

Liu, T.X., Becker, M.W., Jelinek, J., Wu, W.S., Deng, M., Mikhalkevich, N., Hsu, K., Bloomfield, C.D., Stone, R.M., DeAngelo, D.J., Galinsky, I.A., Issa, J.P., Clarke, M.F. & Look, A.T. (2007) Chromosome 5q deletion and epigenetic suppression of the gene encoding alpha-catenin (CTNNA1) in myeloid cell transformation. *Nat Med,* **13,** 78-83.

Liu, W.M., Di, X., Yang, G., Matsuzaki, H., Huang, J., Mei, R., Ryder, T.B., Webster, T.A., Dong, S., Liu, G., Jones, K.W., Kennedy, G.C. & Kulp, D. (2003) Algorithms for large-scale genotyping microarrays. *Bioinformatics,* **19,** 2397-2403.

Lo, H.S., Wang, Z., Hu, Y., Yang, H.H., Gere, S., Buetow, K.H. & Lee, M.P. (2003) Allelic variation in gene expression is common in the human genome. *Genome Res,* **13,** 1855-1862.

Luo, G., Santoro, I.M., McDaniel, L.D., Nishijima, I., Mills, M., Youssoufian, H., Vogel, H., Schultz, R.A. & Bradley, A. (2000) Cancer predisposition caused by elevated mitotic recombination in Bloom mice. *Nat Genet,* **26,** 424-429.

Lutterbach, B., Hou, Y., Durst, K.L. & Hiebert, S.W. (1999) The inv(16) encodes an acute myeloid leukemia 1 transcriptional corepressor. *Proc Natl Acad Sci U S A,* **96,** 12822-12827.

Maley, C.C., Galipeau, P.C., Finley, J.C., Wongsurawat, V.J., Li, X., Sanchez, C.A., Paulson, T.G., Blount, P.L., Risques, R.-A., Rabinovitch, P.S. & Reid, B.J. (2006) Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat Genet,* **38,** 468-473.

Maley, C.C., Galipeau, P.C., Li, X., Sanchez, C.A., Paulson, T.G. & Reid, B.J. (2004) Selectively Advantageous Mutations and Hitchhikers in Neoplasms: p16 Lesions Are Selected in Barrett's Esophagus. *Cancer Res,* **64,** 3414-3427.

Malpas, J.S. & Scott, R.B. (1968) Rubidomycin in acute leukaemia in adults. *Br Med J,* **3,** 227-229.

Marsh, S. & McLeod, H.L. (2004) Cancer pharmacogenetics. *Br J Cancer,* **90,** 8-11.

Matheny, C.J., Speck, M.E., Cushing, P.R., Zhou, Y., Corpora, T., Regan, M., Newman, M., Roudaia, L., Speck, C.L., Gu, T.L., Griffey, S.M., Bushweller, J.H. & Speck, N.A. (2007)

Disease mutations in RUNX1 and RUNX2 create nonfunctional, dominant-negative, or hypomorphic alleles. *Embo J,* **26,** 1163-1175.

Mathew, C.G. (2006) Fanconi anaemia genes and susceptibility to cancer. *Oncogene,* **25,** 5875-5884.

Matsuzaki, H., Loi, H., Dong, S., Tsai, Y.Y., Fang, J., Law, J., Di, X., Liu, W.M., Yang, G., Liu, G., Huang, J., Kennedy, G.C., Ryder, T.B., Marcus, G.A., Walsh, P.S., Shriver, M.D., Puck, J.M., Jones, K.W. & Mei, R. (2004) Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res,* **14,** 414-425.

Mayer, R.J., Davis, R.B., Schiffer, C.A., Berg, D.T., Powell, B.L., Schulman, P., Omura, G.A., Moore, J.O., McIntyre, O.R., Frei, E. & The Cancer and Leukemia Group, B. (1994) Intensive Postremission Chemotherapy in Adults with Acute Myeloid Leukemia. *N Engl J Med,* **331,** 896-903.

Mead, A.J., Linch, D.C., Hills, R.K., Wheatley, K., Burnett, A.K. & Gale, R.E. (2007) FLT3 tyrosine kinase domain mutations are biologically distinct from and have a significantly more favorable prognosis than FLT3 internal tandem duplications in patients with acute myeloid leukemia. *Blood,* **110,** 1262-1270.

Mele, A., Szklo, M., Visani, G., Stazi, M.A., Castelli, G., Pasquini, P., Mandelli, F. & Italian Leukemia Study Group (1994) Hair Dye Use and Other Risk Factors for Leukemia and Pre-leukemia: A Case-Control Study. *Am. J. Epidemiol.,* **139,** 609-619.

Merlo, L.M., Pepper, J.W., Reid, B.J. & Maley, C.C. (2006) Cancer as an evolutionary and ecological process. *Nat Rev Cancer,* **6,** 924-935.

Micallef, I.N., Lillington, D.M., Apostolidis, J., Amess, J.A., Neat, M., Matthews, J., Clark, T., Foran, J.M., Salam, A., Lister, T.A. & Rohatiner, A.Z. (2000) Therapy-related myelodysplasia and secondary acute myelogenous leukemia after high-dose therapy with autologous hematopoietic progenitor-cell support for lymphoid malignancies. *J Clin Oncol,* **18,** 947-955.

Michels, S.D., McKenna, R.W., Arthur, D.C. & Brunning, R.D. (1985) Therapy-related acute myeloid leukemia and myelodysplastic syndrome: a clinical and morphologic study of 65 cases. *Blood,* **65,** 1364-1372.

Mikhail, F.M., Sinha, K.K., Saunthararajah, Y. & Nucifora, G. (2006) Normal and transforming functions of RUNX1: A perspective. *Journal of Cellular Physiology,* **207,** 582-593.

Milligan, D.W., Wheatley, K., Littlewood, T., Craig, J.I. & Burnett, A.K. (2006) Fludarabine and cytosine are less effective than standard ADE chemotherapy in high-risk acute myeloid leukemia, and addition of G-CSF and ATRA are not beneficial: results of the MRC AML-HR randomized trial. *Blood,* **107,** 4614-4622.

Mills, K. (2005) Microarray studies. AML16 Investigator Meeting, http://www.download.bham.ac.uk/bctu/aml16/Trial/DocsForExistingCentres/Microarray_kenMills.pdf

Minucci, S. & Pelicci, P.G. (2007) Determinants of Oncogenic Transformation in Acute Promyelocytic Leukemia: The Hetero-Union Makes the Force. *Cancer Cell,* **12,** 1-3.

Mistry, A.R., Felix, C.A., Whitmarsh, R.J., Mason, A., Reiter, A., Cassinat, B., Parry, A., Walz, C., Wiemels, J.L., Segal, M.R., Ades, L., Blair, I.A., Osheroff, N., Peniket, A.J., Lafage-Pochitaloff, M., Cross, N.C., Chomienne, C., Solomon, E., Fenaux, P. & Grimwade, D. (2005) DNA topoisomerase II in therapy-related acute promyelocytic leukemia. *N Engl J Med,* **352,** 1529-1538.

Mitelman, F. (1995) *ISCN 1995 An International System for Human Cytogenetic Nomenclature.* S. Karger, Basel.

Mitelman, F., Johansson, B. & Mertens, F. (2004) Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. *Nat Genet,* **36,** 331-334.

Miyagawa, K., Hayashi, Y., Fukuda, T., Mitani, K., Hirai, H. & Kamiya, K. (1999) Mutations of the WT1 gene in childhood nonlymphoid hematological malignancies. *Genes Chromosomes Cancer,* **25,** 176-183.

Mohty, M., de Lavallade, H., Ladaique, P., Faucher, C., Vey, N., Coso, D., Stoppa, A.M., Gastaut, J.A. & Blaise, D. (2005) The role of reduced intensity conditioning allogeneic stem cell transplantation in patients with acute myeloid leukemia: a donor vs no donor comparison. *Leukemia,* **19,** 916-920.

Monaghan, K.G., van Dyke, D.L., Wiktor, A. & Feldman, G.L. (1997) Cytogenetic and clinical

findings in a patient with a deletion of 16q23.1: First report of bilateral cataracts and a 16q deletion. *American Journal of Medical Genetics,* **73,** 180-183.

Moore, G.E., Gerner, R.E. & Franklin, H.A. (1967) Culture of normal human leukocytes. *Jama,* **199,** 519-524.

Morison, I.M., Ellis, L.M., Teague, L.R. & Reeve, A.E. (2002) Preferential loss of maternal 9p alleles in childhood acute lymphoblastic leukemia. *Blood,* **99,** 375-377.

Moynahan, M.E. & Jasin, M. (1997) Loss of heterozygosity induced by a chromosomal double-strand break. *Proc Natl Acad Sci U S A,* **94,** 8988-8993.

Mullighan, C.G., Goorha, S., Radtke, I., Miller, C.B., Coustan-Smith, E., Dalton, J.D., Girtman, K., Mathew, S., Ma, J., Pounds, S.B., Su, X., Pui, C.H., Relling, M.V., Evans, W.E., Shurtleff, S.A. & Downing, J.R. (2007) Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature,* **446,** 758-764.

Murthy, S.K., DiFrancesco, L.M., Ogilvie, R.T. & Demetrick, D.J. (2002) Loss of heterozygosity associated with uniparental disomy in breast carcinoma. *Mod Pathol,* **15,** 1241-1250.

Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C.M. & Canaani, E. (2002) ALL-1 Is a Histone Methyltransferase that Assembles a Supercomplex of Proteins Involved in Transcriptional Regulation. *Molecular Cell,* **10,** 1119-1128.

Nannya, Y., Sanada, M., Nakazaki, K., Hosoya, N., Wang, L., Hangaishi, A., Kurokawa, M., Chiba, S., Bailey, D.K., Kennedy, G.C. & Ogawa, S. (2005) A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res,* **65,** 6071-6079.

National Cancer Institute (2006) *Mitelman Database of Chromosome Aberrations in Cancer.* (ed. by F. Mitelman, B. Johansson & F. Mertens). John Wiley & Sons, Inc.

National Center for Biotechnology Information (2005) *Gene Expression Omnibus (GEO).*

National Center for Biotechnology Information (2007) *dbSNP.* NCBI.

NCI & NCBI (2001) *Cancer Chromosomes.*

Nowell, P.C. (1976) The clonal evolution of tumor cell populations. *Science,* **194,** 23-28.

Nowell, P.C. & Hungerford, D.A. (1960) A minute chromosome in human chronic granulocytic leukemia. *Science,* **132,** 1497.

Nyvold, C.G., Stentoft, J., Braendstrup, K., Melsvik, D., Moestrup, S.K., Juhl-Christensen, C., Hasle, H. & Hokland, P. (2006) Wilms' tumor 1 mutation accumulated during therapy in acute myeloid leukemia: biological and clinical implications. *Leukemia,* **20,** 2051-2054.

O'Driscoll, M. & Jeggo, P.A. (2006) The role of double-strand break repair - insights from human genetics. *Nat Rev Genet,* **7,** 45-54.

Office for National Statistics (2005) *Cancer statistics registrations.* In: Series MB1, Vol. no. 34. HMSO, London.

Office for National Statistics (2006) *Cancer statistics registrations.* In: Series MB1, Vol. no. 35. HMSO, London.

Offman, J., Opelz, G., Doehler, B., Cummins, D., Halil, O., Banner, N.R., Burke, M.M., Sullivan, D., Macpherson, P. & Karran, P. (2004) Defective DNA mismatch repair in acute myeloid leukemia/myelodysplastic syndrome after organ transplantation. *Blood,* **104,** 822-828.

Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V.M., Su, L., Xu, G. & Zhang, Y. (2005) hDOT1L Links Histone Methylation to Leukemogenesis. *Cell,* **121,** 167-178.

Oki, Y., Kantarjian, H.M., Zhou, X., Cortes, J., Faderl, S., Verstovsek, S., O'Brien, S., Koller, C., Beran, M., Bekele, B.N., Pierce, S., Thomas, D., Ravandi, F., Wierda, W.G., Giles, F., Ferrajoli, A., Jabbour, E., Keating, M.J., Bueso-Ramos, C.E., Estey, E. & Garcia-Manero, G. (2006) Adult acute megakaryocytic leukemia: an analysis of 37 patients treated at M.D. Anderson Cancer Center. *Blood,* **107,** 880-884.

Okuda, T., van Deursen, J., Hiebert, S.W., Grosveld, G. & Downing, J.R. (1996) AML1, the Target of Multiple Chromosomal Translocations in Human Leukemia, Is Essential for Normal Fetal Liver Hematopoiesis. *Cell,* **84,** 321-330.

Otto, F., Lübbert, M. & Stock, M. (2003) Upstream and downstream targets of RUNX proteins. *Journal of Cellular Biochemistry,* **89,** 9-18.

Owen, C., Barnett, M. & Fitzgibbon, J. (2008) Familial myelodysplasia and acute myeloid

leukaemia--a review. *Br J Haematol,* **140,** 123-132.

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

Pandolfi, P.P. (2001) Oncogenes and tumor suppressors in the molecular pathogenesis of acute promyelocytic leukemia. *Hum. Mol. Genet.,* **10,** 769-775.

Parada, L.F., Tabin, C.J., Shih, C. & Weinberg, R.A. (1982) Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus ras gene. *Nature,* **297,** 474-478.

Parcells, B.W., Ikeda, A.K., Simms-Waldrip, T., Moore, T.B. & Sakamoto, K.M. (2006) FLT3 in Normal Hematopoiesis and Acute Myeloid Leukemia. *Stem Cells.*

Parmentier, C. (2003) Use and risks of phosphorus-32 in the treatment of polycythaemia vera. *European Journal of Nuclear Medicine and Molecular Imaging,* **30,** 1413-1417.

Paschka, P., Marcucci, G., Ruppert, A.S., Mrozek, K., Chen, H., Kittles, R.A., Vukosavljevic, T., Perrotti, D., Vardiman, J.W., Carroll, A.J., Kolitz, J.E., Larson, R.A. & Bloomfield, C.D. (2006) Adverse prognostic significance of KIT mutations in adult acute myeloid leukemia with inv(16) and t(8;21): a Cancer and Leukemia Group B Study. *J Clin Oncol,* **24,** 3904-3911.

Paulsson, K., Morse, H., Fioretos, T., Behrendtz, M., Strombeck, B. & Johansson, B. (2005) Evidence for a single-step mechanism in the origin of hyperdiploid childhood acute lymphoblastic leukemia. *Genes Chromosomes Cancer.*

Paulsson, K., Panagopoulos, I., Knuutila, S., Jee, K.J., Garwicz, S., Fioretos, T., Mitelman, F. & Johansson, B. (2003) Formation of trisomies and their parental origin in hyperdiploid childhood acute lymphoblastic leukemia. *Blood,* **102,** 3010-3015.

Peterson, L.F. & Zhang, D.E. (2004) The 8;21 translocation in leukemogenesis. *Oncogene,* **23,** 4255-4262.

Piller, G. (2001) Leukaemia - a brief historical review from ancient times to 1950. *Br J Haematol,* **112,** 282-292.

Pogoda, J.M., Preston-Martin, S., Nichols, P.W. & Ross, R.K. (2002) Smoking and risk of acute

myeloid leukemia: results from a Los Angeles County case-control study. *Am J Epidemiol,* **155,** 546-553.

Prado, F., Cortes-Ledesma, F., Huertas, P. & Aguilera, A. (2003) Mitotic recombination in Saccharomyces cerevisiae. *Curr Genet,* **42,** 185-198.

Preudhomme, C., Warot-Loze, D., Roumier, C., Grardel-Duflos, N., Garand, R., Lai, J.L., Dastugue, N., Macintyre, E., Denis, C., Bauters, F., Kerckaert, J.P., Cosson, A. & Fenaux, P. (2000) High incidence of biallelic point mutations in the Runt domain of the AML1/PEBP2alpha B gene in Mo acute myeloid leukemia and in myeloid malignancies with acquired trisomy 21. *Blood,* **96,** 2862-2869.

Purdie, K.J., Lambert, S.R., Teh, M.T., Chaplin, T., Molloy, G., Raghavan, M., Kelsell, D.P., Leigh, I.M., Harwood, C.A., Proby, C.M. & Young, B.D. (2007) Allelic imbalances and microdeletions affecting the PTPRD gene in cutaneous squamous cell carcinomas detected using single nucleotide polymorphism microarray analysis. *Genes Chromosomes Cancer,* **46,** 661-669.

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

Ramsay, G. (1998) DNA chips: State-of-the art. *Nat Biotech,* **16,** 40-44.

Rauscher, G.H., Shore, D. & Sandler, D.P. (2004) Hair dye use and risk of adult acute leukemia. *Am J Epidemiol,* **160,** 19-25.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W. & Hurles, M.E. (2006) Global variation in copy number in the human genome. *Nature,*

444, 444-454.

Rege-Cambrin, G., Giugliano, E., Michaux, L., Stul, M., Scaravaglio, P., Serra, A., Saglio, G. & Hagemeijer, A. (2005) Trisomy 11 in myeloid malignancies is associated with internal tandem duplication of both MLL and FLT3 genes. *Haematologica,* **90,** 262-264.

Richardson, C., Moynahan, M.E. & Jasin, M. (1998) Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. *Genes Dev,* **12,** 3831-3842.

Rohatiner, A.Z., Bassan, R., Raimondi, R., Amess, J.A., Arnott, S., Personen, A., Rodeghiero, F., Barbui, T., Bradburn, M.J., Carter, M. & Lister, T.A. (2000) High-dose treatment with autologous bone marrow support as consolidation of first remission in younger patients with acute myelogenous leukaemia. *Ann Oncol,* **11,** 1007-1015.

Rombouts, E.J.C., Pavic, B., Lowenberg, B. & Ploemacher, R.E. (2004) Relation between CXCR-4 expression, Flt3 mutations, and unfavorable prognosis of adult acute myeloid leukemia. *Blood,* **104,** 550-557.

Rosenberg, P.S., Alter, B.P., Bolyard, A.A., Bonilla, M.A., Boxer, L.A., Cham, B., Fier, C., Freedman, M., Kannourakis, G., Kinsey, S., Schwinzer, B., Zeidler, C., Welte, K., Dale, D.C. & for the Severe Chronic Neutropenia International, R. (2006) The incidence of leukemia and mortality from sepsis in patients with severe congenital neutropenia receiving long-term G-CSF therapy. *Blood,* **107,** 4628-4635.

Roumier, C., Eclache, V., Imbert, M., Davi, F., MacIntyre, E., Garand, R., Talmant, P., Lepelley, P., Lai, J.L., Casasnovas, O., Maynadie, M., Mugneret, F., Bilhou-Naberra, C., Valensi, F., Radford, I., Mozziconacci, M.J., Arnoulet, C., Duchayne, E., Dastugue, N., Cornillet, P., Daliphard, S., Garnache, F., Boudjerra, N., Jouault, H., Fenneteau, O., Pedron, B., Berger, R., Flandrin, G., Fenaux, P. & Preudhomme, C. (2003) M0 AML, clinical and biologic features of the disease, including AML1 gene mutations: a report of 59 cases by the Groupe Francais d'Hematologie Cellulaire (GFHC) and the Groupe Francais de Cytogenetique Hematologique (GFCH). *Blood,* **101,** 1277-1283.

Roumier, C., Lejeune-Dumoulin, S., Renneville, A., Goethgeluck, A.S., Philippe, N., Fenaux, P.

& Preudhomme, C. (2006) Cooperation of activating Ras//rtk signal transduction pathway mutations and inactivating myeloid differentiation gene mutations in M0 AML: a study of 45 patients. *Leukemia,* **20,** 433-436.

Rucker, F.G., Bullinger, L., Schwaenen, C., Lipka, D.B., Wessendorf, S., Frohling, S., Bentz, M., Miller, S., Scholl, C., Schlenk, R.F., Radlwimmer, B., Kestler, H.A., Pollack, J.R., Lichter, P., Dohner, K. & Dohner, H. (2006) Disclosure of Candidate Genes in Acute Myeloid Leukemia With Complex Karyotypes Using Microarray-Based Molecular Characterization. *J Clin Oncol.*

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S. & Altshuler, D. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* **409,** 928-933.

Sakatani, T., Kaneda, A., Iacobuzio-Donahue, C.A., Carter, M.G., de Boom Witzel, S., Okano, H., Ko, M.S., Ohlsson, R., Longo, D.L. & Feinberg, A.P. (2005) Loss of imprinting of Igf2 alters intestinal maturation and tumorigenesis in mice. *Science,* **307,** 1976-1978.

Schmid, D., Heinze, G., Linnerth, B., Tisljar, K., Kusec, R., Geissler, K., Sillaber, C., Laczika, K., Mitterbauer, M., Zochbauer, S., Mannhalter, C., Haas, O.A., Lechner, K., Jager, U. & Gaiger, A. (1997) Prognostic significance of WT1 gene expression at diagnosis in adult de novo acute myeloid leukemia. *Leukemia,* **11,** 639-643.

Schoch, C., Kohlmann, A., Dugas, M., Kern, W., Schnittger, S. & Haferlach, T. (2006) Impact of trisomy 8 on expression of genes located on chromosome 8 in different AML subgroups. *Genes, Chromosomes and Cancer,* **45,** 1164-1168.

Scholl, C., Bansal, D., Dohner, K., Eiwen, K., Huntly, B.J.P., Lee, B.H., Rucker, F.G., Schlenk, R.F., Bullinger, L., Dohner, H., Gilliland, D.G. & Frohling, S. (2007) The homeobox gene

CDX2 is aberrantly expressed in most cases of acute myeloid leukemia and promotes leukemogenesis. *J. Clin. Invest.*, **117,** 1037-1048.

Schubbert, S., Shannon, K. & Bollag, G. (2007) Hyperactive Ras in developmental disorders and cancer. *Nat Rev Cancer,* **7,** 295-308.

Schubert, E.L., Hsu, L., Cousens, L.A., Glogovac, J., Self, S., Reid, B.J., Rabinovitch, P.S. & Porter, P.L. (2002) Single nucleotide polymorphism array analysis of flow-sorted epithelial cells from frozen versus fixed tissues for whole genome analysis of allelic loss in breast cancer. *Am J Pathol,* **160,** 73-79.

Schull, W.J. (1998) The somatic effects of exposure to atomic radiation: The Japanese experience, 1947-1997

. *PNAS,* **95,** 5437-5441.

Scott, R.B. (1957) Leukaemia. *Lancet,* **272,** 1053-1057.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T.C., Trask, B., Patterson, N., Zetterberg, A. & Wigler, M. (2004) Large-scale copy number polymorphism in the human genome. *Science,* **305,** 525-528.

Seedhouse, C., Bainton, R., Lewis, M., Harding, A., Russell, N. & Das-Gupta, E. (2002) The genotype distribution of the XRCC1 gene indicates a role for base excision repair in the development of therapy-related acute myeloblastic leukemia. *Blood,* **100,** 3761-3766.

Seedhouse, C., Faulkner, R., Ashraf, N., Das-Gupta, E. & Russell, N. (2004) Polymorphisms in genes involved in homologous recombination repair interact to increase the risk of developing acute myeloid leukemia. *Clin Cancer Res,* **10,** 2675-2680.

Seedhouse, C.H., Hunter, H.M., Lloyd-Lewis, B., Massip, A.M., Pallis, M., Carter, G.I., Grundy, M., Shang, S. & Russell, N.H. (2006) DNA repair contributes to the drug-resistant phenotype of primary acute myeloid leukaemia cells with FLT3 internal tandem duplications and is reversed by the FLT3 inhibitor PKC412. *Leukemia,* **20,** 2130-2136.

Sengupta, S. & Harris, C.C. (2005) p53: traffic cop at the crossroads of DNA repair and recombination. *Nat Rev Mol Cell Biol,* **6,** 44-55.

Shearer, P.D., Valentine, M.B., Grundy, P., DeCou, J.M., Banavali, S.D., Komuro, H., Green, D.M., Beckwith, J.B. & Look, A.T. (1999) Hemizygous deletions of chromosome band 16q24 in Wilms tumor: detection by fluorescence in situ hybridization. *Cancer Genet Cytogenet,* **115,** 100-105.

Sievers, E.L., Larson, R.A., Stadtmauer, E.A., Estey, E., Lowenberg, B., Dombret, H., Karanes, C., Theobald, M., Bennett, J.M., Sherman, M.L., Berger, M.S., Eten, C.B., Loken, M.R., van Dongen, J.J.M., Bernstein, I.D. & Appelbaum, F.R. (2001) Efficacy and Safety of Gemtuzumab Ozogamicin in Patients With CD33-Positive Acute Myeloid Leukemia in First Relapse. *J Clin Oncol,* **19,** 3244-3254.

Silva, F.P., Morolli, B., Storlazzi, C.T., Anelli, L., Wessels, H., Bezrookove, V., Kluin-Nelemans, H.C. & Giphart-Gassler, M. (2003) Identification of RUNX1/AML1 as a classical tumor suppressor gene. *Oncogene,* **22,** 538-547.

Slovak, M.L., Kopecky, K.J., Cassileth, P.A., Harrington, D.H., Theil, K.S., Mohamed, A., Paietta, E., Willman, C.L., Head, D.R., Rowe, J.M., Forman, S.J. & Appelbaum, F.R. (2000) Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group study. *Blood,* **96,** 4075-4083.

Smith, B.D., Levis, M., Beran, M., Giles, F., Kantarjian, H., Berg, K., Murphy, K.M., Dauses, T., Allebach, J. & Small, D. (2004a) Single-agent CEP-701, a novel FLT3 inhibitor, shows biologic and clinical activity in patients with relapsed or refractory acute myeloid leukemia. *Blood,* **103,** 3669-3676.

Smith, M.L., Arch, R., Smith, L.L., Bainton, N., Neat, M., Taylor, C., Bonnet, D., Cavenagh, J.D., Andrew Lister, T. & Fitzgibbon, J. (2005) Development of a human acute myeloid leukaemia screening panel and consequent identification of novel gene mutation in FLT3 and CCND3. *Br J Haematol,* **128,** 318-323.

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

Snaddon, J., Smith, M.L., Neat, M., Cambal-Parrales, M., Dixon-McIver, A., Arch, R., Amess,

J.A., Rohatiner, A.Z., Lister, T.A. & Fitzgibbon, J. (2003) Mutations of CEBPA in acute myeloid leukemia FAB types M1 and M2. *Genes Chromosomes Cancer,* **37,** 72-78.

Song, W.-J., Sullivan, M.G., Legare, R.D., Hutchings, S., Tan, X., Kufrin, D., Ratajczak, J., Resende, I.C., Haworth, C., Hock, R., Loh, M., Felix, C., Roy, D.-C., Busque, L., Kurnit, D., Willman, C., Gewirtz, A.M., Speck, N.A., Bushweller, J.H., Li, F.P., Gardiner, K., Poncz, M., Maris, J.M. & Gilliland, D.G. (1999) Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet,* **23,** 166-175.

Stephens, K., Weaver, M., Leppig, K.A., Maruyama, K., Emanuel, P.D., Le Beau, M.M. & Shannon, K.M. (2006) Interstitial uniparental isodisomy at clustered breakpoint intervals is a frequent mechanism of NF1 inactivation in myeloid malignancies. *Blood,* **108,** 1684-1689.

Stern, C. (1936) Somatic crossing over and segregation in drosophila melanogaster. *Genetics,* **21,** 625-730.

Stone, R.M., DeAngelo, D.J., Klimek, V., Galinsky, I., Estey, E., Nimer, S.D., Grandin, W., Lebwohl, D., Wang, Y., Cohen, P., Fox, E.A., Neuberg, D., Clark, J., Gilliland, D.G. & Griffin, J.D. (2005) Patients with acute myeloid leukemia and an activating mutation in FLT3 respond to a small-molecule FLT3 tyrosine kinase inhibitor, PKC412. *Blood,* **105,** 54-60.

Strefford, J.C., van Delft, F.W., Robinson, H.M., Worley, H., Yiannikouris, O., Selzer, R., Richmond, T., Hann, I., Bellotti, T., Raghavan, M., Young, B.D., Saha, V. & Harrison, C.J. (2006) Complex genomic alterations and gene expression in acute lymphoblastic leukemia with intrachromosomal amplification of chromosome 21. *Proc Natl Acad Sci U S A,* **103,** 8167-8172.

Suciu, S., Mandelli, F., de Witte, T., Zittoun, R., Gallo, E., Labar, B., De Rosa, G., Belhabri, A., Giustolisi, R., Delarue, R., Liso, V., Mirto, S., Leone, G., Bourhis, J.-H., Fioritoni, G., Jehn, U., Amadori, S., Fazi, P., Hagemeijer, A. & Willemze, R. (2003) Allogeneic compared with autologous stem cell transplantation in the treatment of patients younger than 46 years with acute myeloid leukemia (AML) in first complete remission (CR1): an intention-to-treat analysis of the EORTC/GIMEMAAML-10 trial. *Blood,* **102,** 1232-1240.

Summers, K., Stevens, J., Kakkas, I., Smith, M., Smith, L.L., MacDougall, F., Cavenagh, J.,

Bonnet, D., Young, B.D., Lister, T.A. & Fitzgibbon, J. (2007) Wilms' tumour 1 mutations are associated with FLT3-ITD and failure of standard induction chemotherapy in patients with normal karyotype AML. *Leukemia,* **21,** 550-551.

Sung, P. & Klein, H. (2006) Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat Rev Mol Cell Biol,* **7,** 739-750.

Super, H.J., McCabe, N.R., Thirman, M.J., Larson, R.A., Le Beau, M.M., Pedersen-Bjergaard, J., Philip, P., Diaz, M.O. & Rowley, J.D. (1993) Rearrangements of the MLL gene in therapy-related acute myeloid leukemia in patients previously treated with agents targeting DNA-topoisomerase II. *Blood,* **82,** 3705-3711.

Takata, M., Sasaki, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A. & Takeda, S. (1998) Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *Embo J,* **17,** 5497-5508.

Tallman, M.S., Andersen, J.W., Schiffer, C.A., Appelbaum, F.R., Feusner, J.H., Ogden, A., Shepherd, L., Willman, C., Bloomfield, C.D., Rowe, J.M. & Wiernik, P.H. (1997) All-trans-Retinoic Acid in Acute Promyelocytic Leukemia. *N Engl J Med,* **337,** 1021-1028.

Talpaz, M., Shah, N.P., Kantarjian, H., Donato, N., Nicoll, J., Paquette, R., Cortes, J., O'Brien, S., Nicaise, C., Bleickardt, E., Blackwood-Chirchir, M.A., Iyer, V., Chen, T.T., Huang, F., Decillis, A.P. & Sawyers, C.L. (2006) Dasatinib in imatinib-resistant Philadelphia chromosome-positive leukemias. *N Engl J Med,* **354,** 2531-2541.

Tanner, S.M., Austin, J.L., Leone, G., Rush, L.J., Plass, C., Heinonen, K., Mrozek, K., Sill, H., Knuutila, S., Kolitz, J.E., Archer, K.J., Caligiuri, M.A., Bloomfield, C.D. & de la Chapelle, A. (2001) BAALC, the human member of a novel mammalian neuroectoderm gene lineage, is implicated in hematopoiesis and acute leukemia. *Proceedings of the National Academy of Sciences,* **98,** 13901-13906.

Teh, M.T., Blaydon, D., Chaplin, T., Foot, N.J., Skoulakis, S., Raghavan, M., Harwood, C.A., Proby, C.M., Philpott, M.P., Young, B.D. & Kelsell, D.P. (2005) Genomewide single nucleotide polymorphism microarray mapping in basal cell carcinomas unveils uniparental disomy as a key

somatic event. *Cancer Res,* **65,** 8597-8603.

Thiede, C., Steudel, C., Mohr, B., Schaich, M., Schakel, U., Platzbecker, U., Wermke, M., Bornhauser, M., Ritter, M., Neubauer, A., Ehninger, G. & Illmer, T. (2002) Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. *Blood,* **99,** 4326-4335.

Tischfield, J.A. (1997) Loss of heterozygosity or: how I learned to stop worrying and love mitotic recombination. *Am J Hum Genet,* **61,** 995-999.

Valk, P.J., Verhaak, R.G., Beijen, M.A., Erpelinck, C.A., Barjesteh van Waalwijk van Doorn-Khosrovani, S., Boer, J.M., Beverloo, H.B., Moorhouse, M.J., van der Spek, P.J., Lowenberg, B. & Delwel, R. (2004) Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med,* **350,** 1617-1628.

van Dongen, J.J., van der Velden, V.H., de Ridder, D., Langerak, A.W. & Staal, F.J.T. (2005) Laboratory diagnosis of leukemia: can we replace current molecular diagnostics by novel flow cytometry? *Hematology (EHA Educ Program),* **1,** 36-40.

Virappane, P., Gale, R., Hills, R., Kakkas, I., Summers, K., Stevens, J., Allen, C., Green, C., Quentmeier, H., Drexler, H., Burnett, A., Linch, D., Bonnet, D., Lister, T.A. & Fitzgibbon, J. (2008) Mutation of the Wilms' Tumor 1 Gene Is a Poor Prognostic Factor Associated With Chemotherapy Resistance in Normal Karyotype Acute Myeloid Leukemia: The United Kingdom Medical Research Council Adult Leukaemia Working Party. *J Clin Oncol,* JCO.2008.2016.0333.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Nakamura, Y., White, R., Smits, A.M. & Bos, J.L. (1988) Genetic alterations during colorectal-tumor development. *N Engl J Med,* **319,** 525-532.

Vyas, P. & Roberts, I. (2006) Down myeloid disorders: A paradigm for childhood preleukaemia and leukaemia and insights into normal megakaryopoiesis. *Early Human Development,* **82,** 767-773.

Wagner, K.-D., Wagner, N. & Schedl, A. (2003) The complex life of WT1. *J Cell Sci,* **116,** 1653-1658.

Walker, B.A., Leone, P.E., Jenner, M.W., Li, C., Gonzalez, D., Johnson, D.C., Ross, F.M., Davies, F.E. & Morgan, G.J. (2006) Integration of global SNP-based mapping and expression arrays reveals key regions, mechanisms, and genes important in the pathogenesis of multiple myeloma. *Blood,* **108,** 1733-1743.

Wang, J., Hoshino, T., Redner, R.L., Kajigaya, S. & Liu, J.M. (1998) ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. *Proceedings of the National Academy of Sciences,* **95,** 10860-10865.

Wang, W., Warren, M. & Bradley, A. (2007) From the Cover: Induced mitotic recombination of p53 in vivo. *Proceedings of the National Academy of Sciences,* **104,** 4501-4505.

Weinstein, I.B. & Joe, A.K. (2006) Mechanisms of disease: Oncogene addiction--a rationale for molecular targeting in cancer therapy. *Nat Clin Pract Oncol,* **3,** 448-457.

White, V.A., McNeil, B.K. & Horsman, D.E. (1998) Acquired homozygosity (isodisomy) of chromosome 3 in uveal melanoma. *Cancer Genet Cytogenet,* **102,** 40-45.

Whitman, S.P., Archer, K.J., Feng, L., Baldus, C., Becknell, B., Carlson, B.D., Carroll, A.J., Mrozek, K., Vardiman, J.W., George, S.L., Kolitz, J.E., Larson, R.A., Bloomfield, C.D. & Caligiuri, M.A. (2001) Absence of the wild-type allele predicts poor prognosis in adult de novo acute myeloid leukemia with normal cytogenetics and the internal tandem duplication of FLT3: a cancer and leukemia group B study. *Cancer Res,* **61,** 7233-7239.

Whitman, S.P., Liu, S., Vukosavljevic, T., Rush, L.J., Yu, L., Liu, C., Klisovic, M.I., Maharry, K., Guimond, M., Strout, M.P., Becknell, B., Dorrance, A., Klisovic, R.B., Plass, C., Bloomfield, C.D., Marcucci, G. & Caligiuri, M.A. (2005) The MLL partial tandem duplication: evidence for recessive gain-of-function in acute myeloid leukemia identifies a novel patient subgroup for molecular-targeted therapy. *Blood,* **106,** 345-352.

Whitman, S.P., Ruppert, A.S., Marcucci, G., Mrozek, K., Paschka, P., Langer, C., Baldus, C.D., Wen, J., Vukosavljevic, T., Powell, B.L., Carroll, A.J., Kolitz, J.E., Larson, R.A., Caligiuri, M.A. & Bloomfield, C.D. (2007a) Long-term disease-free survivors with cytogenetically normal acute myeloid leukemia and MLL partial tandem duplication: a Cancer and Leukemia

Group B study. *Blood.*

Whitman, S.P., Ruppert, A.S., Radmacher, M.D., Mrozek, K., Paschka, P., Langer, C., Baldus, C.D., Wen, J., Racke, F., Powell, B.L., Kolitz, J.E., Larson, R.A., Caligiuri, M.A., Marcucci, G. & Bloomfield, C.D. (2007b) *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.* In: Blood.

Wicker, N., Carles, A., Mills, I.G., Wolf, M., Veerakumarasivam, A., Edgren, H., Boileau, F., Wasylyk, B., Schalken, J.A., Neal, D.E., Kallioniemi, O. & Poch, O. (2007) A new look towards BAC-based array CGH through a comprehensive comparison with oligo-based array CGH. *BMC Genomics,* **8,** 84.

Wodnar-Filipowicz, A. (2003) Flt3 Ligand: Role in Control of Hematopoietic and Immune Functions of the Bone Marrow. *News Physiol Sci,* **18,** 247-251.

Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B. & Kinzler, K.W. (2002) Allelic variation in human gene expression. *Science,* **297,** 1143.

Yan, M., Kanbe, E., Peterson, L.F., Boyapati, A., Miao, Y., Wang, Y., Chen, I.M., Chen, Z., Rowley, J.D., Willman, C.L. & Zhang, D.E. (2006) A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. *Nat Med,* **12,** 945-949.

Yates, J.W., Wallace, H.J., Jr., Ellison, R.R. & Holland, J.F. (1973) Cytosine arabinoside (NSC-63878) and daunorubicin (NSC-83142) therapy in acute nonlymphocytic leukemia. *Cancer Chemother Rep,* **57,** 485-488.

Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N., Wang, Z., Welch, R., Staats, B.J., Calle, E.E., Feigelson, H.S., Thun, M.J., Rodriguez, C., Albanes, D., Virtamo, J., Weinstein, S., Schumacher, F.R., Giovannucci, E., Willett, W.C., Cancel-Tassin, G., Cussenot, O., Valeri, A., Andriole, G.L., Gelmann, E.P., Tucker, M., Gerhard, D.S., Fraumeni, J.F., Jr., Hoover, R., Hunter, D.J., Chanock, S.J. & Thomas, G. (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet,* **39,** 645-649.

Yokoyama, A., Somervaille, T.C.P., Smith, K.S., Rozenblatt-Rosen, O., Meyerson, M. & Cleary, M.L. (2005) The Menin Tumor Suppressor Protein Is an Essential Oncogenic Cofactor for MLL-Associated Leukemogenesis. *Cell,* **123,** 207-218.

Zhao, X., Li, C., Paez, J.G., Chin, K., Janne, P.A., Chen, T.H., Girard, L., Minna, J., Christiani, D., Leo, C., Gray, J.W., Sellers, W.R. & Meyerson, M. (2004) An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res,* **64,** 3060-3071.

Zwaan, C.M., Kaspers, G.J., Pieters, R., Ramakers-Van Woerden, N.L., den Boer, M.L., Wunsche, R., Rottier, M.M., Hahlen, K., van Wering, E.R., Janka-Schaub, G.E., Creutzig, U. & Veerman, A.J. (2000) Cellular drug resistance profiles in childhood acute myeloid leukemia: differences between FAB types and comparison with acute lymphoblastic leukemia. *Blood,* **96,** 2879-2886.