Genetic and epigenetic consequences of radiation exposure in human and mouse leukaemogenesis

A thesis submitted for the degree of Doctor of Philosophy

By

Gráinne O'Brien

Department of Life Sciences, Brunel University London

December 2019

Abstract

Ionising radiation (IR) is a well-known carcinogen. For example, there is a dose-dependent increase of cancer incidence seen in the atomic bomb survivors in Japan. Acute Myeloid Leukaemia (AML) is one of the most common cancers to occur in humans following IR exposure. It can also be induced by radiotherapy treatment - so called therapy-related or secondary AML. Although widely studied, the underlying mechanisms of radiation-induced AML (rAML) are yet to be fully characterised. The main purpose of this research is to examine the target cells for rAML development, the hematopoietic stem and progenitor cells (HSPC). Previous studies have allowed classification of HSPC into three sub groups based on their repopulating abilities (long term HSC, short term HSC and haematopoietic progenitor cells (HPC)). We aim to characterise the response and sensitivity of these sub-populations of HSPC to IR. Recent work has focused on analysing the gene expression profiles of these sub-populations. We will expand on this by studying modifications of gene expression and methylation changes in these sub-populations following ionising radiation exposure in order to improve our understanding of the mechanisms of radiation-induced leukaemogenesis. Mouse and human samples of rAML will be used during this project with the aim to characterise the molecular mechanisms of rAML induction, assessing the suitability of the mouse model for humans and making for a first time an interspecies comparison analysis.

Declaration of originality

I declare that work presented in this thesis is my own. Work done by others or in collaboration has been appropriately acknowledged in the materials and methods and results chapters.

Some of the work presented here has been accepted for publication in a peer reviewed journal.

Table of Contents

List of figures

List of tables

Abbreviations

7-AAD - 7 amino-actinomycin D
ACGH - Array Comparative Genomic Hybridisation
ALL - acute lymphoblastic leukaemia
AMKL - acute megakaryoblastic leukaemia
AML – acute myeloid leukaemia
ASXL1 - ASXL Transcriptional Regulator 1
ATLL - Adult T-cell leukaemia/lymphoma
AWERB - Animal Welfare Ethical Review Body
AXIN2 - axin 2
BAALC - BAALC Binder Of MAP3K1 And KLF4
Bcl-2 - B-cell lymphoma 2
BHQ1 - Black Hole Quencher 1
Bmi1 – BMI1 proto-Oncogene, polycomb ring finger
BRCA1 - breast cancer type 1 susceptibility protein 1
BrdU – Bromodeoxyuridine
CA - cortistatin A
CBFB - Core-Binding Factor Subunit Beta
CBP - CREB binding protein
Ccna2 - Cyclin A2
Ccnb1 - Cyclin B1
Ccnd3 - Cyclin D3
CD – Cluster of differentiation
Cdkn1c – Cyclin Dependent Kinase Inhibitor 1C
cDNA - complementary DNA
CEBPA - CCAAT Enhancer Binding Protein Alpha
CFU - colony forming unit
CGIs - CpG islands
CHIP - Clonal hematopoiesis of indeterminate potential
CLL - Chronic lymphocytic leukaemia
CML - chronic myeloid leukaemia
CN-AML - cytogenetically normal AML
CR – complete remission
CT – Computed tomography
CTBP - C-terminal binding protein
CTIP/RBBP8 - retinoblastoma binding protein 8
CYP2B6 - Cytochrome P450 Family 2 Subfamily B Member 6
DDQ1 - Deep Dark Quencher 1
DKK - Dickkopf WNT Signaling Pathway Inhibitor 1
DNA - Deoxyribonucleic acid
DNMT3A - DNA (cytosine-5)-methyltransferase 3A
dNTP - deoxyribonucleotide triphosphate
DRU – distal regulatory unit
EGFR – epidermal growth factor receptor
ELN - European LeukemiaNet
ERG - ETS Transcription Factor
ETS - E26 transformation-specific
EZH2 - Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit
FAB - French-American-British

FACS - Fluorescence-activated cell sorting
FAM - 3' 6-carboxyfluorescein
FISH - fluorescent in situ hybridization
FITC - fluorescein isothiocyante
FLT3 - Fms Related Tyrosine Kinase 3
FOV - field of view
GATA1 - GATA Binding Protein 1
GBM - glioblastoma multiforme
G-CSF - granulocyte colony stimulating factor
GM-CSF - granulocyte-macrophage colony-stimulating factor
HAT - histone acetyltransferases
HCT - hematopoietic cell transplantation
HDAC - histone deacetylases
HER2 - human epidermal growth factor
HEX - Hexachlorofluorescein
HOX – Homeobox
HPRT - Hypoxanthine-Guanine phosphoribosyltransferase 1
HSC – hematopoietic stem cell
HSPC - hematopoietic stem and progenitor cell
IDH1 – Isocitrate Dehydrogenase (NADP(+)) 1
IDH2 - Isocitrate Dehydrogenase (NADP(+)) 2
IFN- interferon
Il1r1 - Interleukin 1 Receptor Type 1
IL – Interleukin
IMDM - Iscove's Modified Dulbecco's Media
IPSCs - induced pluripotent stem cells
IR - ionising radiation
ITD - internal tandem duplication
Itga6 – Integrin subunit alpha 6
JAK - Janus Kinase
Jun - Jun Proto-Oncogene, AP-1 Transcription Factor Subunit
KIT - KIT Proto-Oncogene
KRAS - Kirsten Rat Sarcoma Viral Oncogene Homolog
Lin- - lineage negative
LSK - Lin-Sca1+c-kit+
LSS – life span study
M-CSF - macrophage colony-stimulating factor
Mac-1 - Macrophage-1 antigen
MBD - methyl-CpG-binding domain
Mcm5 - minichromosome maintenance complex component 5
MDS - myelodysplastic syndrome
MEL - murine erythroleukaemia
miRNA – microRNA
MLL - mixed lineage leukemia
Mllt4 - Mixed-lineage leukaemia 4
MPP - multipotent progenitor
MQRT-PCR – multiplex quantitative real time polymerase chain reaction
mRNA – messenger RNA
MTase – methyltransferase
MUC1 - mucin 1

Myb - MYB Proto-Oncogene, Transcription Factor
MYH11 - Myosin Heavy Chain 11
NFκB - nuclear factor kappa B subunit 1
NICR - Northern Institute for Cancer Research
Nmi - n-Myc and STAT interactor
NPM1 - nucleophosmin 1
NQ01 - NAD(P)H:quinone oxidoreductase
NRAS - Neuroblastoma RAS Viral (V-Ras) Oncogene Homolog
NsSNP - nonsynonymous single nucleotide polymorphism
Pbx1 - PBX Homeobox 1
PCR – polymerase chain reaction
PE - phycoerythrin
PHF6 - PHD Finger Protein 6
PIAS - protein inhibitors of activated STATs
PML - Promyelocytic Leukaemia
PP2A - protein phosphatase 2A
PROVEAN - PROtein Variation Effect Analyzer
PRU - proximal regulatory unit
PTEN - phosphatase and tensin homolog
qPCR – quantitative polymerase chain reaction
rAML – radiation-induced acute myeloid leukaemia
RARA - Retinoic Acid Receptor Alpha
RINe - RNA Integrity Number equivalent
RNA - Ribonucleic acid
RR - relative risk
RUNX1 - Runt-related transcription factor 1
SATB1 - special AT-rich sequence binding protein 1
sAML - secondary AML
sca1 - Stem cell antigen-1
Sesn2 - Sestrin 2
SET - SET Nuclear Proto-Oncogene
SETBP1 - SET Binding Protein 1
Sfpi1/PU.1 – Spi-1 Proto-Oncogene
SFRP1 - Secreted Frizzled Related Protein 1
SLAM – signaling lymphocytic activation molecule
SNP - single nucleotide polymorphisms
Socs2 - suppressor of cytokine signaling 2
SP1 - specificity protein 1
SRA - SET- and Ring finger-associated
SRSF2 - Serine And Arginine Rich Splicing Factor 2
STAT - Signal Transducer and Activator Of Transcription
t-AML - therapy-related AML
TBE - Tris-Borate-EDTA
TET2 - Tet Methylcytosine Dioxygenase 2
TEX - Texas Red
TGF-β - transforming growth factor beta
Thy1.1 - Thy-1 Membrane Glycoprotein
tMDS - therapy-related acute myeloid leukaemia
TNFRSF5 - Tumor Necrosis Factor Receptor Superfamily Member 5
TP53 - Tumor Protein P53

TRD - transcriptional repression domain
URE – upstream regulatory element
VEP - Variant Effect Predictor
WHO - World Health Organisation
WGA - whole genome amplification
WT1 - Wilms Tumor 1
WTA - whole transcriptome amplification

Acknowledgements

1 Introduction

1.1 Human blood

Human blood functions to transport nutrients to cells, transport waste away from cells and protect the body from foreign substances or infections. Blood cells develop by the process of haematopoiesis, whereby a hematopoietic stem cell, the originating cell of all blood cells, differentiates first into a progenitor which differentiates further into a more specialised cell type (Figure 1). Hematopoietic stem cells reside in the bone marrow and are multipotent, in that they can self-renew and can give rise to any blood cell. Hematopoietic stem cells can differentiate into a myeloid or lymphoid progenitor. Myeloid progenitors can differentiate into many types of cells such as thrombocytes,
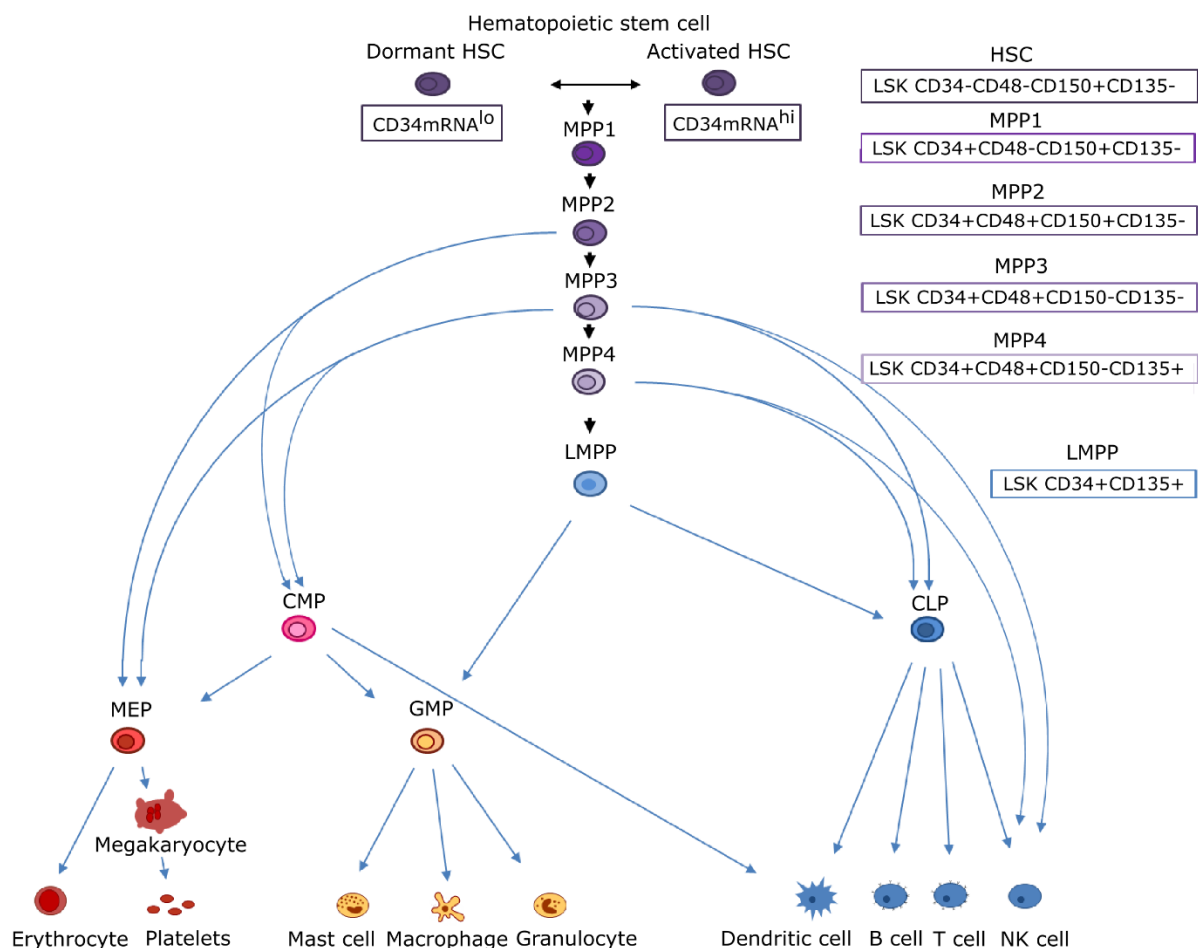
Figure 1. The development of blood cells.

This image shows the formation, development, and differentiation of blood cells from a stem cell to differentiation by either myeloid or lymphoid pathways. Adapted from Wilson et al. 2008 and Wilson et al. 2015.

erythrocytes, granulocytes, monocytes, macrophages and dendritic cells. Lymphoid progenitors can differentiate into T cells, B cells and natural killer cells. Lymphoid and myeloid cells function to protect the body against infectious disease and foreign bodies.

1.2 Acute myeloid leukaemia

Leukaemia is a cancer of the blood that results in an overproduction of immature blood cells, known as blast cells. This accumulation of abnormal immature cells leaves little room for the replenishment of normal functioning blood cells which leads to a decrease in erythrocytes, platelets and normal white blood cells. This results in a lower immune response and a higher susceptibility to illness.

Leukaemia can be categorised into 4 main groups depending on what cell type is affected, myeloid or lymphoid, and the speed of development of the disease, acute or chronic (Foucar 2010). Chronic leukaemia's do not necessarily show symptoms during the early stages of the disease and can take years to progress. They can sometimes only be discovered during routine blood testing and initial treatment may include monitoring the disease, tyrosine kinase inhibitors and, in some cases at a later stage, stem cell transplants. Acute leukaemia's are aggressive cancers that progress rapidly and require immediate treatment. Acute Lymphoblastic Leukaemia (ALL) is a rare blood cancer affecting lymphocytes in the blood. Abnormal lymphocytes can build up in the spleen and lymph nodes enlarging them causing swollen lymph nodes and abdominal pain, leading to weight loss and frequent infections. If untreated ALL can lead to death within weeks or months. Acute Myeloid Leukaemia (AML) is caused by the rapid development of cancer in the immature myeloid cells (monocytes, macrophages, neutrophils, basophils, eosinophils, erythrocytes, dendritic cells and platelets) which are responsible for fighting bacterial infections, stemming bleeding and oxygen transport. AML has a very rapid onset of disease and can lead to fatalities within weeks or months.

Symptoms of AML include fatigue, weight loss, shortness of breath, easy bruising and frequent infections. To diagnose AML, a blood test is taken to determine the presence of abnormal blast cells in the blood, accompanied by a reduced white blood cell count (Figure 2). A bone marrow sample is also taken from the back of the pelvic bone and can include both a bone marrow aspirate and bone marrow biopsy. A blast cell count

of 20% or more in the blood or bone marrow will confirm a diagnosis of acute myeloid leukaemia (Dohner et al. 2017). A haematologist examines both blood and bone marrow samples to determine the type of leukaemia. Further tests such as immunohistochemistry, cytogenetic analysis or fluorescent in situ hybridization (FISH) can also be performed to classify the type of AML.
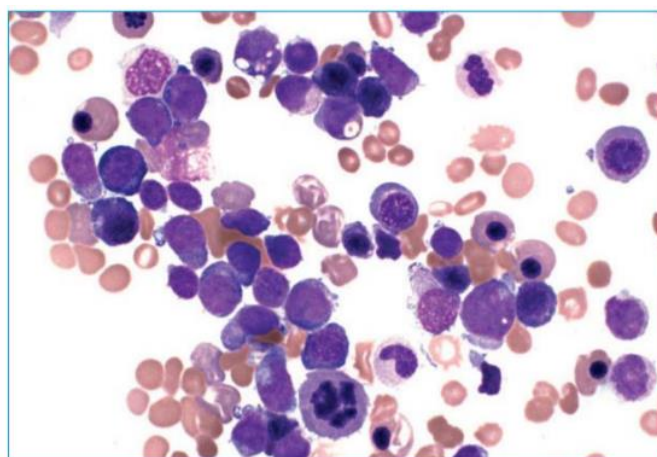
Figure 2. AML bone marrow aspirate smear

This smear is from a therapy-related acute myeloid leukaemia patient showing abundant myeloid blasts and abnormal erythroid cells. This image is taken from Foucar et al. 2008.

The overall incidence of AML in the U.K. is 3.4 cases per 100,000 with around 70% occurring in people over the age of 60 (British Committee for Standards in Haematology et al. 2006). AML is generally a cancer affecting the elderly with a poorer prognosis and survival with increasing age. Treatment options include chemotherapy, radiotherapy and allogeneic hematopoietic-cell transplantation, however, for older patients, supportive care or new investigational therapies can often be the only options due to frailty and ill health. The survival rate in patients below the age of 60 is 35-40% while for older patients over 60 years the survival rate drops to a low rate of 5-15% (Dohner, Weisdorf, and Bloomfield 2015).

AML can be classified using two systems, the French-American-British (FAB) system and the World Health Organisation (WHO) classification system. The FAB system which separates AML into subtypes M0 to M7 based on the cell type affected. Diagnosis includes morphologic assessment of the blasts for lineage determination, which in diagnosis of AML blasts include myeloblasts, monoblasts, promonocytes,

erythroblasts and megakaryoblasts. Flow cytometry immunophenotyping is also carried out to assess the immaturity of the blasts and cytogenetic analysis. The WHO classification system groups subtypes based on cytogenetic analysis and patient history (Foucar 2010). AML can be classified into 9 biologic sub-groups based on the chromosomal abnormalities present by analysis of a patients karyotype as well as two sub-groups based on clinical information such as previous blood disorders or radiotherapy and two sub-groups based on Down syndrome diagnosis. This classification requires cytogenetic analysis of a patient's chromosomes to look for aberrations. Leukaemia can develop due to an accumulation of chromosomal aberrations such as translocations, insertions and deletions (Figure 3). Translocations occur when sections of chromosomes are swapped with no loss of genetic material and remain fully functional (balanced) or they are swapped with loss of genetic material (unbalanced). Inversions also involve chromosomal breaks, this time the break occurs in one chromosome where the broken section is reversed and reattached to the chromosome incorrectly. Common translocations or inversions that occur in AML include t(8;21)(q21;q22) which affects the Runt-related transcription factor 1 (*RUNX1)* gene, also known as *AML1*, and the gene *RUNX1T1*, also known as *ETO* producing the fusion protein *AML1-ETO* and inv(16)/t(16;16)(p13.1;q22), a translocation or inversion affecting genes core-binding factor subunit beta (*CBFB)* and myosin heavy chain 11 (*MYH11)* and t(15;17)(q22;q21) which is a fusion of the two genes promyelocytic leukaemia (*PML*) and retinoic acid receptor alpha (*RARA*) (Mrozek et al. 1997). Cytogenetic analysis can categorise patients into those with favourable (t(8;21), inv(16), t(15;17)), intermediate (normal cytogenetics) or unfavourable cytogenetics (3q21q26 abnormalities, 5q-/-5, 7q-/-7, 11q23 abnormalities, 12p abnormalities, 17p abnormalities, complex aberrant abnormalities). Cytogenetic analysis provides the most accurate determination of diagnosis and prognosis for an AML patient (Mrozek et al. 1997).
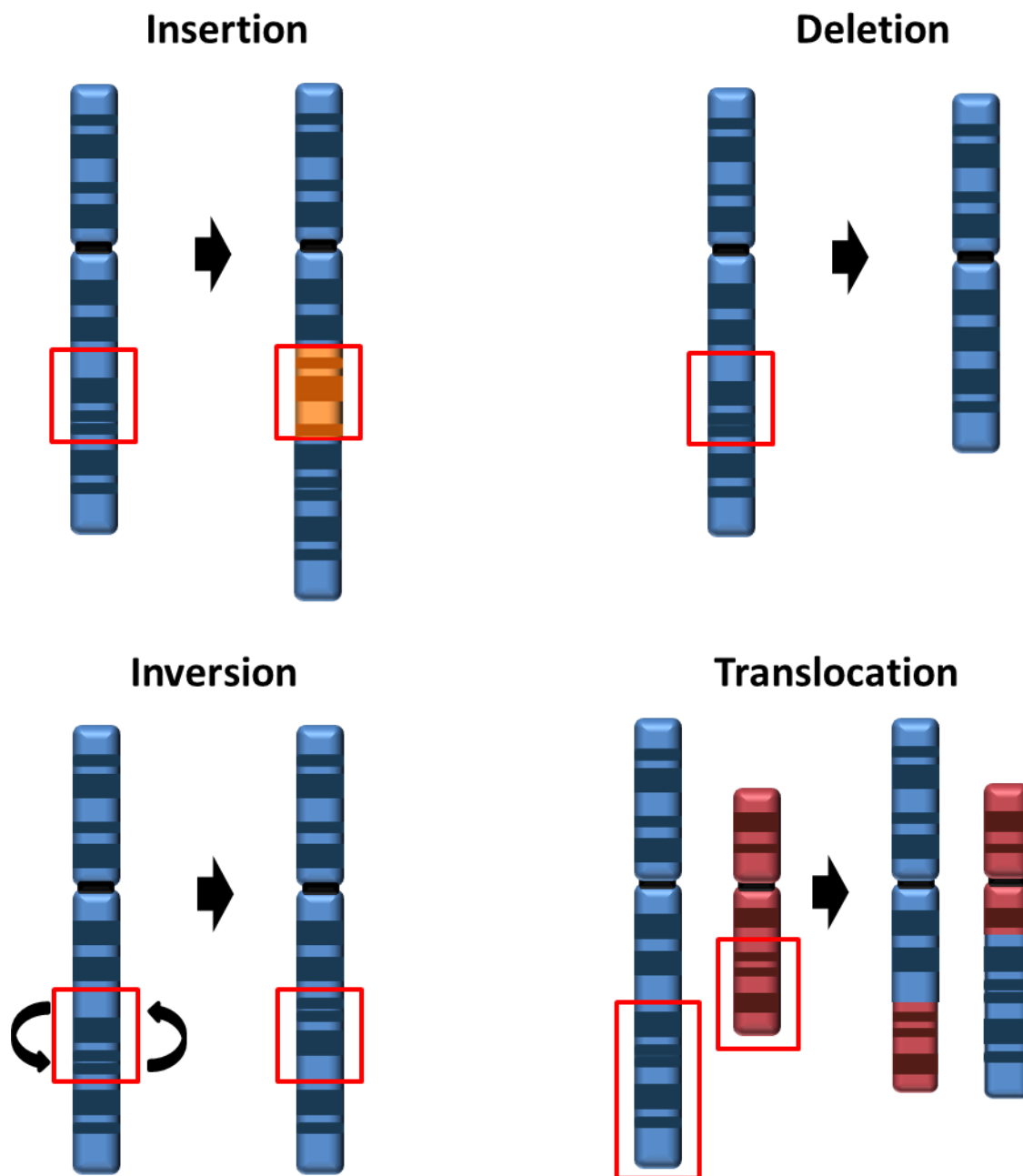
Figure 3. Chromosomal aberrations in AML.

Chromosomal aberrations such as insertions, deletions, inversions, where a part of the chromosome inverts and reattaches to the same chromosome, and translocations where 2 chromosomes exchange parts of each other's chromosomes, are common in AML cases.

1.3 Risk factors for AML development

There are a number of factors that increase a person's risk of developing AML such as inherited genetic disorders, environmental exposures and cancer therapies.

1.3.1 Genetic

People with the blood disorders known as myelodysplastic syndromes (MDS) or severe congenital neutropenia have a high risk of acquiring more genetic mutations which will progress the disease into AML. MDS is the term used to categorise a group of pre-leukaemic blood disorders that produce immature blood cells and is characterised by symptoms such as anaemia, neutropenia and thrombocytopenia. Many people with the disorder are asymptomatic with onset of the disease occurring gradually, often late in life and 40% of cases lead to the development of AML (Heaney and Golde 1999). Due to their similar characteristics, appearance, and presentation, AML and MDS are often combined together in leukaemic analysis. For people with severe congenital neutropenia who have had granulocyte colony stimulating factor (G-CSF) therapy for 15 years, the incidence of development of MDS and AML is 34% (Foucar 2010). Other disorders such as Shwachman-Diamond syndrome, Dyskeratosis congenital and Fanconi anaemia also report an elevated incidence of development of MDS/AML (Foucar 2010). It has been reported that there is a higher incidence of AML, up to 10 to 18 times greater, among the Down syndrome population (Evans and Steward 1972).

Another factor which increases the risk of AML development is the natural aging process. The incidence of AML increase with time with rates of 1.3 per 100,000 people aged less than 65 years and 12.2 per 100,000 people aged over 65 years old (De Kouchkovsky and Abdul-Hay 2016). The mean age of de novo AML patients is 65 years, which illustrates the time required to gradually acquire mutations which lead to the occurrence of the disease (Milligan et al. 2006). The frequency of mutations has been seen to increase consistently with increasing age in people without hematologic malignancies (Jaiswal et al. 2014). In a large scale study common AML mutations were present in 2% of individuals without leukaemia studied, rising to 5-7% in people over 70 years old, illustrating this background number of mutations present in normal individuals (Xie et al. 2014).

1.3.2 Environmental

Benzene, a known carcinogen, has been used in the petro-chemical industry and other manufacturing sectors such as shoes, rubber goods and paint for decades. Benzene levels are now strictly regulated, especially in gasoline production in western countries.

A large cohort consisting of over 73,000 workers from hundreds of benzene factories in China from 1972-1987 has shown a significantly elevated incidence of AML (Linet et al. 2015).

Ionising radiation induces DNA damage in cells that can result in cytogenetic aberrations such as translocations or somatic mutations that can lead to the development of cancer. After exposure to ionising radiation, AML is one of the most common malignancies to occur in humans (Hsu et al. 2013; Krestinina et al. 2013). This is particularly evident in the atomic bomb survivors of Hiroshima and Nagasaki which compromises one of the largest human radiation exposure cohort of over 93,000 survivors. Physicians of the atomic bomb survivors during the 1940s noticed an increased rate of leukaemia in survivors who were near the hypocentre. Since then many reports and studies have been published on this cohort with many factors such as cancer incidence, mortality, sex, and age at exposure examined (Ron et al. 1994; Pierce et al. 1996; Bizzozero, Johnson, and Ciocco 1966; Preston et al. 1994). 176 cases of AML have been identified from these Life Span Study (LSS) cohort studies, with radiation exposure accounting for 38% of these cases (Hsu et al. 2013). Other populations of exposed individuals such as the Techa River Cohort, where village residents near the Mayak plutonium production complex in Russia were exposed to multiple radionucleotides dumped into the Techa river, have long been studied as a source of protracted low dose rate exposure samples. Similar to the atomic bomb survivors, nearly 50% of the leukaemia samples studied have been estimated to be due to this radiation exposure (Krestinina et al. 2013). Radiation workers, although a much smaller cohort, also serve as a valuable source of radiation exposed individuals for analysis. Accidental exposure of IR among radiation workers before 1950 resulted in a higher incidence of leukaemia due to the high level of radiation exposure (Yoshinaga et al. 2004). The development of leukaemia has also been linked to exposure to sources of lower doses of radiation through the use of computed tomography (CT) scans in childhood in recent studies in the U.K. and Australia. Paediatric exposures to low doses of 50 mGy of IR from CT scans are also thought to triple the risk of leukaemia with an excess relative risk per mGy of 0.036 (Pearce et al. 2012) and 0.039 (Mathews et al. 2013) reported. Increases in leukaemia have been reported in patients after x-ray treatment for other conditions such as ankylosing spondylitis (Weiss, Darby, and Doll 1994) and metropathia haemorrhagica (Darby et al. 1994).

1.3.3 Medical

Secondary AML includes the development of AML from MDS or therapeutic treatment for a primary malignancy, known as therapy-related AML (t-AML). T-AML describes the development of leukaemia after a wide range of cytotoxic treatments such as chemotherapy drugs, radiotherapy or a combination of all. T-AML cases can be separated into two groups depending on what treatment the patient received. The most common subtype occurs after exposure to alkylating agents and/or radiation and is characterised by unbalanced chromosomal aberrations affecting chromosomes 5 and/or 7 with a latency period of 5-10 years (Pedersen-Bjergaard et al. 1990). The second group is caused by treatment with topoisomerase II targeting agents. It is characterised by translocations involving chromosome bands 11q23 or 21q22 with a shorter latency period of 1-5 years (Greaves 1997). Separation of patients into these categories, however, is not easy since most patients receive a combination of these treatments. Of a reported 372 AML cases, 13% developed due to chemotherapy and/or radiation treatment for another disease and with a larger study group of 4230 AML cases, 14% were t-AML (Mauritzson et al. 2002). Exposure to chemotherapy and radiotherapy treatment for other malignancies such as Hodgkin's Lymphoma, non-Hodgkin's Lymphoma, lung cancer, breast cancer and prostate cancer has the highest incidence of the development of secondary AML (Schoch et al. 2004; Pedersen-Bjergaard et al. 1990; Mauritzson et al. 2002). In a study by Mauritzson et al., nearly a quarter of t-AML/t-MDS cases developed after treatment for non-malignant diseases such as arthritis (Mauritzson et al. 2002). In comparison to de novo AML, t-AML is a disease with a faster onset, an often resistance to chemotherapy, higher relapse rates and lower complete remission (CR) rates. T-AML usually presents within 3 to 5 years after receiving treatment for cancer with an increased risk caused by exposure to radiotherapy and chemotherapy and an older age at treatment (Preston et al. 1994). It is considered a lethal disease with a poor outcome, particularly for those with unfavourable cytogenetics. This is a severe disease in the elderly, an already fragile group who may have other concurrent malignancies, previous high dose treatments which damaged vital organs and in general have a compromised immune system.

T-AML is commonly reported after a combination of chemotherapy and radiotherapy due to the recommended treatment strategy, nonetheless, cases of t-AML following radiotherapy alone are also prevalent. In a case-control study of 150,000 women with

cervical cancer, a two-fold increase in the risk of leukaemia development was determined in cervical cancer patients after radiotherapy (Boice et al. 1987). Breast cancer patients who received radiotherapy alongside lumpectomy had a 2.38 (P=0.006) relative risk (RR) which was relative to mastectomy patients who did not receive radiation therapy (Smith, Bryant, et al. 2003). Another study determined breast cancer patients had a 3.9 RR of developing AML after receiving radiotherapy treatment (Le Deley et al. 2007). For Hodgkin's lymphoma, however, some studies have reported radiotherapy as having no or little effect on t-AML development (Swerdlow et al. 1992; Henry-Amar and Dietrich 1993). A large meta-analysis on 1,183 patients with Hodgkin's lymphoma from 12 trials concluded that combined chemotherapy and radiotherapy does have a significantly adverse effect on mortality in comparison to chemotherapy alone (Loeffler et al. 1998).

T-AML patients can also be categorised based on favourable, intermediate or unfavourable cytogenetics which, similar for de novo AML, gives strong indications of prognosis and overall survival (Kern et al. 2002; Armand et al. 2007). Normal cytogenetics are seen at a higher rate in de novo AML cases (Schoch et al. 2004; Mauritzson et al. 2002; Kayser et al. 2011). The spectrum of chromosomal abnormalities is similar among de novo AML and t-AML, it is the frequency of these abnormalities that differs. The incidence of karyotype aberrations is present at a much higher rate of 86% in t-AML in comparison to 57.6% in de novo AML (Schoch et al. 2004). This was again seen in a study by Kayser et al. in 2011 where abnormal karyotype was seen in 75% of t-AML cases in comparison to 51% in de novo AML cases and in a study by Mauritzson et al. in 2002 where chromosome abnormalities were present in 68% of t-AML patients in comparison to 50% of de novo AML patients (Kayser et al. 2011; Mauritzson et al. 2002). The favourable risk category seems to occur at similar (Kayser et al. 2011) or slightly higher (Mauritzson et al. 2002) rates in t-AML in comparison to de novo AML cases. This increase in abnormal karyotypes in t-AML is mostly seen in the unfavourable cytogenetics category with cases much higher at 46.2% in t-AML cases in comparison to 20.4% for de novo AML cases (Schoch et al. 2004). Higher unfavourable aberrations rates could be seen in a large study involving 306 t-AML patients with abnormal karyotypes affecting chromosomes 5 or 7 present in 76% of cases (Smith, Le Beau, et al. 2003). Also, in other radiation-induced cancers such as radiation-induced osteosarcoma, spindle cell sarcoma, angiosarcoma and breast cancer, higher rates of balanced inversions and small

deletions have been detected in comparison to naïve cancers (Behjati et al. 2016). Cytogenetic analysis is currently the best method at predicting prognosis. However, when comparing de novo AML with t-AML survival in all risk categories, the outcome was poorer in all t-AML cases (Pagano et al. 2005; Schoch et al. 2004). Therefore, cytogenetics does not cover all aspects of prognosis and other techniques, such as genetic or epigenetic analysis, may provide more information on risk, prognosis and overall survival.

The prognosis for therapy-related AML, with high rates of unfavourable cytogenetics, is poor. T-AML cases in general have a worse response to treatment than de novo cases (Ballen and Antin 1993). The most successful treatment for t-AML cases is allogeneic hematopoietic cell transplantation (HCT), which has reported success rates of survival of 20-30%, with a 3 year overall survival of 35% (Kroger et al. 2006) and a 5 year disease free survival of 24.4% (Anderson et al. 1997). However very few patients are eligible for HCT due to their poor condition after initial treatment. Complete remission rates are reportedly similar between de novo AML and t-AML cases (Pagano et al. 2005; Kayser et al. 2011). The overall survival rate of t-AML, however, is much lower at 10 months in comparison to 15 months for de novo AML (Schoch et al. 2004) with a lower median survival rate of 4 months for t-AML after diagnosis also reported (Mauritzson et al. 2002). Relapse is around 3 times higher in t-AML cases in comparison to de novo AML cases (Schoch et al. 2004) and relapse has been identified as the main cause of treatment failure (Dohner, Weisdorf, and Bloomfield 2015). In a large study involving 306 t-MDS and t-AML patients, 5 year overall survival is calculated at just 10% (Smith, Le Beau, et al. 2003).

At the moment, the outlook for t-AML patients is poor with few treatment options when viable, poor complete remission, high relapse rates and low overall survival rates. For many, the treatment plan is supportive care. Also, there is a growing incidence of t-AML due to aging population, increased detection and treatment of cancers and an increased use of chemotherapy and/or radiotherapy during treatment (Leone et al. 1999). Therefore, investigation into the development of t-AML, the genetic risks associated with the likely development of t-AML and successful treatment options are vital for the successful treatment of these patients.

1.4   Leukaemogenesis

Leukaemogenesis has long been considered a multistep process, requiring the sequential acquisition of mutations for the development of leukaemia. It has been shown that co-operating mutations are required in order for leukaemogenesis to occur, described as a two-hit theory proposed by Kelly and Gilliland (Kelly and Gilliland 2002). Both stages are necessary; an arrest in myeloid differentiation resulting in immature cells and an increase in cell proliferation, resulting in an infiltration of immature cells into the bone marrow and blood. Without mutations affecting both stages, myeloid development may be affected, or late onset-leukaemia may occur. Fusion proteins AML1-ETO and CBFB-MYH11, resulting from t(8;21) and inv(16)t(16;16), have been seen in murine models to lead to myeloid impairment but not development of leukaemia (Castilla et al. 1999; Yuan et al. 2001). In both cases, exposure to a DNA alkylating mutagen was required to induce leukaemogenesis, again showing the requirement for at least two mutational events. Other mouse studies have illustrated this theory of co-operating mutations with mouse models. A mouse model with a nucleophosmin 1 (*Npm1*) mutation and internal tandem duplication (ITD) in the gene Fms Related Tyrosine Kinase 3 (*Flt3*) resulted in the development of leukaemia (Mupo et al. 2013), while these mouse models on their own result in delayed-onset AML (Vassiliou et al. 2011) and late-onset disease similar to chronic myelomonocytic leukaemia (Lee et al. 2007). The long latency period in mouse models with one mutation indicate the requirement for at least a second mutation for the efficient development of AML.

Cytogenetic analysis by karyotype and recent advances in techniques such as whole genome sequencing has allowed further analysis into AML development and architecture. Recent work by the Cancer Genome Atlas reported an average of 5 recurrent genetic mutations per patient (Cancer Genome Atlas Research et al. 2013), illustrating a clonal development of AML. Clonal hematopoiesis of indeterminate potential (CHIP), is the development of a subpopulation of blood cells possessing a genetic mutation. Clonal development of preleukaemic cells leads to multiple clones within an individual with the dominant clone causing the cancer. In a large blood sampling study involving 12,380 participants, high numbers of mutations in the genes DNA (cytosine-5)-methyltransferase 3A (*DNMT3A),* Tet Methylcytosine Dioxygenase 2 (*TET2)* and ASXL Transcriptional Regulator 1 (*ASXL1)* were detected in participants,

suggesting them to be initiating mutations, with somatic mutations increasing with age (Genovese et al. 2014). Also, the detection of clonal haematopoiesis in these patients strongly linked with a future blood cancer diagnosis. Cytogenetic subclones have been detected in 15.8% of AML patients, most common in the unfavourable cytogenetics group, and their development is a more branched, rather than a linear, acquisition of mutations (Bochtler et al. 2013). Generally cytogenetic subclones were significantly more evident in older patients with a median age of 58 in comparison to AML patients with no subclones with a median age of 52 (Bochtler et al. 2013). Normal healthy human hematopoietic cells acquire mutations with increasing age with the number of mutations in healthy individuals similar to the number of mutations in AML patients of the same age, therefore, although AML patients can contain mutations, only some initiate leukaemogenesis (Welch et al. 2012). Mutations in one gene in particular, tumour protein P53 (*TP53)*, have been identified in 9 out of 19 healthy elderly cancer-free donors (Wong et al. 2015). This is also seen in genetic mosaicism, defined by an occurrence of two or more distinct karyotypes in an individual, where the frequency of karyotype mutations is seen to increase with increasing age in cancer-free individuals (Jacobs et al. 2012).

As the body ages, the hematopoietic system progressively loses its function with increased self-renewal and decreased differentiation which can lead to several diseases, such as leukaemia, being more prevalent in the older population. This develops through functional, transcriptional and epigenetic changes that occur in stem cells such as a decrease in expression of epigenetic regulators (Sun et al. 2014). Karyotype evolution from diagnosis to relapse is more commonly seen in patients with unfavourable aberrations which have been suggested to be due to genomic instability (Kern et al. 2002). Cytogenetic subclones have inferior complete remission rates of 30% in comparison to the 40% in cases without subclones (Bochtler et al. 2013). The analysis of samples at diagnosis and relapse has illustrated how multiple clones can co-exist at diagnosis with cytotoxic treatment often unable to eliminate all clones. This can lead to relapse with the reoccurrence developing from sub clones at diagnosis resistant to treatment. Genomic profiling of AML patient samples at diagnosis and relapse determined the relapse to be due to progression of a founder clone rather than the emergence of a novel clone (Parkin et al. 2013). DNA sequencing analysis of primary and relapsed AML patient samples in a study by Ding et al. has indicated two major pathways of clonal development in AML. Either the founding clone survives

chemotherapy and gains mutations and evolves into the relapsed clone or a subclone survives chemotherapy and gains more mutations to develop into the relapsed clone (Ding et al. 2012). Mutations in *DNMT3A* have been detected in non-leukaemic T cells as well as hematopoietic stem cells (HSCs) from AML patients at the time of diagnosis and relapse (Shlush et al. 2014). These mutations are thought to be present in pre-leukaemic cells, that survived chemotherapy, can persistent during remission and are capable of expanding with acquired mutations to cause relapse. It is also possible, however, that cancer therapies might exert selective pressures on hematopoietic stem cells (HSCs) such that certain mutant populations have a selective advantage under cytotoxic conditions. If the mutant clones survive longer, they may accumulate more mutations with time. This could be why the expansion of mutant clones, is associated with an increased risk of developing hematologic malignancies (Genovese et al. 2014; Jaiswal et al. 2014; Xie et al. 2014). This highlights the weakness of current cytotoxic treatment therapies and the need to further investigate the contribution of cytogenetic and genetic mutations to t-AML development in order to improve treatment success and reduce relapse rates.

1.5 Genetic mutations in AML

The screening of AML patients for genetic mutations developed due to the number of patients with no chromosomal aberrations, known as cytogenetically normal AML (CN-AML). Cytogenetic analysis allows for prediction of outcome and survival by categorising them into favourable, intermediate or adverse risk groups, with all cytogenetically normal patients being categorised as intermediate risk, and it also guides physicians in their choice of treatment. However, about 42-48% of AML patients have a cytogenetically normal karyotype (Byrd et al. 2002; Grimwade et al. 1998; Mrozek, Heerema, and Bloomfield 2004; Schlenk et al. 2008). This has led to the inclusion of genetic mutations of three genes, *NPM1*, CCAAT/enhancer binding protein alpha (*CEBPA*) and *FLT3*, into the WHO scheme in 2008 (Khasawneh and Abdel-Wahab 2014). They are particularly useful in providing prognostic information for cytogenetically normal cases, which are classified as intermediate-risk AML patients, so that they can be classified further into favourable, intermediate or adverse risk patients (Dohner et al. 2010). In order to provide prognostic information, genetic mutations were screened for and hundreds of mutations have now been identified.

Only 5 genes, *FLT3*, *NPM1*, *CEBPA*, *DNMT3A* and *KIT*, were mutated in more than 10% of an AML patient group of 197 people (Kihara et al. 2014). This reinforces the idea that leukaemia develops from several co-operating mutations in a limited number of genes.

Commonly mutated genes such as *FLT3*, *NPM1* and *DNMT3A* have been identified as well as a large number of genes that are less frequently mutated (Figure 4) (Patel et al. 2012; Cancer Genome Atlas Research et al. 2013). Genetic mutations can be grouped into class I mutations which result in cell proliferation and survival (*FLT3*, *KIT*, Kirsten Rat Sarcoma Viral Oncogene Homolog *(KRAS)*, Neuroblastoma RAS Viral (V-Ras) Oncogene Homolog *(NRAS)*, *TP53*) and class II mutations which block myeloid differentiation (*NPM1*, *CEBPA*, *RUNX1*).
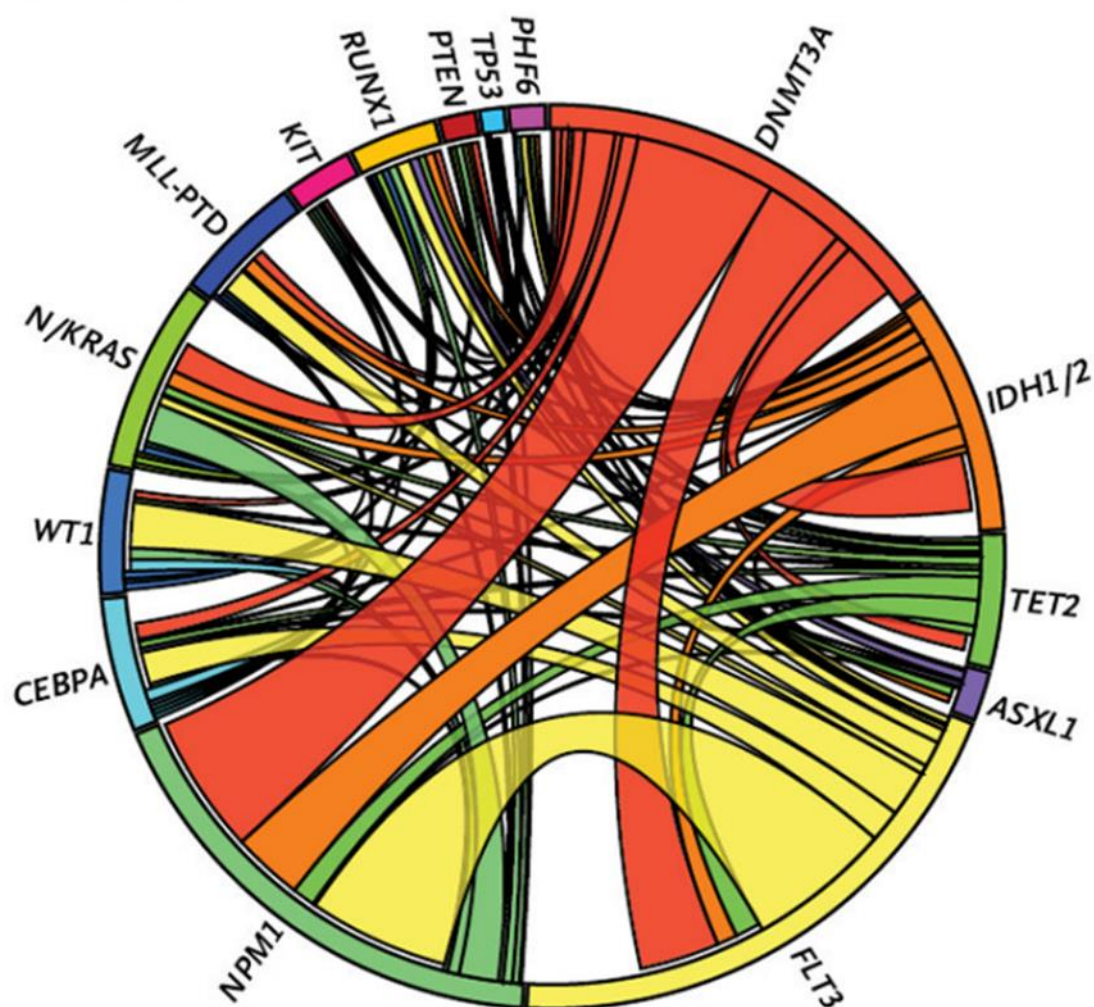
Figure 4. Mutational spectrum and heterogeneity of commonly mutated AML genes.

In the Circos diagram, the width of the ribbons indicates the percentage of patients with the mutation. Pairwise mutations are shown linking two mutated genes. Taken from (Patel et al. 2012).

Recent work has scanned the genome trying to identify genomic changes occurring in AML patients. A large study of 200 de novo AML patients by Ley and colleagues found 23 genes which were frequently mutated, each having different functions such as epigenetic regulators (*DNMT3A*, Isocitrate Dehydrogenase (NADP(+)) 1 and 2 (*IDH1*, *IDH2*) *TET2*, Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit (*EZH2)*), signalling proteins (*FLT3*, *KIT*, *NRAS*, *KRAS*) or tumour suppressors (*TP53*, Wilms Tumour 1 (*WT1)*, PHD Finger Protein 6 (*PHF6)*) (Cancer Genome Atlas Research et al. 2013). Some class I and class II mutations have been found to occur together while others occur exclusive of one another. *FLT3-ITD*, *DNMT3A* and *NPM1* mutations often occur together in 6% of patients (Papaemmanuil et al. 2016). This can be further broken down with *FLT3-ITD* mutations commonly occur alongside *NPM1* exon 12 mutations, and *NPM1* mutations also occur alongside *DNMT3A* mutations (Cancer Genome Atlas Research et al. 2013). Also, the fusion genes PML-RARA and MYH11-CBFB were present in samples that lacked *NPM1* and *DNMT3A* mutations, while patients with *RUNX1* and *TP53* mutations did not have *FLT3* and *NPM1* mutations. In a larger study involving 1540 patients, an association could be seen for *NPM1* mutations to associate with mutations affecting codon 12/13 of the gene *NRAS* rather that codon 61 (Papaemmanuil et al. 2016). Also, a mutation affecting the codon 140 of gene *IDH2* appeared to co-occur with *NPM1* mutations, while a mutation affecting the codon 172 did not (Papaemmanuil et al. 2016). By comparing mutations in patients where the initiating event (PML-RARA) is known with mutations where the initiating event is not known, Welch and colleagues were able to identify potential initiating (*NPM1*, *DNMT3A*, *IDH1*) and co-operating mutations (Welch et al. 2012). Analysis of pre-leukaemic patient samples obtained years before the development of AML has allowed for the identification of early mutations. *DNMT3A* mutations being present in pre-leukaemic stem cells demonstrate how they are early mutations that arise earlier than other mutations not found in pre-leukaemic samples such as *NPM1* and *FLT3* (Shlush et al. 2014). *TP53* mutations have also been identified at low frequencies in bone marrow samples 3-6 years before the development of t-AML (Wong et al. 2015).

The use of genetic information for patient prognosis has identified mutations which are associated with good survival rates and can provide information for physicians for deciding the most suitable treatment type. These mutations provide valuable prognosis information for complete remission and relapse, depending on where the mutation is located, and which genes are also mutated. *RUNX1* mutations have been previously reported to occur alongside *FLT3-ITD* or *FLT3-TKD*. Mutations in the gene TP53 have been associated with an unfavourable risk cytogenetic profile (Cancer Genome Atlas Research et al. 2013) and these mutations occur in 70% of cases with complex karyotype (Rucker et al. 2012). *TP53* mutations have been associated with very poor prognosis and have been identified as an unfavourable factor for achieving complete remission (Kihara et al. 2014), while mutation in the gene *NPM1* has been associated with a favourable factor for complete remission (Kao et al. 2014; Kihara et al. 2014). Some of these mutations may be included in clinical guidelines such as WHO or European LeukemiaNet (ELN) recommendations. Others will require more thorough analysis involving larger studies to investigate the implication of the mutation. It has also been recently suggested to include an AML classification group to include mutations specifically affecting epigenetic regulator genes such as *ASXL1*, *EZH2*, *TET2*, *DNMT3A*, *IDH1/2*. In 2013, the Cancer Genome Atlas project revitalised these classes to a total of 9 classes including epigenetic mutations. Mutations in genes that regulate epigenetic control, such as *TET2* and *DNMT3A*, have been reported in 40-44% of AML patients (Kao et al. 2014; Cancer Genome Atlas Research et al. 2013). *TET2* mutations were present in 13.2% of 486 de novo AML patients and if present in patients with unfavourable genotypes, conferred a worse overall survival (Chou et al. 2011). In another large study involving 357 AML patients, *TET2* mutations again showed a correlation with poor overall survival (Aslanyan et al. 2014). *DNMT3A* mutations, mainly affecting the codon R882, are found to be associated with an inter-mediate risk cytogenetic profile and also a poor survival in comparison to patients without *DNMT3A* mutations (12.3 months vs 41.1 months) (Cancer Genome Atlas Research et al. 2013). Many pre-leukaemic mutations have been reported in epigenetic regulatory genes such as *IDH1*, *IDH2* and *DMT3A*, illustrating the importance of epigenetic changes during leukaemic development (Corces-Zimmerman et al. 2014).

1.5.1 SNPs

It is not just somatic mutations which are associated with AML development but also single nucleotide polymorphisms (SNPs). SNPs are genetic variations in a single nucleotide of DNA present at a frequency of > 1% of a population. SNPs can occur across the genome on coding, or non-coding sites and depending on the change, can result in a change of protein or not. SNPs are associated with a number of diseases and large GWA studies have identified SNPs which are associated with an increased risk of diseases such as diabetes and coronary artery disease (Wellcome Trust Case Control 2007).

The enzyme NAD(P)H:quinone oxidoreductase (NQO1) functions to break down chemotherapeutic drugs and carcinogens such as some of those in cigarette smoke and benzene. When benzene is metabolized, the NQO1 enzyme catalyzes the reduction of benzoquinones to hydroxyl metabolites which are less toxic. A polymorphism in this gene results in reduced expression in heterozygous patients and complete loss of function in homozygous patients which has been shown to lead to hematotoxicity in benzene-poisoned workers in China (Rothman et al. 1997). A higher prevalence of this polymorphism of 1.4-fold was seen in 56 t-AML patients with 11% being homozygous and 41% being heterozygous for the C $\rightarrow$ T polymorphism (Larson et al. 1999). This polymorphism was again seen 24% of a Japanese cohort of 58 t-AML patients in comparison to 15% of de novo AML (Naoe et al. 2000). Polymorphisms in other detoxification genes that metabolise toxic compounds have also been investigated. A polymorphism in the cytochrome P450 superfamily B member, *CYP2B6*, which functions to metabolise toxic compounds and protect cells against oxidative damage has also been investigated. The $G^{516}T$ SNP was present at a higher rate in AML patients in comparison to controls, with homozygous polymorphisms were present at a much higher rate of 20% of secondary AML patients (Daraki et al. 2014). These results suggest that the presence of these polymorphisms genetically increase the risk of developing t-AML.

1.6 Epigenetic changes in AML

Epigenetic research has expanded over the past decade with gene expression and DNA methylation changes in cancer development being widely investigated (Wajed, Laird, and DeMeester 2001). Epigenetic research has come to the forefront due to

problematic issues for AML patients such as cytogenetically normal AML cases, high relapse rates of AML, resistance of secondary AML cases to chemotherapy and dismal prognosis of secondary AML cases. There are AML cases where no mutation in the well-known targeted genes has been detected (Patel et al. 2012; Shen et al. 2011). In a large study of 485 adults with cytogenetically normal AML, 14.8% of patients were found to have no mutations in the frequently mutated genes *NPM1*, *CEBPA*, *FLT3*, *NRAS* and *WT1* genes (Dohner et al. 2010). There are therefore other factors, whether genetic or epigenetic, that result in the development of AML.

Transcriptional mRNA and miRNA changes in leukemic cells have recently been investigated. Expression of 44 core genes in leukaemia stem cells have been associated with shorter survival in cytogenetically normal AML patients (Eppert et al. 2011). Further work on these samples has shown that this gene expression signature is associated with mutations in genes *FLT3*, *WT1* and *RUNX1* with high *miR-155* expression (Metzeler et al. 2013). Overexpression of the gene SET, Nuclear Proto-Oncogene (*SET*), was found in 60 out of 214 de novo AML patients which associated with poor overall survival and was particularly evident in patients with a normal karyotype (Cristobal et al. 2012). This over-expression of SET was found to increase proliferation of leukaemic cells by inhibiting the tumour suppressor protein phosphatase 2A (PP2A).

DNA methylation is an epigenetic mechanism which is common throughout development in order to control the transcriptional activity of gene expression, as needed. DNA methylation is essential for normal development and involved in numerous processes such as gene repression, imprinting and X-chromosome inactivation (Bogdanovic and Veenstra 2009). DNA methylation involves the transfer of a methyl group to the 5th carbon atom of the nucleotide cytosine to form 5-methylcytosine (Figure 5). This tends to happen to areas of DNA that contain stretches of cytosine and guanine nucleotides, known as CpG islands, where the modified cytosine lies next to a guanosine base and so there is also a cytosine diagonally across from the modified cytosine. The nucleotides cytosine and adenine can be methylated, however, methylation of cytosine is the most common. DNA methylation can occur across the whole genome, not just in the promoter region. DNA methylation is also found in exons, thought to regulate alternative splicing (Maunakea et al. 2013; Shayevitch et al. 2018), however it seems to be centered around CpG islands, which can be commonly found next to transcription start sites.
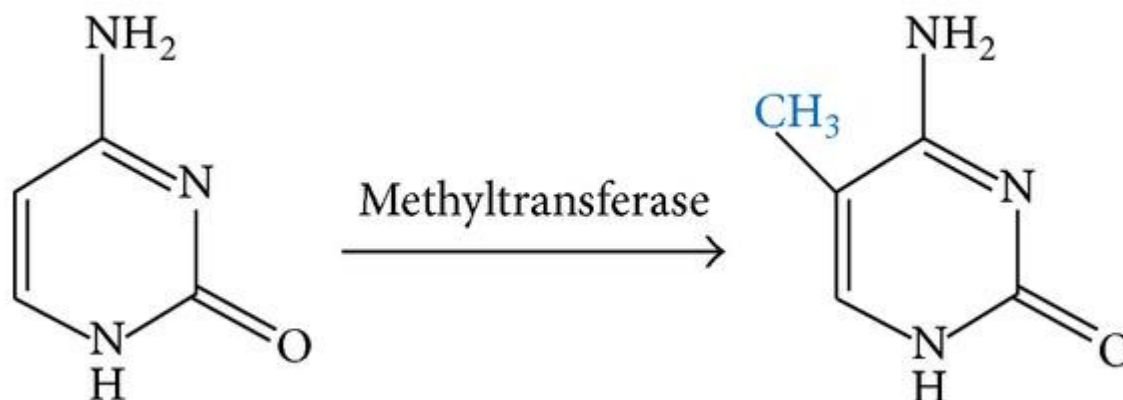
Figure 5. Methylation of cytosine in carbon 5.

This figure is copied from (Ghavifekr Fakhr et al. 2013).

DNA methylation patterns are regulated by several groups of proteins such as DNA methylation writers, readers and editors. DNA methylation first occurs due to members of the DNA methyltransferase family, DNMT1, DNMT3A and DNMT3B. DNMT3A and DNMT3B are responsible for de novo methylation, so they are able to methylate unmethylated CpG regions, while DNMT1 functions to maintain the methylation signature during replication (Bogdanovic and Veenstra 2009). DNMT3A and DNMT3B are essential in embryonic development with murine studies reporting knock outs of both genes to be embryonic lethal (Okano et al. 1999). Aberrant DNA methylated has been thought to contribute to cancer development for many years and in AML patients, mutations in genes involved in DNA methylation have been identified with *DNMT3A* mutations found to be present in 22% of AML patients (Ley et al. 2010).

DNA methylation patterns are then modulated by DNA methylation binding proteins which include methyl-CpG-binding domain (MBD) proteins, the Kaiso family proteins and the SET- and Ring finger-associated (SRA) domain family. These family members contain a MBD domain which binds to methylated CpG dinucleotides and most of the proteins also contain a transcriptional repression domain (TRD). The MBD proteins guide proteins with histone modifying activity or chromatin remodelling activity to specific sites. MeCP2 and MBD1 both interact with the chromatin remodelling binding partner HP1 (Du et al. 2015). MBD proteins are expressed at different levels during development. MBD2a and MBD2t bind to and regulate expression of the pluripotency genes OCT4 and NANOG (Lu et al. 2014). Their expression levels change during

development with overexpression of MBD2a causing differentiation by silencing OCT4 and NANOG expression while MBD2t overexpression maintains pluripotency. Many of these proteins have dual roles in histone and DNA methylation. MBD1 is involved in directing histone methylation to sites of DNA methylation, while MBD3 regulates histone deacetylation and de novo DNA methylation by interacting with the NuRD/Mi-2 complex and recruiting DNMT1 and DNMT3B (Morey et al. 2008). Finally, DNA methylation editors, the TET family members are responsible for the removal of the DNA methylation mark by converting 5-mC to 5-hmC, which inhibits binding of MBD proteins (Pastor, Aravind, and Rao 2013).

DNA hypermethylation of gene promoters, results in gene silencing while hypomethylation, a general decrease in methylation level, is very common in cancers. Abnormal DNA methylation is found in many diseases, including leukaemia (Claus and Lubbert 2003). Both hypermethylation and hypomethylation are evident in leukaemic cases, with hypermethylation of tumour suppressor genes leading to their silencing evident in cancers. Reprogramming of leukemic AML samples to induced pluripotent stem cells (iPSCs) resets epigenetic patterns but still results in the reacquisition of epigenetic patterns upon differentiation, illustrating their persistence in leukemic development (Chao et al. 2017). Abnormal methylation changes have been evident in early stages of AML, defined by blast percentages in the bone marrow, which can increase dramatically as the disease progresses (Sonnet et al. 2014). Mutations in epigenetic genes such as *DNMT3A*, *TET2*, *IDH1* and *IDH2*, as already discussed, have been identified as an important class of mutation in AML development. Mutations in *DNMT3A* have been found to be present in AML patients with significantly reduced gene methylation, of Homeobox (*Hox)* genes in particular, in comparison to healthy donor samples (Hajkova et al. 2012). Also, 4 weeks after exposure to ^{56}Fe ions, a decrease in expression in the *Dnmt* genes Dnmt1, Dnmt3a and Dnmt3a was reported in murine HSPCs with an increase in global HSPC methylation, suggesting hypermethylation of these genes (Miousse et al. 2014). Disruption of epigenetic regulation, therefore, appears to be important for leukaemogenic development.

Different areas of the genome are found to be methylated; promoters, gene bodies and CpG islands (CGIs). Analysis of methylation changes in these regions across AML subtypes has revealed that differentially methylated regions, mostly CGIs, are individual to each AML subtype (Saied et al. 2012). Promoter methylation levels of secondary AML cases showed an increase in comparison to de novo AML cases

which may be responsible for their different response to chemotherapy (Figueroa et al. 2009). Distinct subsets of AML revealed defined methylation patterns in comparison to normal bone marrow and DNA methylation patterns of 15 genes were also capable of predicting overall survival in AML patients (Figueroa et al. 2010). In a study by Agrawal et al. increased estrogen receptor α (ERα) and p15^{INK4B} methylation levels in bone marrow samples from AML patients in remission was found to correlate with a high relapse risk (Agrawal et al. 2007). DNA methylation therefore, like genetic mutations, has possibilities to be used for providing prognostic information for AML patients after treatment and survival and relapse risk estimation.

Another epigenetic mechanism which regulates transcription are histone modifications. Histones are proteins around which DNA winds, packaging the DNA into compact structures. There are 5 main families of histones: H1, H2A, H2B, H3 and H4. Histones H2A, H2B, H3 and H4 are known as the core histones which form the nucleosome. while histones H1 are known as linker histones which act to bind the nucleosome together. Histone modifications are an epigenetic mechanism which has important roles in DNA repair, replication and recombination (Bannister and Kouzarides 2011). Modifications to the histone proteins change the structure of the proteins and can therefore regulate access to DNA for gene transcription. Some modifications can disrupt histone-DNA interactions, loosening and unwinding the nuclesosome leaving a more open chromatin conformation for gene transcription. While other modifications strengthen the histone-DNA modifications tightening the structure and preventing gene activation. Many different types of modifications can occur such as methylation, acetylation, phosphorylation, deamination and ubiquitination.

Lysine acetylation occurs on the N-terminal histone tails that extends out from the nucleosome, removing lysine's positive charge and this modification then weakens its binding to the negatively charged DNA. This results in a dissociation of the histone from the DNA and a more open chromatin structure, allowing transcription factors access to the DNA. Histone acetylation and deacetylation are catalysed by histone acetyltransferases (HAT) which add acetyl groups or histone deacetylases (HDAC) which remove them. In therapy-related AML cases, it was found that the HAT CBP/P300 has an increased activity after the chromosomal translocation t(11;16)(q23;p13.3) (Sobulo et al. 1997) and directs cells towards a specific leukemic phenotype (Santillan et al. 2006).

Methylation modifications can also occur with the addition of a methyl group to the lysine or arginine residues of H3 and H4. Methylation differs from acetylation in that it does not change the charge or interaction with DNA. Lysines can be mono-, di- or tri-methylated while arginine can be mono- or di-methylated and the level of methylation appear to have different effects. A study by Barski et al. 2007 mapped 20 histone methylations of lysine and arginine residues by ChIP-seq analysis (Barski et al. 2007). They found that monomethylation of H3K27, H3K9, H4K20, H3K79 and H2BK5 are associated with gene activation while trimethylation of H3K27, H3K9 and H3K79 are associated with repression.

Murine studies have identified enzymes such as Moz, Mll1 and Ezh2 which perform essential roles in the modifications of histones in hematopoiesis (Butler and Dent 2013). Genetic mutations in genes involved in chromatin regulations have been identified in AML cases, such as *EZH2* and *ASXL1*, highlighting the interplay between cancer genetics and epigenetics (Cancer Genome Atlas Research et al. 2013). Recent research by Yi et al. has identified subtypes of AML through an integrated epigenetic analysis of human patient samples. One subtype with mutated *NPM1* were associated with an open chromatin confirmation, active enhancers with H3K4me1 and H3K27ac signals and hypomethylated CpG regions while samples with mutated *RUNX1* were associated with closed chromatin, promoters with H3K27me3 signals and DNA methylated regions (Yi et al. 2019). Histone modification, therefore, is another epigenetic mechanism disrupted in AML cases which can lead to cancer progression.

1.7 Radiation-induced AML in the mouse

Mouse models of human cancers are invaluable in giving us a better understanding of the molecular mechanisms of tumour initiation and development. They are useful for the study of human diseases due to the similarity of protein-coding genes in both species and the ability to alter transcriptional expression of a gene of interest using gene knock-in/knock-down techniques. There are also several inbred mouse strains of rAML available which develop leukaemia following radiation exposure, such as CH3/He, RF and SJL/J, with the inbred CBA/Ca strain being particularly well-characterised model of radiation induced AML (Verbiest et al. 2015). The CBA/Ca mouse is the mouse model of choice for radiation induced AML for several reasons. CBA/Ca mice have low background rates of AML (less than 1 in 1000) (Rithidech et

al. 1995) and there is a consistent induction rate of 15-20% following exposure to an optimal leukaemic dose of 3 Gy whole body exposure (Major and Mole 1978). Moreover, the histopathological features, such as splenomegaly, are very similar to those of human AML. Molecular findings such as a partial deletion of chromosome 2 (Clark et al. 1996) and a later accompanying point mutation in the gene Spi-1 Proto-Oncogene (*Sfpi1)* (Cook et al. 2004) have been established as characteristic features in the CBA/Ca mouse after irradiation. In contrast, the C57/BL6 mouse model is considered a model resistant to the development of AML after radiation exposure. Chromosome 2 deletions have been shown to be at a lower frequency in F1 hybrid CBA X C57 mice in comparison to CBA mice (Clark et al. 1996). Analysis of bone marrow cells from CBA and C57 mice after ^{56}Fe ions or γ-rays show C57 mice to have fewer cells with *Sfpi1* loss and the numbers of C57 bone marrow cells with *Sfpi1* loss returned to control levels around one month after exposure, while CBA bone marrow cells with *Sfpi1* loss did not (Peng et al. 2009). There is also a difference in the number of *Sfpi1* point mutations with R235 point mutations reported in 87% of the radiation susceptible cross CBA x SJL mice (Cook et al. 2004), however this drops to 67% in CBA x C57 mice (Suraweera et al. 2005). Therefore, the genetic background does appear to have an influence on the development of AML and provides rAML sensitive and rAML resistant mouse strains of interest in leukaemogenesis studies.

1.7.1 *Sfpi1*

The gene *Sfpi1*, known as *PU.1* in humans, is a member of the E26 transformation-specific (ETS) family that encodes a transcription factor PU.1 which is vital for development of the immune system. It contains a transactivation domain which interacts with transcription factors, a PEST domain which regulates protein-protein interactions, and a DNA binding domain that regulates its binding to DNA and other proteins such as GATA Binding Protein 1 (GATA1), GATA Binding Protein 1 (GATA2), Jun Proto-Oncogene, AP-1 Transcription Factor Subunit (Jun), MYB Proto-Oncogene, Transcription Factor (Myb) and RUNX1 (Gupta et al. 2009). PU.1 directly activates up to 67 genes such as immunoglobulins and Ig receptors, cytokines and cytokine receptors involved in leukocyte growth and development and inflammation with the majority of the protein products located in the plasma membrane, illustrating the important role of PU.1 in cell communication (Turkistany and DeKoter 2011). In the

rAML mouse model CBA/Ca, AML mainly develops after irradiation through the deletion of one allele of the gene *Sfpi1* and a mutation in the remaining copy affecting codon R135 (Silver et al. 1999). In about 90% of mice that develop rAML, a partial deletion of one copy of chromosome 2 is apparent (Azumi and Sachs 1977; Silver et al. 1999; Brown et al. 2015), which is where the myeloid transcription factor *Sfpi1* lies. This is commonly thought of as the first hit in radiation leukaemogenesis. In mice containing chromosome 2 deletions, 70-80% were also found to contain a point mutation at codon R135 in the DNA-binding domain of the protein (Cook et al. 2004) on the second allele which is believed to be the second hit needed for the development of leukaemia. This particular mechanism seems to be highly specific to this mouse model and one that is not commonly found in human therapy-related AML cases (Suraweera et al. 2005). From *Sfpi1* gene knockout studies, homozygous mutations of the DNA binding domain resulted in an almost complete halt in progenitor production evident from clonogenic assays with no viable embryos after 18 days of gestation (Scott et al. 1994). This shows that *Sfpi1* is necessary for myeloid development and so is a critical factor of the hematopoietic system.

However other mutations must also contribute to the development of rAML in cases lacking *Sfpi1* mutations. In the CBA/Ca mouse, *Flt3 ITDs* were detected specifically in cases with neither chromosome 2 deletions nor *Sfpi1* mutation, suggesting a different mechanism of rAML development (Finnon et al. 2012). A nonsynonymous single nucleotide polymorphism (nsSNP) in the gene C-terminal binding protein (CTBP)-interacting protein (CTIP)/retinoblastoma binding protein 8 (*RBBP8*) in the mouse has been identified which disrupts its homologous recombination repair function and so may play a role in the development and susceptibility to rAML (Patel et al. 2016). Although the Q418 nsSNP has not been reported in human cases, it points towards other regions of this gene that might affect the risk of AML and worth investigating.

1.7.2 Epigenetic changes

Several factors, including regulation of transcriptional expression, microRNA (miRNA) and interactions with proteins tightly regulate the expression of *Sfpi1*. *Sfpi1* contains upstream distal regulatory regions. Epigenetic analysis in human cases has revealed the promoter and -17 kb upstream regulatory element of the *PU.1* gene to be highly methylated in classical Hodgkin lymphoma cells (Yuki et al. 2013) and in myeloma

cells (Tatetsu et al. 2007) with reduced *PU.1* expression. Deletion of an upstream regulatory element (URE) which is necessary for expression of the gene has been shown to reduce *Sfpi1* expression down to 20% of wild type levels in a murine mouse model and results in the development of AML within 3-8 months (Rosenbauer et al. 2004) with even a minimal reduction in expression of 35% leading to the transformation to AML (Will et al. 2015). Epigenetic regulation of *PU.1* expression is mediated by the microRNA-155 (*miR-155*), having both pro-tumorigenic and anti-tumorigenic functions. Increased expression of *miR-155* decreases *Sfpi1* mRNA levels which develops into a myeloproliferative disorder (O'Connell et al. 2008) while *miR-155* deficient cells produces an up-regulation of *PU.1* (Vigorito et al. 2007). It is also regulated by its many interactions with many proteins such as transcription factors, chromatin remodelling factors and proteins that function in cell cycle regulation (Gupta et al. 2009). Despite the very low frequency of *Sfpi1* mutations in human AML cases, a recent review by Verbiest et al. has highlighted the possible epigenetic mechanisms, such as *miR-155*, which may instead affect *Sfpi1* expression (Verbiest et al. 2015). A recent study by Badie et al. has developed a radiation-induced AML gene expression signature in a murine rAML mouse model, compromising of 17 genes/ proteins which are distinguishable from normal control cells (Badie et al. 2016). This illustrates that the development of AML involves a mixture of genetic mutations and epigenetic changes which up until now has not been fully investigated.

1.8 Hematopoietic stem and progenitor cells

AML is thought to develop from initial mutations occurring in the hematopoietic stem and progenitor cells (HSPC). Hematopoietic stem cells are multipotent self-renewing cells that reside in the bone marrow and give rise to all blood cells through the process of hematopoiesis. This self-renewing property was first demonstrated by Till and McCulloch in 1960 after the transplantation of bone marrow cells into lethally irradiated mice led to reconstitution of the bone marrow (McCulloch and Till 1960). Bone marrow transplantation assays remain the standard technique to confirm the presence of stem cells. HSCs reside in the stem cell niche in the pelvis, sternum, femur, spine and in smaller populations in umbilical cord blood with a few detectable in the peripheral blood. HSCs can now be separated into long-term HSCs and short-term HSCs. Long term HSCs, also known as dormant HSCs, are defined by their ability to give rise to

all blood cell lineages and also must be able to reconstitute the bone marrow niche of a lethally irradiated niche following transplantation, resulting in long-term survival. They are located beside osteoblasts in the endosteal niche and they lie in a quiescent state until there is a requirement for blood cells to be replenished (Figure 6). These dormant HSCs then become activated HSCs and relocate to the vascular niche beside sinusoids. The activated HSCs can divide either asymmetrically producing another stem cell or symmetrically into a multipotent progenitor (MPP). MPPs have a limited self-renewal capacity and can be further separated into MPP1, MPP2 or MPP3, depending on their phenotype. MPPs can follow a myeloid path (Common Myeloid Progenitor) or lymphoid path (Common Lymphoid Progenitor) of development. Genetic alterations such as insertions, deletions, translocations, inversions or mutations may occur in these stem cells or progenitors after radiation exposure potentially inhibiting further development. This causes them to remain in an immature state and increases proliferation, leading to an overproduction of immature white blood cells and a reduction in functioning leukocytes, red blood cells and platelets. This leads to symptoms such as fatigue, bleeding and shortness of breath.
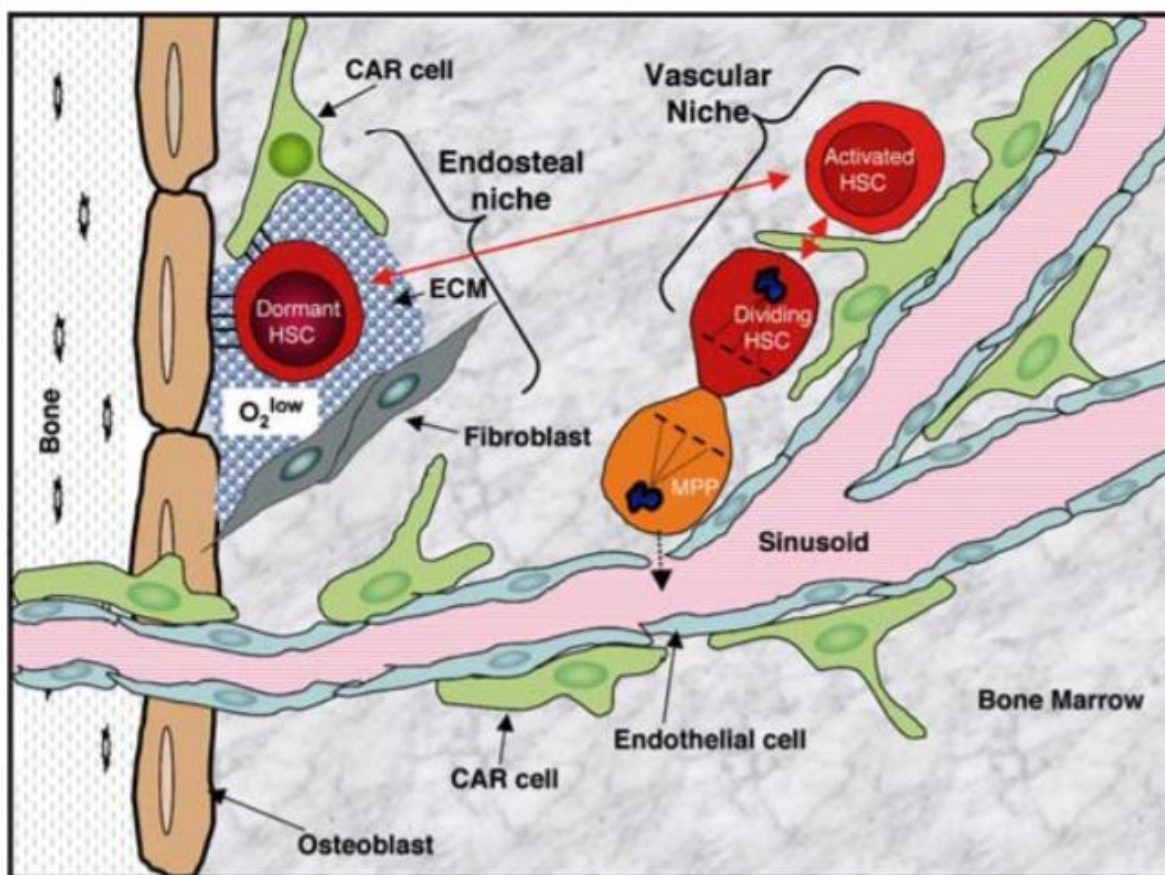
Figure 6. Model of the stem cell niche.

Dormant HSCs lie in a dormant state in the endosteal niche and relocate to the vascular niche upon activation. The activated HSC can divide to form a HSC or MPP. CAR = Cxcl12-abundant Reticular, ECM = extra-cellular matrix. This figure is copied from (Wilson et al. 2007).

1.8.1 HSPC sub-populations

In order to study HSPC, mouse HSCs were first isolated in 1988 by a lab in Stanford University (Spangrude, Heimfeld, and Weissman 1988) using lineage negative (Lin-) and populations positive for stem cell antigen-1 (sca1) cell surface markers with a Thy-1 Membrane Glycoprotein (Thy1.1)lo expression. C-kit was discovered to play an important role in haematopoiesis by the mapping of c-kit to the W locus in the stem cell defective W/W mice and subsequent transplantation assays with Lin-, c-kit+ cells resulted in reconstitution of lethally irradiated mice (Okada et al. 1991). Further development has led to the use of Lin-, Sca1+, c-kit+ markers, known as LSK, to isolate HSCs (Adolfsson et al. 2001).

Other markers have been investigated to further classify the LSK cell population. The signalling lymphocytic activation molecule (SLAM) family cell surface receptors, such as CD150 and CD48 are expressed in different sub-sets of HSCs. CD150 transcriptional expression is up-regulated in HSCs in comparison to MPPs and in transplantation assays recipients of CD150+ cells resulted in long-term reconstitution while recipients of CD150- cells resulted in short-term reconstitution (Kiel et al. 2005). CD150 therefore has been identified as a marker of long-term HSCs. Transplantation of CD48+ cells however resulted in B and T cell reconstitution, while CD48- cells resulted in long-term reconstitution (Kiel et al. 2005). CD34 has been identified as a marker of long-term HSCs. Mouse HSC colony forming unit (CFU) assays showed CD34 expression to positively correlate with colony growth. However, CD34- HSCs cultured with IL-3 and stem cell factor resulted in 80% of the cells forming multi-lineage colonies (Osawa et al. 1996). The tyrosine kinase receptor *Flt3* was found to be expressed in 60% of the LSK population and was investigated as it was indicated to play an important role in the growth of HSCs (Lyman and Jacobsen 1998). Transplantation assays have revealed that Lin-, Sca1+, c-kit+, flt3- cells produced multilineage reconstitution whereas Lin-, Sca1+, c-kit+, flt3+ cells reconstituted B and T lymphocytes with only a short-term myeloid reconstitution (Adolfsson et al. 2001). *Flt3*, also known as CD135, therefore appears to be expressed on lymphoid progenitors which do not have self-renewal capacity. HSC can now be further categorised into long-term HSCs (Kiel et al. 2005), intermediate-term HSCs (Benveniste et al. 2010), common lymphoid progenitors (Kondo, Weissman, and Akashi 1997), common myeloid progenitor (Akashi et al. 2000) by differences in expression of Sca-1, macrophage-1 antigen (Mac-1) (Morrison and Weissman 1994) CD34 (Osawa et al. 1996), *Flt3* (Adolfsson et al. 2001), CD150, CD48 and CD49B (Wagers and Weissman 2006).

Taking together the previous studies over the past couple of decades, a model of HSC hierarchy has been proposed by Wilson et al. illustrating the loss or acquisition of various cell surface markers as a HSC evolves (Figure 7). Cycling activity, measured by Ki67 expression and DNA content, showed that the LSK+, CD34-, CD48-,CD150+, CD135- population had 70% of the population in G_0 and less than 2% actively cycling with other cell populations showing larger populations cycling and less in G_0 (Wilson et al. 2008). This specific population was therefore designated HSC. This was further

confirmed by a label-retaining assay where DNA was labelled in vivo with bromodeoxyuridine (BrdU) to detect nondividing cells. The HSC population was found to retain the most BrdU labelled cells, again showing it to be the most quiescent. Using a mathematical model of BrdU decay, the HSC population was determined to be composed of two populations, a small dormant population and a larger activated population. HSCs can reversibly switch between being a dormant HSC and an activated HSC with self-renewing properties when needed due to bone marrow injury or G-CSF stimulation and then differentiate into MPPs (Wilson et al. 2008). These advancements in stem cell characterisation and isolation will allow further investigation into the development of radiation-induced leukaemia and insight into the stem cell niche as a whole.
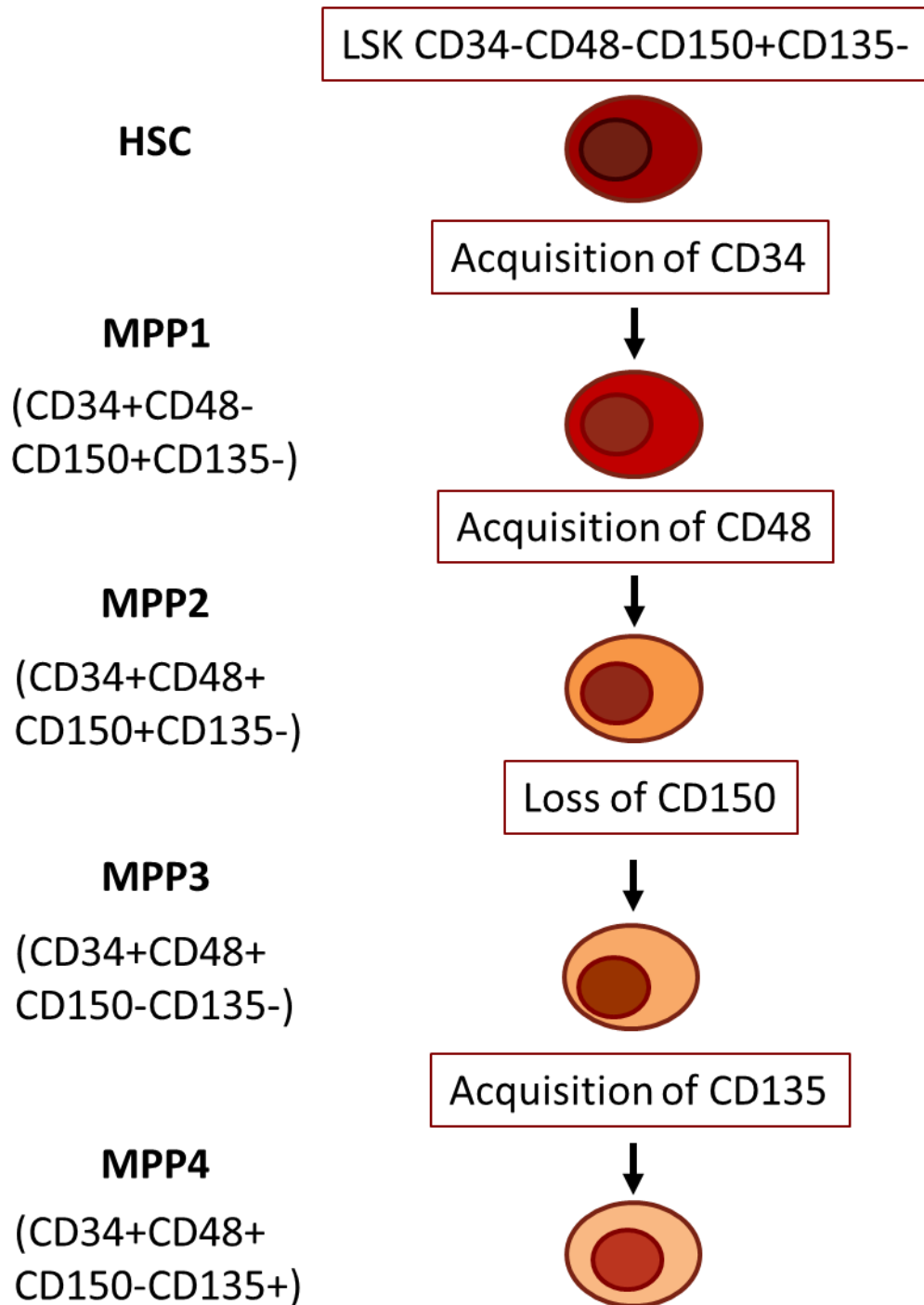
Figure 7. Proposed model of stem cell development in the mouse.

Adapted from (Wilson et al. 2007; Wilson et al. 2008).

1.8.2 HSC heterogenicity

Traditionally, transcriptional analysis of HSPCs has been by bulk analysis, however, single cell analysis in recent years has revealed the heterogeneity of a pool of cells in a population (Figure 8). Single cell transcriptional analysis of murine long-term HSC, lymphoid progenitors and erythroid progenitors has identified heterogeneity in transcription factor expression among the cell types (Moignard et al. 2013). Some genes such as meis homeobox 1 (*Meis1*) showed different expression levels between cell types, while other genes showed a range of expression levels within a single population. Single cell RNA sequencing analysis of 2000 single cells from chronic myeloid leukaemia patient samples revealed heterogeneity within the CML population while also identifying rare BCR-ABL+ populations persisting during remission (Giustacchini et al. 2017).

To isolate HSCs, differentiated cells are usually first removed from the total bone marrow sample by immunobead subtraction of lineage positive cells so that the remaining cells are lineage negative. Cell surface markers can then be used to separate the cells into sub-populations. Recent advances in technology, in particular, single cell capture methods, has led to an increased analysis and understanding of hematopoietic stem and progenitor cell populations. There are numerous technologies to research single cell genomics and transcription such as digital droplet polymerase chain reaction (PCR), Fluorescence-activated cell sorting (FACS) single cell sorting, microfluidics, laser capture microdissection or microwell capture. These technologies allow for the analysis of single cell RNA and DNA sequencing. This will allow further investigation of these distinct sub-populations in the development of radiation induced leukaemia and advance our understanding of haematopoiesis.
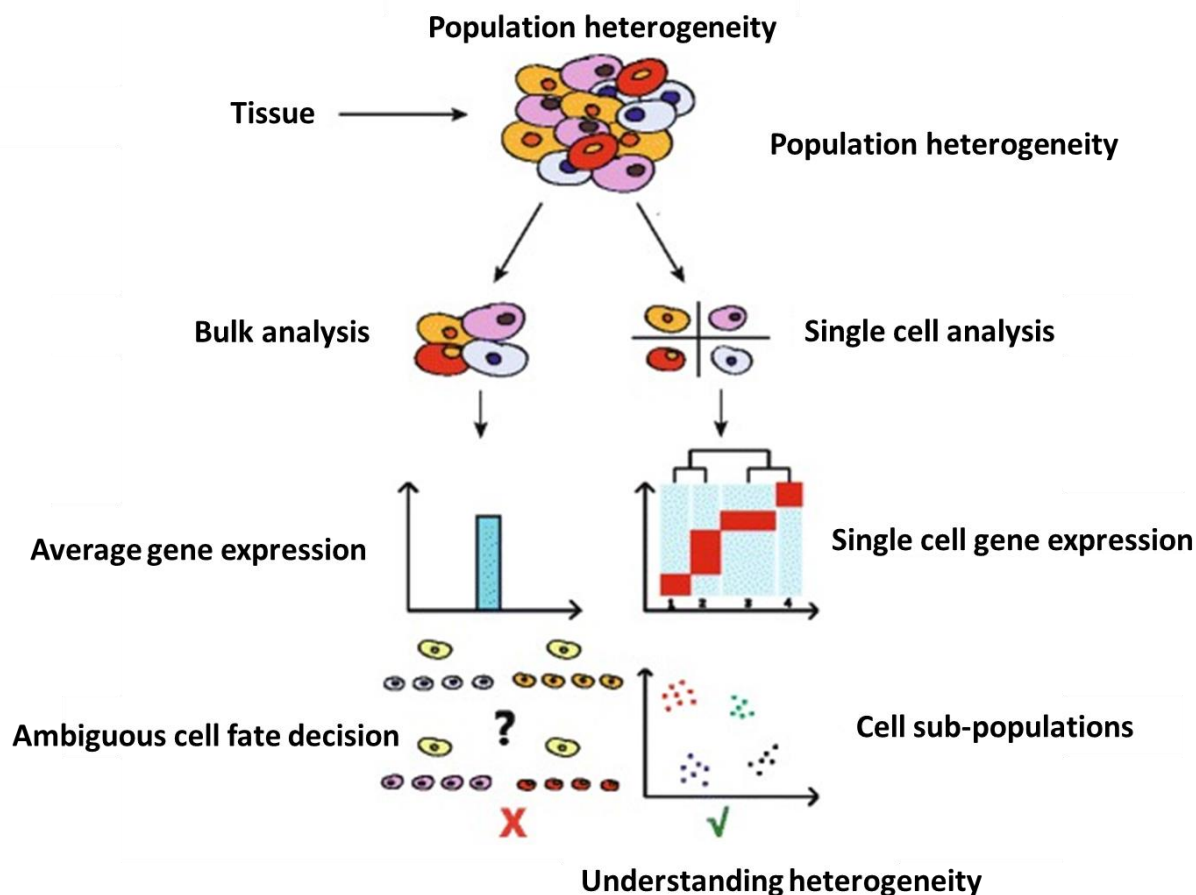
Figure 8. Heterogeneity revealed by single cell analysis.

Using traditional bulk analysis of a pool of cells, gene expression analysis reports an average expression level of all cells. On a single cell level, transcriptional expression is revealed per cell and cells populations can be further separated into sub-populations based on their expression levels. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/). Taken from (Ye, Huang, and Guo 2017).

1.9 Objectives of the project

The main purpose of this research is to examine the target cells for rAML development, the hematopoietic stem and progenitor cells where the initiation of leukaemic events occurs. Previous studies have allowed classification of HSPC into three sub groups based on their repopulating abilities i.e. long term HSC, short term HSC and haematopoietic progenitor cells (Zhong et al. 2005). In order to do so, we aimed to

characterise the response and sensitivity of these sub-populations of HSPC to IR. We will also analysed the genetic and epigenetic alterations that occur in human de novo AML and t-AML samples in order to improve our understanding of the mechanisms of radiation-induced leukaemogenesis as outlined in Figure 9. Mouse and human samples of rAML were used during the course of this PhD project, assessing the validity of the mouse model for humans and making for a first time an interspecies comparison analysis.
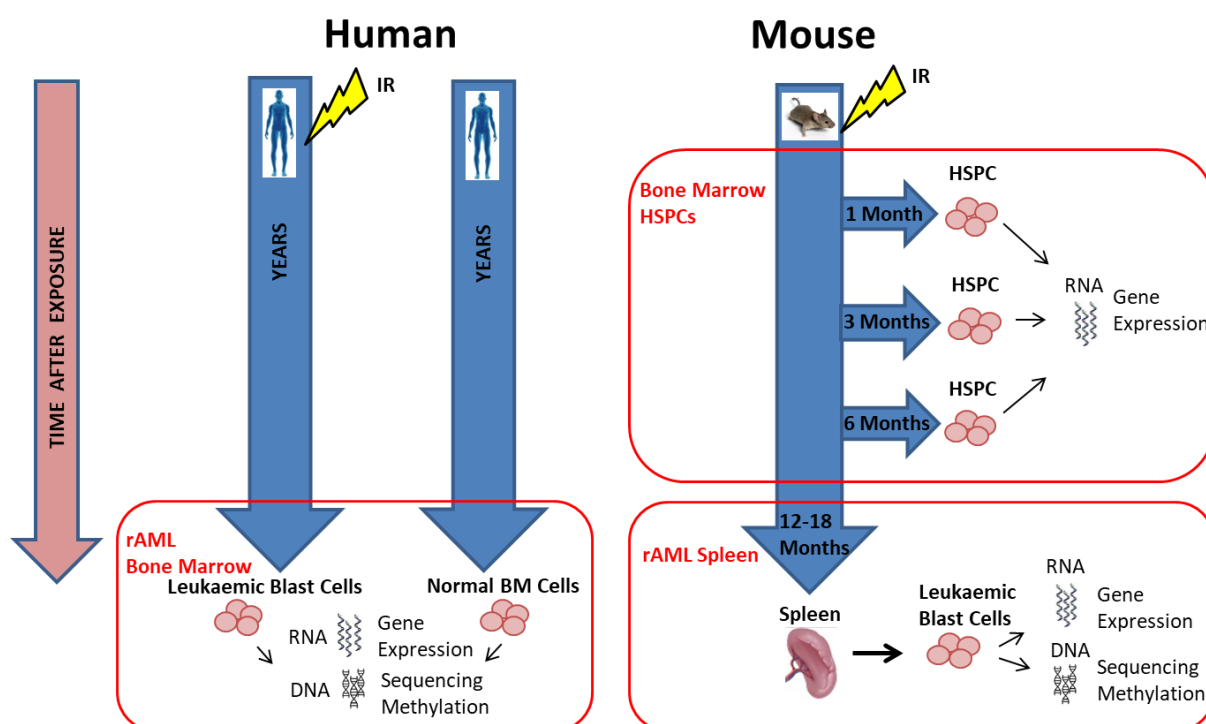
Figure 9. Experimental plan for mouse and human samples.

Human DNA and bone marrow samples from de novo AML and t-AML patients will be collected and analysed by gene expression, DNA sequencing and DNA methylation analysis. HSPC will be isolated from CBA/Ca and C57BL/6 mice at various times after IR exposure and separated into HSPC sub-populations. RNA will be extracted from the samples and gene expression performed. RNA and DNA will also be isolated from rAML spleen tissues and gene expression, DNA sequencing and DNA methylation performed.

1. Human bone marrow samples were obtained from de novo AML patients and AML after chemotherapy and radiotherapy patients which were analysed by DNA sequencing of previously identified genes mutated in AML.

2. Human AML patient samples were also analysed by gene expression and DNA methylation analysis.

3. A historical bank of murine rAML spleen samples were used to identify genetic mutations, epigenetic changes and identify genes of interest in the development of rAML.

4. An interspecies comparison of *PU.1/Sfpi1* expression in human and mouse AML cases was performed.

5. A protocol was developed for the single cell transcriptional analysis of HSPC populations after irradiation

6. mRNA gene expression changes were analysed in murine HSPC sub-populations in response to irradiation by molecular counting (nCounter analysis) and genes of interest identified and validated by quantitative PCR.

2 Materials and methods

2.1 Common reagents

Pipettes, pipette tips, 1.5 ml microcentrifuge tubes and 0.5 ml microcentrifuge tubes were purchased from STARLAB (STARLAB UK Ltd., Milton Keyes, UK). 15 ml and 50 ml conical bottom tubes were purchased from SARSTEDT (SARSTEDT Ltd, Leicester, UK). Centrifugation for microcentrifuge tubes was performed using a MicroStar 17R centrifuge (VWR International Ltd., Lutterworth, UK) and centrifugation for conical bottom tubes was performed in Sorvall Legend RT (Thermo Fisher Scientific, Paisley, U.K.).

2.2 Samples

2.2.1 Mice

All CBA/Ca and C57/BL6 mice animal procedures conformed to the United Kingdom Animals (Scientific Procedures) Act 1986, Amendment Regulations 2012. Experimental protocols were approved by the Home Office and institutional Animal Welfare Ethical Review Body (AWERB) (PPL 30/3355; 21 December 2015). CBA/Ca and C57/BL6 mice were provided with sterile water and food *ad libitum* and subject to a 12 h light/12 h dark cycle.

2.2.2 Mouse spleen samples

A bank of 123 radiation-induced AML samples (104 spleen and 19 DNA samples) were held at Public Health England, Chilton, Oxfordshire. The samples were a combination of 95 CBA/H mice, 30 of which were exposed to 3 Gy X-rays, 3 exposed to 4.5 Gy X-rays and 62 of which were exposed to neutrons and 28 F1 CBA/H x C57BL/Lia which were exposed to 3 Gy X-rays. Mice were irradiated at 10-12 weeks of age at MRC Harwell using a Pantak X-ray source 250kVp, 11 mA at a dose rate of 0.887 Gy/Min to give a single whole body dose. The 62 neutron-induced AMLs were irradiated with fast fission neutrons from a 235U converter in the Low Flux Reactor at NRG, Petten, Netherlands.

AMLs were diagnosed using the criteria described in the Bethesda Proposals for Classification of Non-lymphoid Hematopoietic Neoplasms in Mice (Kogan et al. 2002). Mice were examined daily for signs of illness and were euthanized with a rising concentration of CO_2 when moribund. Mice found to have increased white blood cell

counts in the peripheral blood and to display splenomegaly upon dissection were treated as suspect AMLs. Diagnosis was confirmed by examination of blood films, immunophenotyping and transplantation of tumours into recipient hosts. All animals were bred and handled according to UK Home Office Animals (Scientific Procedures) Act 1986 and with guidance from the local animal welfare body. All spleen samples were stored in RNALater (Thermo Fisher Scientific, Paisley, U.K.) at the time of sacrifice.

2.2.3 Human AML patient samples

DNA from 21 historical AML patients was obtained from the National Centre for Research "Demokritos", Athens, Greece with informed consent from each patient. Bone marrow aspirates were obtained from 5 normal donors (ALLCELLS, Alameda, CA) with informed consent and ethical approval from the Alpha Independent Review Board. Bone marrow aspirates from 7 AML patients were also obtained from an existing biobank in the Northern Institute for Cancer Research (NICR), Newcastle with informed consent and ethical approval from the Newcastle and North Tyneside Research Ethics Committee (REC 17/NE/0361) and with the appropriate material transfer agreement in place. Age, gender, cytogenetic data were obtained where possible. Approval for human cell work was obtained from the Brunel Research Ethics Committee (reference 11466-TISS-Oct/2018- 14402-4).

2.3 X-ray exposure

In vivo irradiations were performed at the Centre for Radiation, Chemical and Environmental Hazards (Public Health England, Harwell, UK) using an X-ray source (AGO, Reading, UK; 250 kVp and 13 mA) with a dose rate of 4.9 mGy/min. All x-ray exposures were performed by Mr Paul Finnon.

2.4 Tissue harvest and cell preparation

All mouse euthanasia and dissections were performed by Dr Rosemary Finnon. Mice were sacrificed using a rising concentration of CO_2. Femurs, tibias, iliac crests and spine were dissected from the mice. The bones were cleaned including removal of the spinal cord and excess tissue. The bones and spine were then crushed in a small volume Iscove's Modified Dulbecco's Media (IMDM) using a pestle and mortar and

disaggregated with an 18 G needle and then passed through a 40 micron cell strainer to produce a single cell bone marrow suspension. A cell count was performed using an equal volume of cell suspension with 0.4% Trypan Blue (Thermo Fisher Scientific, Paisley, U.K.) using a Neubauer haemocytometer (Nanoentek, Korea). All reagents were obtained from Sigma-Aldrich, Dorset, UK, unless otherwise stated.

2.5 Immunomagnetic negative selection of haematopoietic stem and progenitor cells

Lineage depletion of the bone marrow cells was performed by immunomagnetic negative selection using the EasySep^{TM} Mouse Hematopoietic Progenitor Cell Isolation Kit (Stem Cell Technologies, Grenoble, France). All reactions were performed on the RoboSep^{TM} and RoboSep^{TM}-S (Stem Cell Technologies, Grenoble, France). Bone marrow cell suspensions were prepared at 1 x 10^8 cells/ml EasySep™ Buffer in a 5 ml Falcon™ polystyrene round-bottom tube (BD Biosciences, Oxford, U.K.). Normal Rat Serum was added at 50 µl/ml of cells. The samples and kit reagents were then loaded onto the RoboSep^{TM}. Briefly, following 15 minutes incubation with the EasySep^{TM} Mouse Hematopoietic Progenitor Cell Isolation Cocktail at 50 µl/ml of cells, EasySep™ Streptavidin™ RapidSpheres were added at 75 µl/ml of cells and incubated for another 10 min. EasySep™ Buffer was added to bring the cell suspension up to a total volume of 2.5 ml and the mixture pipetted into the EasySep™ Magnet for 3 min, after which the negative fraction containing the HSPCs was pipetted off into a fresh tube. The cell number was determined using a Neubauer haemocytometer (Nanoentek, Korea) with 0.4% Trypan Blue (Thermo Fisher Scientific, Paisley, U.K.).

2.6 Flow cytometry analysis and cell sorting

Following bone marrow lineage depletion, cell suspensions were prepared at 5 x 10^6 cells/ 100 µl. Primary conjugated antibodies were then added, and all reagents were diluted to the manufacturers' recommendations. Two sets of conjugated antibodies were used. Set one was composed of CD27 (LG.3A10) fluorescein isothiocyante (FITC) (BioLegend, San Diego, USA) and CD201 phycoerythrin (PE) Figure 10. Set two was composed of Sca1 PECy7 (D7; Affymetrix, High Wycombe, UK), cKit-APCe780 (2B8; Affymetrix), CD150-APC (TC15-12F12.2; BioLegend, San

Diego, USA), CD48 FITC (HM48-1; BioLegend, San Diego, USA), CD135 PE (A2F10; BioLegend, San Diego, USA) and CD34 Alexa Fluor 700 (RAM 34; Invitrogen, Carlsbad, USA) Figure 11. Following incubation for 30 min on ice, cells were washed twice in EasySep™ Buffer. Flow cytometry acquisition and sorting was performed using a MoFlo XDP (Beckman Coulter, High Wycombe, UK) equipped with four lasers [355nm (UV); 488nm (Blue); 561nm (Yellow-Green); 640nm (Red)] and analysed using Summit 5.4 software (Beckman Coulter)]. Single colour compensation controls were prepared for each fluorochrome (UltraComp eBeads®; Affymetrix). Cells were gated for size, shape and granularity, using forward and side scatter parameters. Regions were drawn around major populations of cells on a forward scatter/side scatter (FSC/SSC) dot plot. Dead cells, clumps and debris were excluded from further analysis. For dead cell exclusion, 7-amino-actinomycin D (7-AAD; BD Biosciences, Oxford, U.K.) was added 10 mins before analysis. Operation of the MoFlo XDP sorter was performed by Mr Andrew Worth, Oxford University, UK.
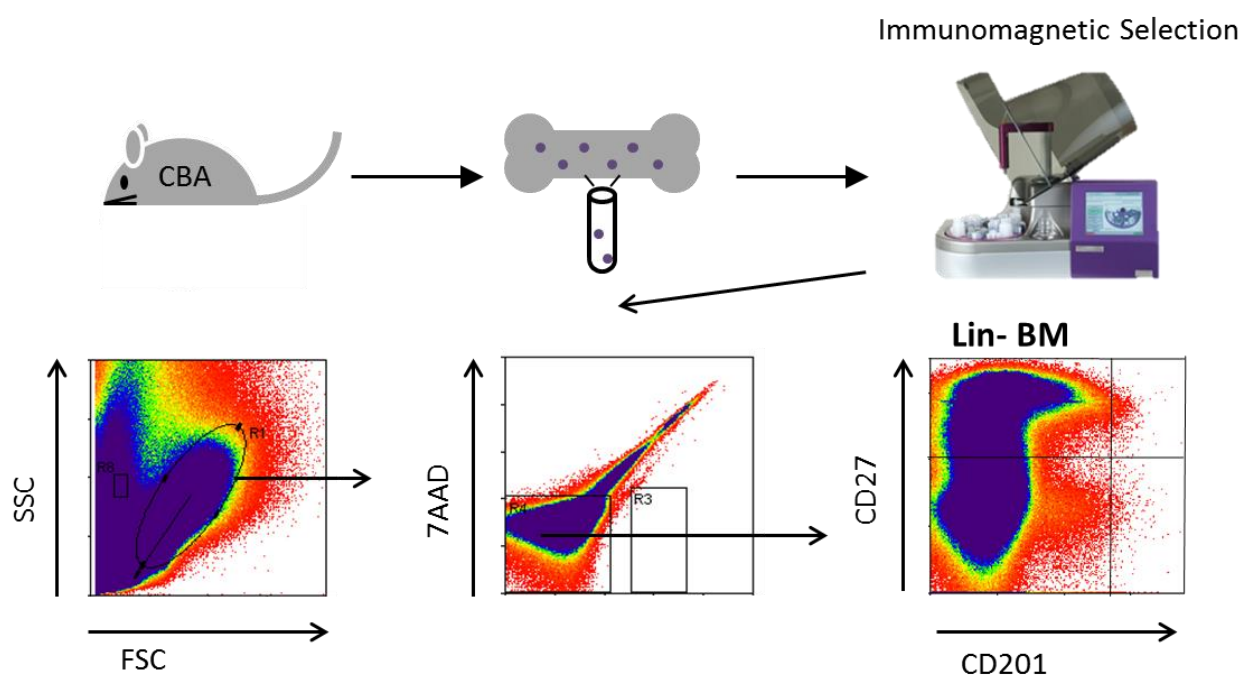
Figure 10. Flow cytometry cell sorting set one protocol.

The bone marrow of CBA/Ca mice was crushed obtaining a cell suspension. Lineage depleted cells were isolated from the cell suspension by immunomagnetic selection on the RoboSep. Cells were gated for size, shape and granularity using forward and

side scatter and dead cells excluded using 7AAD. Long-term HSCs were isolated that were CD27+/CD201+.
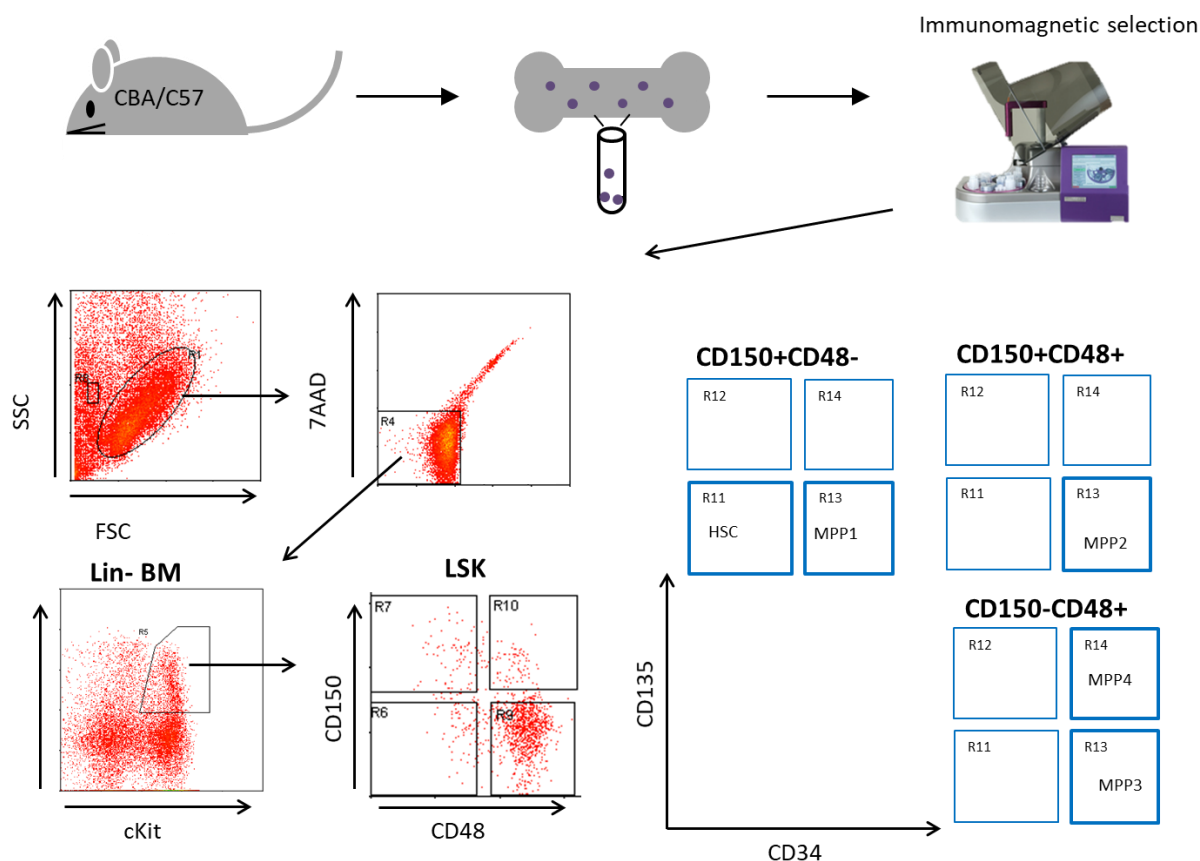
Figure 11. Flow cytometry cell sorting set two protocol.

The bone marrow of CBA/Ca mice was crushed obtaining a cell suspension. Lineage depleted cells were isolated from the cell suspension by immunomagnetic selection on the RoboSep. Cells were gated for size, shape and granularity using forward and side scatter and dead cells excluded using 7AAD. LSK cells were isolated by gating cells Sca-1+/cKit+. Five populations of cells were isolated HSCs (Sca-1+, cKit+, CD150+, CD48-, CD135-, CD34-), MPP1 (Sca-1+, cKit+, CD150+, CD48-, CD135-, CD34+), MPP2 (Sca-1+, cKit+, CD150+, CD48+, CD135-, CD34+), MPP3 (Sca-1+, cKit+, CD150-, CD48+, CD135-, CD34+) and MPP4 (Sca-1+, cKit+, CD150-, CD48+, CD135+, CD34+).

2.7 RNA extraction

RNA was extracted from murine spleen tissues using the miRNeasy Mini Kit (Qiagen Ltd, Crawley U.K.) according to manufacturer's guidelines, including the optional DNase I treatment step. RNA was extracted from murine sorted hematopoietic stem and progenitor cells using the following kits; the miRNeasy Mini kit (Qiagen Ltd, Crawley U.K.), the Single Cell RNA Purification Kit (Norgen Biotek Corp., Canada) and the ReliaPrep™ RNA MiniPrep Systems (Promega, Southampton, U.K.) according to manufacturer's guidelines and including the optional DNase I treatment step. RNA was extracted from human bone marrow cells using the miRNeasy Mini Kit (Qiagen Ltd, Crawley U.K.) according to manufacturer's guidelines.

RNA quantity was measured using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Paisley, U.K.). RNA quality was assessed using the 2200 TapeStation (Agilent Technologies Ltd., Wokingham, UK) according to manufacturer's protocol.

2.8 RNA quantity and quality measurement

RNA quantity was measured using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Paisley, U.K.) according to manufacturer's instructions.

2.8.1 2200 TapeStation assessment

RNA quality was assessing using the 2200 TapeStation (Agilent Technologies Ltd., Wokingham, UK) according to manufacturer's instructions. Briefly, R6K Screentape and R6K Buffer were placed at room temperature for 30 min. 1 µl RNA sample was mixed with 4 µl R6K buffer in an 8 well strip (ABgene/Thermo Scientific, Paisley, U.K.). Samples were vortexed using a MixMate vortexer (Eppendorf UK Ltd., Stevenage, UK) at 2500 rpm for 15 sec and spun down at 3000 rpm using a Sigma 2-16P centrifuge (Sciquip Ltd. Newtown, UK). The strip was then placed at 72 °C for 3 min in a Thriller® thermoshaker-incubator (Peqlab Ltd., Sarisbury Green, UK) and cooled on ice for 2 min. The strip was again centrifuged for 1 min at 3000 rpm to collect sample from inside the lid. The strip was then placed on the Tapestation instrument for analysis, producing RNA Integrity Number equivalent (RINe) which indicated the level of RNA quality.

2.9 DNA extraction

DNA was extracted from murine spleen tissue and human bone marrow cells using the DNeasy® Blood & Tissue kit (Qiagen, Manchester, UK) according to the manufacturer's instructions. DNA quantity was measured by NanoDrop 2000 (Thermo Fisher Scientific, Paisley, U.K.) and quality assessed by agarose gel electrophoresis using 1.3% agarose gel.

2.10 Agarose gel electrophoresis

Agarose gel electrophoresis was used to assess DNA quality, to detect insertions and to assess PCR amplicons of primer designs. Briefly, 0.65 g of Certified Molecular Biology Grade Agarose (Bio-Rad Laboratories Ltd., Hemel Hempstead, UK) agarose was placed in a beaker with 50 ml 1X Tris-Borate-EDTA (TBE) buffer (Sigma-Aldrich, Dorset, UK) and heated in a microwave for 20 sec after which time the beaker was removed from the microwave and the contents gently shaken, followed by another 10 sec on high power. After allowing to cool slightly, 5 µl GelRed Nucleic Acid Gel Stain 10,000X (Cambridge Bioscience Ltd., Cambridge, UK) was added to the beaker, mixed by gently shaking and poured into an electrophoresis tray (Bio-Rad Laboratories Ltd., Hemel Hempstead, UK) and allowed to solidify for 45 min. The gel was then placed in a Mini-Sub® Cell GT electrophoresis system (Bio-Rad Laboratories Ltd., Hemel Hempstead, UK) with 1 X TBE buffer.

To load the gel, 3 µl sample was mixed with 1 µl loading buffer, consisting of 40 % sucrose and 0.24% bromophenol blue (Sigma-Aldrich, Dorset, UK) in water, and pipetted into the gel wells with 1 µl 1 Kb Plus DNA ladder (Life Technologies) loaded into one lane. The gel was then run at 100 V for 40 min and viewed on a U:Genius3 gel documentation system (Syngene Europe, Cambridge, UK). Good quality DNA was characterised by a strong band above the 12 Kb band with an absence or minimal amount of smearing throughout the gel. PCR products of specific primer designs were characterised by a single sharp peak at the appropriate band size.

2.11 Array Comparative Genomic Hybridisation (aCGH)

Both a custom-made array consisting of 1.4 M loci (NimbleGen Custom CGH 3 x 1.4 M, Roche NimbleGen, Madisonm WI, USA) and G3 Mouse CGH Microarrays (4 x 180

K, Agilent Technologies, Wilmington, DE, USA) were used due to the discontinuation of the NimbleGen product. Briefly, 0.5 μg of AML DNA was labelled with a Cy3 fluorescent dye and control DNA was labelled with a Cy5 fluorescent dye (NimbleGen Dual Colour DNA labelling kit, SureTag DNA labelling kit, Agilent Technologies) according to manufacturer's instructions. Equal amounts of AML DNA and control DNA were mixed and added to the CGH array slide/mixer assembly according to manufacturer's instructions. The slide was then placed in the Hybridisation System for the recommended time, washed with the appropriate buffer (NimbleGen Hybridisation Wash Buffer Kit, Agilent Oligo ACGH/ChIP-on-Chip Wash buffers) according to manufacturer's instructions and dried using an Array-It slide dryer (Arrayit Corporation, Sunnyvale, CA, USA). Arrays were stored in a dark dessicator. Roche array slides were scanned using a NimbleGen MS 200 Microarray Scanner and Agilent slides were scanned with a SureScan Microarray scanner. Analysis using the Roche arrays was performed using DEVA software. Analysis using Agilent slides was performed using Genomics Workbench 7.0. aCGH analysis was performed on the 115 mouse spleen samples by Dr. Natalie Brown.

2.12 Reverse transcription

2.12.1 mRNA

Reverse transcription reactions were performed using a High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, USA) according to the manufacturer's protocol. In brief, 35 μl RNA sample (700 ng RNA/sample) was added to 5 μl reverse transcription buffer, 2 μl dNTP mix, 5 μl random primers, 2.5 μl Multiscribe® Reverse Transcriptase and 0.5 μl water. Reactions were performed on a Verti-96 well Thermal Cycler (Applied Biosystems, Foster City, USA), using the following conditions: 25 °C for 10 min, 37 °C for 120 min and 85 °C for 5 min.

2.12.2 miRNA

To prepare miRNA samples for first strand complementary DNA (cDNA) synthesis, samples were first polyadenylated using the qScript^{TM} microRNA cDNA Synthesis Kit (Quanta BioSciences, Gaithersburg, U.S.A.). A mixture of 100 ng total RNA in 7 μl RNase-free water, 2 μl Poly(A) Tailing buffer and 1 μl Poly(A) polymerase were added together on ice and placed in a thermocycler at 37 °C for 20 min and 70 °C for 5 min.

Following polyadenylation, a 10 µl cDNA synthesis master mix consisting of 9 µl cDNA buffer and 1 µl reverse transcriptase was added to each sample. The samples were placed in a thermocycler at 42 °C for 20 min and 85 °C for 5 min. Controls included a –RT reaction to identify genomic DNA contamination and a –poly(A) reaction to identify non-specific binding of universal primers. Samples were then diluted 50 times before QPCR analysis.

2.13 DNA mutation and insertion sanger sequencing

For DNA assays, a set of PCR primers were designed. Assays were designed using PrimerQuest software (Integrated DNA technologies, Leuven, Belgium). The primers were first checked against the Ensemble database (https://www.ensembl.org/index.html) to avoid the presence of SNPs. The primers were tested for specificity with a SYBR Green assay consisting of 1 µl cDNA, 5 µl PerfeCTa SYBR® Green SuperMix (Quanta Biosciences, Gaithersburg, USA), 0.15 µl (10 Mm) forward primer, 0.15 µl (10 µM) reverse primer and 3.7 µl RNase-free water and a control sample with 1 µl water in place of cDNA. The primers were run on a RotorGene (Qiagen, Manchester, UK) with cycling conditions of 95 °C for 2 min, 40 cycles of 95°C for 10 sec and 60 °C for 60 sec followed by a melt curve. Melt curve results were examined for the presence of a single sharp peak in the cDNA sample and no peak in the control. The samples were run on a 1.3% agarose gel to examine for the presence of a single sharp band of appropriate size. Primer designs for human and mouse sanger sequencing are listed in appendices A and B, respectively. Primer designs listed also includes primers designed taken from Bonadies et al. and Finnon et al. (Bonadies, Pabst, and Mueller 2010; Finnon et al. 2012).

For the PCR reaction, 1 µl of extracted DNA (25 ng DNA/µl) was used in 9 µl PCR master mix [1 µl 10x PCR buffer, 2 µl 5x Q-solution, 1.6 µl 1.25 mM dNTP, 3.15 µl dH$_2$0, 0.5 µl 10 µM forward   primer, 0.5 µl 10 µM reverse primer, 0.25 µl *Taq* DNA Polymerase (Qiagen, Manchester, UK) and amplified at the following conditions: 4 min at 95 °C, then 35 cycles of 30 sec at 95 °C, 30 sec at 57 °C and 30 sec at 72 °C, followed by 10 min at 72 °C. PCR product (1 ng/µl per 100 bp) and primers (3.2 pmol/µl) were sent for sequencing by Source BioScience (University of Oxford, UK). Sequencing files were analysed using Chromas 2.6 software (Technelysium Pty Ltd, Brisbane, Australia). For detection of insertions, 2% agarose gels were prepared, and

PCR products were run at 80 V to allow for the detection of an increase in amplicon size.

2.14 Haloplex DNA sequencing

Human DNA (225 ng) from 16 patients (6 de novo AML, 6 AML after chemotherapy, 3 AML after radiotherapy, 1 AML after MDS) were sent to Oxford Gene Technology, Begbroke, Oxfordshire for sequencing analysis by Haloplex^TM Targeted Enrichment System using the ClearSeq AML panel (Agilent Technologies Ltd., Wokingham, UK). Briefly, the DNA was fragmented by restriction enzymes and ClearSeq biotinylated probes specific to the target region were hybridized to the fragments forming circular DNA molecules. These circular molecules were separated from the rest of the sample using magnetic streptavidin beads and amplified by PCR ready for MiSeq 150bp PE Illumina sequencing analysis.

2.15 Pyrosequencing

2.15.1 Mutational analysis

Human and mouse mutations were confirmed by pyrosequencing through use of the Pyromark Q48 (Qiagen, Manchester, UK) according to manufacturer's guidelines. The PyroMark is a novel method of sequencing able to quantify the percentage of the mutation present in the sample which is based on the emission of light with each added nucleotide with sensitivity of detection down to 2%. Briefly, primers were designed by the Pyromark Q48 Advanced Software, validated by SYBRGreen analysis and amplified by PCR using Pyromark PCR Kit (Qiagen, Manchester, UK). Primer designs for human and mouse mutational pyrosequencing are listed in appendices C and D.

Reactions consisted of 10 ng DNA, 12.5 ul PyroMark PCR Master Mix 2X, 2.5 ul CoralLoad Concentrate (Red) 10X, 0.5 ul forward primer, 0.5 ul reverse primer (one of the primers to be biotinylated) and RNase-free water to a volume of 25 ul. Reactions were performed on a Verti-96 well Thermal Cycler (Applied Biosystems, Foster City, USA), using the following conditions: 95°C for 15 min, 45 cycles of 94°C for 30 sec, 60°C for 30 sec and 72°C for 30 sec. To confirm a strong single band of amplified product, 5 ul PCR product was run on an 1.3% agarose gel. PyroMark Q48 Advanced Reagents were loaded onto the PyroMark with 3 ul Magnetic Beads (Qiagen,

Manchester, UK) and 10 ul PCR reaction. During the reaction, solutions of A, C, G and T nucleotides were sequentially added and removed with an emission of light for the complementary pair and determination of the correct nucleotide. This process continued for the whole strand. Pyrosequencing for *Sfpi1/Pu.1* mutational analysis was performed by Mrs. Lourdes CruzGarcia, CRCE, PHE.

2.15.2 DNA methylation analysis

DNA methylation was analysed by pyrosequencing through use of the Pyromark Q48 (Qiagen, Manchester, UK) according to manufacturer's guidelines. Briefly, primers were designed by the Pyromark Q48 Advanced Software and validated by SYBRGreen analysis. Primer designs for human and mouse DNA methylation pyrosequencing are list in appendices E and F.

Bisulfite conversion was performed on 1 µg DNA using the EpiTect Fast Bisulfite Conversion Kit (Qiagen, Manchester, UK). Briefly, 140 µl bisulfite reactions were prepared with 20 µl DNA, 85 µl bisulfite solution, 35 µl DNA protect buffer in 200 µl PCR tubes (Thermo Fisher Scientific, Paisley, U.K.). Reactions were performed on a Verti-96 well Thermal Cycler (Applied Biosystems, Foster City, USA), using the following conditions: 95°C for 5 min, 60°C for 20 min, 95°C for 5 min and 60°C for 20 min. Mixtures were then passed through MinElute DNA spin columns, washed and the methylated DNA eluted according to the manufacturer's protocol. DNA methylation analysis was then performed on the PyroMark using 50 ng methylated DNA as described previously with cycling conditions of 95°C for 15 min, 45 cycles of 94°C for 30 s, 56°C for 30 s and 72°C for 30 s. A universal methylated mouse DNA standard which was enzymatically methylated at all CpGs by M. SssI methyltransferase from the CpGenome™ Universal Methylated Mouse DNA Standard Set (Merck Millipore, Darmstadt, Germany), was used as a positive control and bisulfite conversion efficiency was evaluated by using normal DNA without bisulfite treatment during pyrosequencing analysis. For human samples a positive and negative methylation control was used from the kit EpiTect Control DNA Set (Qiagen, Manchester, UK). Methylation of human controls was achieved using SssI methylase.

2.16 Multiplex quantitative real time-PCR (MQRT-PCR)

2.16.1 Assay designs

For mRNA assays, a set of short PCR primers and a fluorescent probe were designed. Assays were designed using PrimerQuest software (Integrated DNA Technologies, Leuven, Belgium). When designing TaqMan assays, the primers were first checked against the Ensemble database (https://www.ensembl.org/index.html) to ensure maximum transcript coverage and to avoid the presence of SNPs. The primers were tested for specificity with a SYBR Green assay as previously described.

For mRNA assays for single cell experiments a set of short PCR primers and a fluorescent probe were designed, with at least one primer spanning an exon-exon junction or on separate exons with an intro of at least 1000 bp in between, using Primer Blast software (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primers were tested for specificity with a SYBR Green assay as previously described, testing for specificity on both cDNA and DNA. Designs were chosen which had a single sharp peak in the cDNA samples and no amplification in the DNA sample or control if possible. Designs were included if amplification in the DNA sample was at least 5 Ct after the amplification in the cDNA sample. The samples were also run on a 1.3% agarose gel to examine for the presence of a single sharp band of appropriate size. 3′6-Carboxyfluorescein (FAM)/Black Hole Quencher 1 (BHQ1), 6-Hexachlorofluorescein (HEX)/BHQ1, Texas Red (TEX)/BHQ2, CY5/BHQ3, Atto 680/BHQ3 and Atto 390/Deep Dark Quencher 1 (DDQ1) (Eurogentec Ltd, Fawley, UK) were used as fluorochrome reporters for the hydrolysis probes analysed in multiplexed mRNA reactions. All primers were ordered through IDT (Integrated DNA technologies, Leuven, Belgium) and all probes were ordered through Eurogentec (Eurogentec Ltd, Fawley, UK) except for CY5 probes which were ordered through Sigma (Sigma, Haverhill, UK).

For miRNA assays, primer designs were obtained from the QuantaBio website (Quanta Biosciences, Gaithersburg, USA).

Primer designs for human and mouse MQRT-PCR are listed in appendices G and H respectively. Primer design listed also includes primers designed by Dr Francois Paillier (bioMérieux Ltd, Rhône-Alpes, France).

2.16.2 Standard curve preparation

Preparation of standard curve involved the use of the primers to amplify PCR products for each gene of interest. Standard curve PCR reactions consisted of 10 µl of 10 X PCR Buffer, 20 µl of 5 x Q solution, 4 µl of each deoxyribonucleotide triphosphate (dNTP) (Invitrogen, Carlsbad, USA), 31.5 µl water, 5 µl of each primer at 10 µM, 2.5 µl of Taq Polymerase and 10 µl of cDNA. Reactions were performed on a Verti-96 well Thermal Cycler (Applied Biosystems, Foster City, USA), using the following conditions: 94 °C for 3 min, 35 cycles of 30 s at 94 °C, 30 s at 60 °C, 30 s at 72 °C and 72 °C for 10 min. All reagents unless otherwise stated were obtained from Qiagen (Qiagen, Manchester, UK). PCR products were loaded onto a 1.3% gel and run for 40 min at 100 V. The bands were visualised on a UVT 400-M ultraviolet transilluminator (International Biotechnologies Inc., New Haven, Connecticut, USA) and the band of interest cut out and purified using the MinElute Gel Extraction Kit (Qiagen, Manchester, UK) according to the manufacturer's guidelines. cDNA quantity was measured using the NanoDrop 2000 (Thermo Fisher Scientific, Paisley, U.K.) and molar mass of the PCR product was calculated using the online software Mongo Oligo Mass Calculator v2.06 (http://mods.rna.albany.edu/masspec/Mongo-Oligo).

The number of PCR product copies per µl was calculated with the formula;

$$Copy\ number\ per\ \mu l = \frac{Concentration}{\frac{M}{N_A}}$$

Where M is the molar mass of PCR product and N_A is Avogadro's constant.

The standard curve was prepared by adding 1.5625×10^9 molecules of each PCR product of each gene of interest to water in a total volume of 50 µl. This stock solution was then diluted 1600 times by serial dilution. The standard curve was prepared by using the diluted stock and a further eight 5-fold dilutions.

2.16.3 TaqMan MQRT-PCR

Real-time PCR was performed using a Rotor-Gene Q (Qiagen, Manchester, UK). Reactions were run in triplicate with primer and probe sets for target genes at 300 nM each and 2.5 µl cDNA in 30 µl reaction volume (PerfeCTa® MultiPlex qPCR SuperMix; Quanta Biosciences, Gaithersburg, USA). Cycling parameters were as follows: 2 min

at 95 °C and 45 cycles of (10 sec at 95 °C and 60 sec at 60 °C). Data was collected and analysed using Rotor-Gene Q Series software. Gene target cycle threshold (C_t) values were normalised to an internal control (*Hypoxanthine-Guanine phosphoribosyltransferase 1; HPRT*). C_t values were converted to transcript quantity using standard curves obtained by serial dilution of PCR-amplified DNA fragments of each gene and run in each reaction. The linear dynamic range of the standard curves covering six orders of magnitude (serial dilution from 3.2×10^{-4} to 8.2×10^{-10}) gave PCR efficiencies between 93% and 103% for each gene with $R^2 >0.998$.

2.16.4 miRNA QPCR

miRNA QPCR was performed using a Rotor-Gene Q (Qiagen, Manchester, UK). Assay reactions of 10 µl contained 5 µl PerfeCTa SYBR® Green SuperMix (Quanta Biosciences, Gaithersburg, USA), 0.3 µl Universal primer (Quanta BioSciences, Gaithersburg, U.S.A.), 0.3 µl Assay primer, 3.4 µl RNase-free water and 1 µl cDNA. All reactions were run in triplicate with cycling conditions of 95 °C for 2 min, 45 cycles of 95 °C for 10 sec and 60 °C for 30 sec followed by a melt curve. Data was collected and analysed using Rotor-Gene Q Series software. Housekeeping miRNA were identified and analysed through use of the NormFinder algorithm (Andersen CL, 2004) which determines the most stably expressed control genes and is located on the RefFinder website http://leonxie.esy.es/RefFinder/?type=reference.

2.17 Low cell number/single cell analysis

2.17.1 REPLI-g® Cell WGA & WTA kit

Low cell numbers were amplified using REPLI-g® WGA & WTA Kit according to manufacturer's guidelines. Briefly, cells were sorted by FACS into 12 well strips, topped up to 13 µl with H_2O and 8 µl Lysis Buffer added. The tubes were mixed carefully and centrifuged briefly. gDNA was first removed from the whole transcriptome amplification (WTA) cell lysate by adding 2 µl gDNA wipeout buffer and incubating the mixture at 42 °C for 10 min. An RT master mix was prepared with 4 µl RT/polymerase buffer, 1 µl H_2O, 1 µl random primer, 1 µl oligo DT primer, 1 µl Quantiscript RT enzyme mix and 8 µl of the mixture added to the WTA cell lysate. A whole genome amplification (WGA) master mix was prepared with 4 µl RT/polymerase buffer, 2 µl gDNA wipeout buffer, 1 µl H_2O, 1 µl random primer 1 µl oligo-dT primer, 1 µl WGA ready enzyme. 10

µl of the master mix was added to the WGA lysed cell sample. Both WTA and WGA samples were incubated at 42 °C for 60 min, 95 °C for 3 min and then cooled on ice. A ligation mix of 8 µl ligase buffer and 2 µl ligase mix was prepared and 10 µl ligation mix added to both WGA and WTA samples and incubated at 24 °C for 30 min and 95 °C for 5 min. Finally, a REPLI-g SensiPhi amplification mix was prepared with 29 REPLI-g sc reaction buffer and 1 REPLI-g SensiPhi DNA polymerase and 30 µl added to both WGA and WTA samples. The samples were then incubated at 30 °C for 2 h and 65 °C for 5 min. The amplified cDNA was diluted 1:100 before MQRT-PCR.

2.17.2 REPLI-g® WTA Single Cell kit

The REPLI-G® WTA Single Cell Kit was used to perform whole transcriptome amplification according to manufacturer's guidelines. The protocol was as previously described for WTA in the REPLI-g® Cell WGA & WTA kit.

2.17.3 CellsDirect^{TM} modified Moignard et al. protocol

Single cell MQRT-PCR analysis was performed using the CellsDirect^{TM} One-Step qRT-PCR kit, with a pre-amplification step included as in Moignard et al. 2013 (Moignard et al. 2013), on the Rotor-GeneQ (Qiagen, Manchester, UK). cDNA synthesis and specific target amplification (preamplification) of genes of interest were performed using the CellsDirect^{TM} One-Step qRT-PCR Kit (Invitrogen, Carlsbad, U.S.A.). Single cells were sorted by FACS directly into 12 well strips of a 5 µl lysis mix consisting of 5 µl CellsDirect 2X reaction mix and 0.1 µl SUPERase RNase inhibitor. The strips were centrifuged at 300 x g, 4 °C for 5 min to let the cells go into the CellsDirect 2x reaction mix. The strips were sealed and stored at -80 °C to allow for better lysis efficiency. A Reverse Transcription Specific Target Amplification (RT-STA) master mix was prepared. Forward and reverse primers for each gene of interest were diluted to 20 µM in the same tube. 1.5 µl of each gene primer mix were pooled together and topped up to 150 µl with TE buffer to make up the Assay Mix. When thawed, a preamplification master mix of 2.5 µl Assay Mix, 1.3 µl TE buffer and 0.2 µl SuperscriptIII/Platinum Taq was added to the wells. Cycling parameters were 50 °C for 15 minutes, 95 °C for 2 minutes, 40–50 cycles of: 95°C for 15 seconds, 60 °C for 4 min. The cDNA was diluted 1;5 with water and analysed by MQRT-PCR as previously described.

2.18 nCounter low input analysis

Cells lineage-depleted bone marrow sub-populations were screened for specific cancer gene expression by use of the nCounter PanCancer Pathways Expression Assay Kit (NanoString Technologies®, Inc., Seattle, WA, USA) on the nCounter Analysis System (NanoString Technologies®, Inc., Seattle, WA, USA) according to the manufacturer's guidelines. The nCounter miRNA Sample Preparation Kit involves the ligation of unique tags onto their target miRNAs which then allows detection of the miRNA signal. nCounter Analysis System then utilizes a novel digital colour-coded barcode technology, involving target-specific biotinylated capture probes and barcode containing reporter probes, which can count hundreds of transcripts in a single reaction and gives a direct measurement of gene expression. To screen gene expression in dormant, activating, MPP1, MPP2 and MPP3 populations, 300 cells of each population were sorted into a well with 0.5 µl of iScript^{TM} RT-qPCR Sample Preparation Reagent (Bio-Rad, Hemel Hempstead, U.K.). Samples were spun down to allow the cells to mix with the lysis buffer at the bottom of the tube and 2.5 µl water added. A reverse transcription master mix of 0.5 µl 10X RT enzyme mix and 0.5 µl 10X RT primer mix was added to each sample. Cycling parameters were as follows: 25 °C for 10 min, 42 °C for 60 min, 85 °C for 5 min. A gene specific amplification master mix of 1.5 µl 5X dT amp master mix and 1 µl of PanCancer Pathways Panel primers (IDT, Integrated DNA technologies, Leuven, Belgium) was added to each sample. PanCancer Pathways specific primers contain 1512 pooled oligos with 0.25nm of each individual oligo in a mix at 0.5 µM in 500 µl of DTE buffer pH 7.5. Cycling parameters were as follows: 95 °C for 10 min, 8 cycles of 95 °C for 15 s and 60 °C for 4 min. Samples were then incubated at 95 °C for 2 min and then snap cooled on ice for 2 min prior to hybridisation.

A reaction mixture of the amplified PanCancer Pathways Panel samples was set up with 3 µl Reporter CodeSet, 5 µl hybridization buffer and 2 µl Capture ProbeSet and hybridized for 12-18 h at 65°C in a thermocycler. The samples were then loaded onto the PrepStation which added the mixture to a streptavidin-derivatized cartridge which bound with the biotinylated capture probe end, while washing away the unhybridised probes. An electrical field was then applied to align the complexes across the cartridge and the barcode containing reporter end was anchored was a second biotin-containing oligonucleotide. The cartridge was then placed into the Digital Analyzer and scanned

at 555 field of view (FOV). FOV is the area of the cartridge surface imaged by the Digital Analyzer, with 555 FOV being the most detailed scan. The Digital Analyzer counted each individual fluorescent barcode, with each barcode composed of seven spots made up of four colours, specific for each gene of interest. Data analysis was performed using the nSolver Software. The raw code count data was normalised, and background corrected using a standard curve of spike-in controls. The molecular counts were also normalised to internal controls and reference genes.

2.19 nCounter miRNA analysis

2.19.1 Bioanalyzer miRNA measurement

The miRNA quantity was measured using the Small RNA Analysis Kit (Agilent Technologies Ltd., Wokingham, UK). RNA samples were diluted with RNase-free water to 20 ng/µl and heat denatured with the RNA ladder for 2 min at 70 °C. According to manufacturer's instructions, 9 µl of a gel-dye mix was added to the Bioanalyser chip and using the chip priming station, the plunger was pressed down to exert pressure and held in place to spread the gel-dye mix across the chip. 9 µl RNA conditioning solution was added to the specified well and 5 µl RNA marker was added to all 11 sample wells and the ladder well. 1 µl of prepared ladder and 1 µl of sample were added to their specified wells. The chip was then vortexed and run on the Agilent 2100 Bioanalyzer instrument (Agilent Technologies Ltd., Wokingham, UK).

2.19.2 miRNA concentration

To concentrate miRNA in samples for nCounter analysis, the RNA Clean & Concentrator™-5 was used (Zymo Research, CA, U.S.A). According to manufacturer's guidelines, 100 µl RNA Binding Buffer was added to 50 µl of sample. 150 µl 100% ethanol was then added to the sample and mixed. The sample was passed through a Zymo-Spin™ IC Column by centrifugation for 30 seconds. The column was washed with 400 µl RNA Prep Buffer and centrifuged for 30 seconds, followed by 700 µl RNA Wash Buffer with a 30 second centrifugation and 400 µl RNA Wash Buffer with a 30 second centrifugation. The concentrated RNA was eluted with 15 µl DNase/RNase-free water.

2.19.3 nCounter miRNA assay

Mouse spleen RNA was screened for miRNA expression by use of the nCounter Mouse v1.5 miRNA Expression Assay Kit (NanoString Technologies®, Inc., Seattle, WA, USA) on the nCounter Analysis System (NanoString Technologies®, Inc., Seattle, WA, USA) according to the manufacturer's guidelines. For the miRNA sample preparation protocol, 3 µl of each sample, at 5 ng each, was added to an annealing master mix (13 µl Annealing Buffer, 26 µl nCounter miRNA Tag Reagent and 6.5 µl miRNA Assay Controls) in a 12 well 200 µl strip. Annealing reactions were performed on a Verti-96 well Thermal Cycler (Applied Biosystems, Foster City, USA), using the following conditions: 94 °C for 1 min, 65 °C for 2 min, 45 °C for 10 min and then brought to 48 °C. 2.5 µl of a ligation master mix (19.5 µl PEG and 13 µl Ligation Buffer) was then added to each tube and the strip returned to the thermocycler for 5 min. While left in the thermocycler at 48 °C, 1 µl of ligase was added directly to the wells and the ligase protocol initiated using the following conditions: 48 °C for 3 min, 47 °C for 3 min, 46 °C for 3 min, 45 °C for 5 min, 65 °C for 10 min and held at 4 °C. For the purification protocol, 1 µl Ligation Clean-Up Enzyme was added to each well and placing into the thermocycler with the following cycling conditions; 37 °C for 1 hour, 70 °C for 10 min and the samples were then cooled to 4 °C after which 40 µl RNase-free water was added to each sample. Before the hybridization step, the miRNA samples were denatured at 85 °C for 5 min and then quick-cooled on ice. A reaction mixture of a 5 µl aliquot of the ligated miRNA samples was set up with 10 µl Reporter CodeSet, 10 µl hybridization buffer and 5 µl Capture ProbeSet and hybridized for 12-18 h at 65 °C in a thermocycler and analysed by the nCounter system as previously described.

2.20 Statistical analysis

The deleterious impact of observed mutations to encoded amino acid in Kras protein in murine cases was checked using PredictSNP (Bendl et al. 2014) and PolyPhen-2 (Adzhubei et al. 2010) algorithms. PredictSNP use integrative evaluation across all well-known methods for amino acid changes investigation while PolyPhen-2 use the sequence conservation in human as a leading metric in evaluation. The reference Kras FASTA sequence was NP_067259.4."

Human mutations, identified by Haloplex DNA sequencing, were analyzed with a pipeline to evaluate different types of genetic modifications effects to protein function and human organism (Figure 1). The nonsynonymous SNP changes were investigated using PredictSNP. PredictSNP integrate results from several predictions algorithms giving compressed evaluation, thus it seems to be the most effective (Bendl et al. 2014). The sequence change in gene coding regions, which does not affect amino acid sequence change (synonymous effect), were investigated by PredictSNP2 (Bendl et al. 2016). Its results can be supported by PROVEAN (PROtein Variation Effect ANalyzer) (Choi and Chan 2015) and VEP (Variant Effect Predictor) (McLaren et al. 2016) algorithms. We suggest to first use PredictSNP2 as it integrates results from several other methods. Changes, which are located in non-coding regions, or those which cause protein termination but are not characterized as insertion or deletions (INDELS), were analyzed by PROVEAN and VEP. Finally, INDELS were evaluated by VEP algorithm, which was based only on biological rationality and no computational evaluation was performed. The sequences and mutations position were adjusted to each algorithm based on its reference genome. Analysis of mouse and human mutations by PolyPhen2 and PREDICT SNP algorithms was performed by Ms. Joanna Zyla, Silesian University of Technology, Poland.

Statistical analysis of the biological stat was performed using Minitab 17. P Values ≤ 0.05 were considered statistically significant. To test for significance, the data was first tested for normally using the Anderson-Darling test. Depending on whether the data was normally distributed, parametric (t-test) and non-parametric (Mann-Whitney) tests were used to test for significance.

nCounter data was analysed further using BRB-ArrayTools, an excel add-in, to identify significantly expressed miRNA and mRNA. BRB-ArrayTools was developed by Dr. Richard Simon and the BRB-ArrayTools Development Team.

3 Results

3.1 Human t-AML analysis

3.1.1 Introduction

Radiation-induced AML has been widely reported in numerous studies on the atomic bomb survivors of Hiroshima and Nagasaki (Ron et al. 1994), the Techa River cohort (Krestinina et al. 2013), and radiologist workers (Yoshinaga et al. 2004). Nowadays, cases of radiation-induced AML primarily result from radiotherapy due to treatment for a primary cancer, which is increasing due to an aging population, the increased use of radiotherapy and increased treatment success allowing patients to survive for longer and allowing time for AML to develop. T-AML is of particular concern as the onset is very quick with the relapse rate being 3 times higher in t-AML in comparison to de novo AML and the overall survival and complete remission rates are also much lower (Schoch et al. 2004). Cancer treatment programs mainly consist of a combined approach using both chemotherapy and radiotherapy, and most studies on human t-AML patients combine patients regardless of treatment type due to the varied range of chemotherapy agents and radiation doses. These large studies, although informative, do not allow investigation into the specific effects of radiotherapy alone.

AML can be classified by the WHO classification system using information regarding karyotype, morphology and, more recently, genetic mutations. The addition of genetic mutations into the system has allowed a better assessment of prognosis but there remain cases of AML with a normal karyotype and no detection of previously identified mutations (Dohner and Dohner 2008). Further genetic and epigenetic analysis would allow for the identification of new mutations for better risk stratification, prognosis and better understanding of the mechanisms of radiation-induced AML leukaemogenesis.

In order to fully investigate the development of radiation-induced AML we collected t-AML patient bone marrow samples from de novo AML, AML after radiotherapy, AML after chemotherapy and AML after a combination of radiotherapy and chemotherapy. This allowed us to study genetic and epigenetic modifications in radiation-induced AML patients independently and to also compare them to different treatment groups. DNA samples from AML patients were obtained for 20 patients (8 de novo AML, 8 AML after chemotherapy, 3 AML after radiotherapy and 1 secondary AML after MDS) from National Centre for Research "Demokritos" in Greece. It should be noted that the 3 cases with AML after radiotherapy for a previous malignancy from National Centre for Research "Demokritos" in Greece could also have had chemotherapy or

hormonotherapy but, unfortunately, this information is not available. Bone marrow aspirate samples were obtained from 6 AML patients from an existing biobank in the Northern Institute for Cancer Research (NICR), Newcastle (5 de novo AML, 1 AML after radiotherapy and chemotherapy). Bone marrow aspirate samples were also obtained from 5 normal donors from AllCells®. Details of all patient samples are provided in Table 1.

AML sub-type	ID	Primary cancer	Age (y)	Sex	Karyotype	BM blasts
De novo AML	1	N/A	11	F	45,XX,-7[20]/46,XX,-7,+?22[2]	
	2	N/A	86	M	45,X,-Y[22]/46,XY[2]	
	3	N/A	50	M	46,XY,inv(16)(p13q22)[23]/46,XY[2]	80%
	4	N/A	46	M	45,X,-Y,t(8;21)(q22;q22)[25]	30-40%
	5	N/A	29	F	46,XX,t(15;17)(q22;q21)[25]	90%
	6	N/A	56	M	47,XY,+8[4]/46,XY[21]	70%
	7	N/A	60	F	46,XX,inv(3)(q21q26)[17]/46,XX[2]	
	8	N/A	56	M	46,XY[25]	
	9	N/A	NA	N/A	No Cyto	
	10	N/A	NA	F	46XX,t(15;17)(q24;q21.2)[8]	
	11	N/A	NA	M	XY,t(3;12)(q26;p13),+6,+13,+22[10]	
	12	N/A	NA	N/A	No Cyto	

	13	N/A	NA	M	46,XY,t(11;19)(q23p13.3)MLLT1 fusion	
AML after chemotherapy	14	Unknown	73	M	46,XY[20]	
	15	CLL	55	M	46,XY[25]	
	16	MDS	81	F	45,XX,der(14;22)(q10;q10)[9]/44,XX,add(1)(p34~36),add(3)(q27),-5,add(13)(p11),del(13)(q22q34),-14,-16,-17,del(17)(q11.2q21),add(19)(q11),-20,+mar1,+mar2,+mar3[6]	
	17	CML	41	M	46,XY[25]	
	18	Colon cancer	73	M	46,XY,t(8;21)(q22;q22),del(9)(q13q22)[25]	30%
	19	MDS	74	F	47,XX,+del(22)(q11.2)[8]/47,XX,del(12(p11,2),+del(22)(q11.2)[10]/47,XX,del(12)(p11.2),del(20)(q11.2~13.1),+del(22)(q11.2)[5]	

	20	MDS	84	F	46,XX,del(5)(q13q33)[9]/46,XX,del(5)(q13q33),del(20)(q11.2)[4]/46,XX,del(5)(q13q33),-7[3]/46,XX[3]	
	21	MDS	56	M	46,XY[24]	
AML after radiotherapy	**22**	Breast	58	F	47,XX,+8,inv(9)(p12q13)[2]/46,XX,inv(9)(p12q13)[18]	
	23	Breast	70	F	46,XX,inv(9)(p12q13)x2,t(15;17)(q22;q21)[10]/46,XX,inv(9)(p12q13)x2[6]	
	24	Breast	69	F	45,XX,del(2)(p21),add(3)(q12),-4,del(7)(q22),+8,add(14)(q32),-16,-17,+mar1[17]/46,XX[3]	
AML after radiotherapy & chemotherapy	**25**	ATLL	53	M	78-88,inc[5]/46,XY[3]	
s-AML	**26**	MDS	76	F	46,XX[30]	8%
Normal donor	**27**	N/A	30	M		
	28	N/A	21	M		
	29	N/A	33	F		
	30	N/A	24	M		
	31	N/A	23	M		

Table 1. Age, gender, cytogenic details of 26 AML patients and 5 normal donors.

The group denoted "AML after radiotherapy" may have also had chemotherapy or hormonotherapy. CLL=Chronic lymphocytic leukaemia, MDS=Myelodysplastic syndrome, CML=Chronic Myelogenous Leukaemia, ATLL= Adult T-cell leukaemia/lymphoma, s-AML: secondary AML, NA=Not available.

3.1.2 Human AML Haloplex sequencing

DNA from 16 AML patients (6 de novo AML, 6 AML from chemotherapy, 3 AML from radiotherapy and 1 AML after MDS) were sent to Oxford Gene Technology for Haloplex sequencing, which targets known mutational hotspots in AML. The algorithms Polyphen2, PREDICT SNP, PREDICT SNP 2, PROVEAN and VEP were used on the sequencing results to predict the effect of the mutation on the protein, depending on what type of mutation was present. As depicted in Figure 12, single nucleotide changes involve use of the algorithms Polyphen2, PREDICT SNP or PREDICT SNP 2 while longer mutations involve use of PROVEAN or VEP algorithms.

All novel mutations with a predicted deleterious effect on the protein were listed as shown in Table 2. Mutations with a frequency of less than 5% were removed from the list as these mutations would not contribute to the progression of AML and were also difficult to confirm due to the sensitivity of the techniques. A total of 12 novel mutations were identified by Haloplex sequencing. These novel mutations were re-analysed by either Sanger sequencing or pyrosequencing for confirmation of their presence. Five of the novel mutations, consisting of three frameshift insertions and two point mutations, were confirmed in the four patients in the genes *DNMT3A*, *serine and arginine rich splicing factor 2* (*SRSF2)*, *TET2* and *RUNX1* (Figure 13). 5 of the novel mutations were not confirmed in the genes *SET binding protein 1 (SETBP1)*, *ASXL1*, *EZH2* and *SRSF2* (Figure 14). Clinical information of the patients with novel mutations are detailed in Table 3. All mutations, both previously known, and newly confirmed novel mutations are listed in Table 4. Novel mutations in the genes *DNMT3A*, *SRSF2*, *TET2* and *RUNX1* were identified, each predicting serious consequences to the protein function.
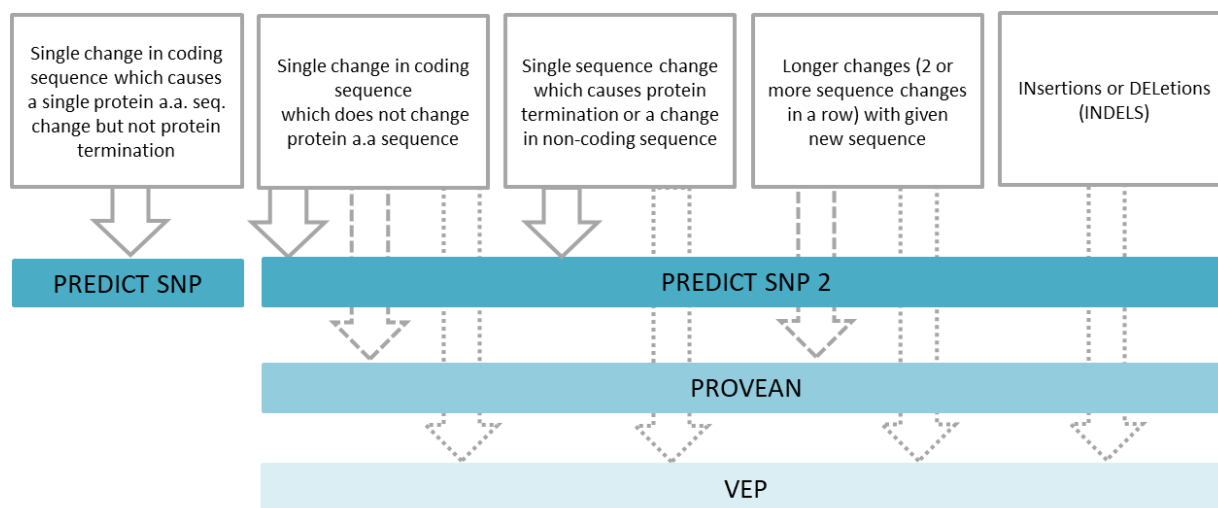
Figure 12. Algorithm flow chart.

For each type of mutation present, a sequence of different algorithms was used to predict whether the effect on the protein was deleterious or not. The colour of algorithm panel represents its effectives. The darker the colour, the more effective the evaluation.

ID	Gene	Ref->Seq	HGVSp	VarFreq	Algorithm
5	*ASXL1*	C->A	p.Gln623Lys	6.54	PREDICT SNP D:65%
22	*ASXL1*	C->A	p.Gln623Lys	21.59	PREDICT SNP D:65%
15	*DNMT3A*	AT->C	p.Ile715ProfsTer64	40.36	VEP High
6	*EZH2*	T->A	p.Phe290Leu	21.53	PREDICT SNP D:51%
8	*EZH2*	T->A	p.Phe290Leu	21.65	PREDICT SNP D:51%
14	*EZH2*	T->A	p.Phe290Leu	22.95	PREDICT SNP D:51%
15	*EZH2*	T->A	p.Phe290Leu	24.86	PREDICT SNP D:51%
17	*EZH2*	T->A	p.Phe290Leu	27.15	PREDICT SNP D:51%
23	*EZH2*	T->A	p.Phe290Leu	26.13	PREDICT SNP D:51%
4	*EZH2*	T->-	p.Phe290LeufsTer31	15.68	VEP High
26	*EZH2*	T->-	p.Phe290LeufsTer31	17.69	VEP High
6	*RUNX1*	-> C	p.Ala338ArgfsTer262	46.56	VEP High
26	*RUNX1*	-->GGCTGAGC	p.Leu144ArgfsTer4	21.96	VEP High
1	*SETBP1*	TCAGA->TCGGG	p.Glu1276Gly	98.97	VEP Moderate
3	*SETBP1*	TCAGA->TCGGG	p.Glu1276Gly	99.77	VEP Moderate
5	*SETBP1*	TCAGA->TCGGG	p.Glu1276Gly	99.84	VEP Moderate
6	*SETBP1*	TCAGA->TCGGG	p.Glu1276Gly	99.16	VEP Moderate
15	*SETBP1*	TCAGA->TCGGG	p.Glu1276Gly	98.92	VEP Moderate
20	*SETBP1*	TCAGA->TCGGG	p.Glu1276Gly	100	VEP Moderate
23	*SETBP1*	TCAGA->TCGGG	p.Glu1276Gly	99.15	VEP Moderate
24	*SETBP1*	TCAGA->TCGGG	p.Glu1276Gly	99.45	VEP Moderate
19	*SRSF2*	C->T	p.Pro96Leu	23.28	PREDICT SNP D:61%
4	*SRSF2*	GGGAC->TGGAT	p.Arg47Leu	79.17	VEP Moderate
4	*SRSF2*	GGGAC->GGGCT	p.Asp48Ala	79.17	VEP Moderate
21	*SRSF2*	AC->GT	p.Asp48Gly	71.79	VEP Moderate
21	*SRSF2*	AC->CT	p.Asp48Ala	71.79	VEP Moderate
22	*SRSF2*	GGGAC->TGGAT	p.Arg47Leu	66.67	VEP Moderate
22	*SRSF2*	GGGAC->GGGCT	p.Asp48Ala	66.67	VEP Moderate
6	*TET2*	C->T	p.Gln821Ter	94.9	PREDICT SNP 2 D:81%

Table 2. Novel mutations in human AML cases from DNA sequencing analysis.

Novel mutations identified in 6 de novo AML, 6 AML from chemotherapy, 3 AML from chemotherapy and 1 AML after MDS, detected using the Haloplex system. Mutated genes, frequency, codons affected, and algorithm predictions are listed. Algorithms predictions (Polyphen2, PREDICT SNP, PREDICT SNP 2, PROVEAN, VEP) are listed as being deleterious (D) or HIGH (high chance of the mutation being deleterious) with Ter (termination of protein) based on the type of algorithm used.
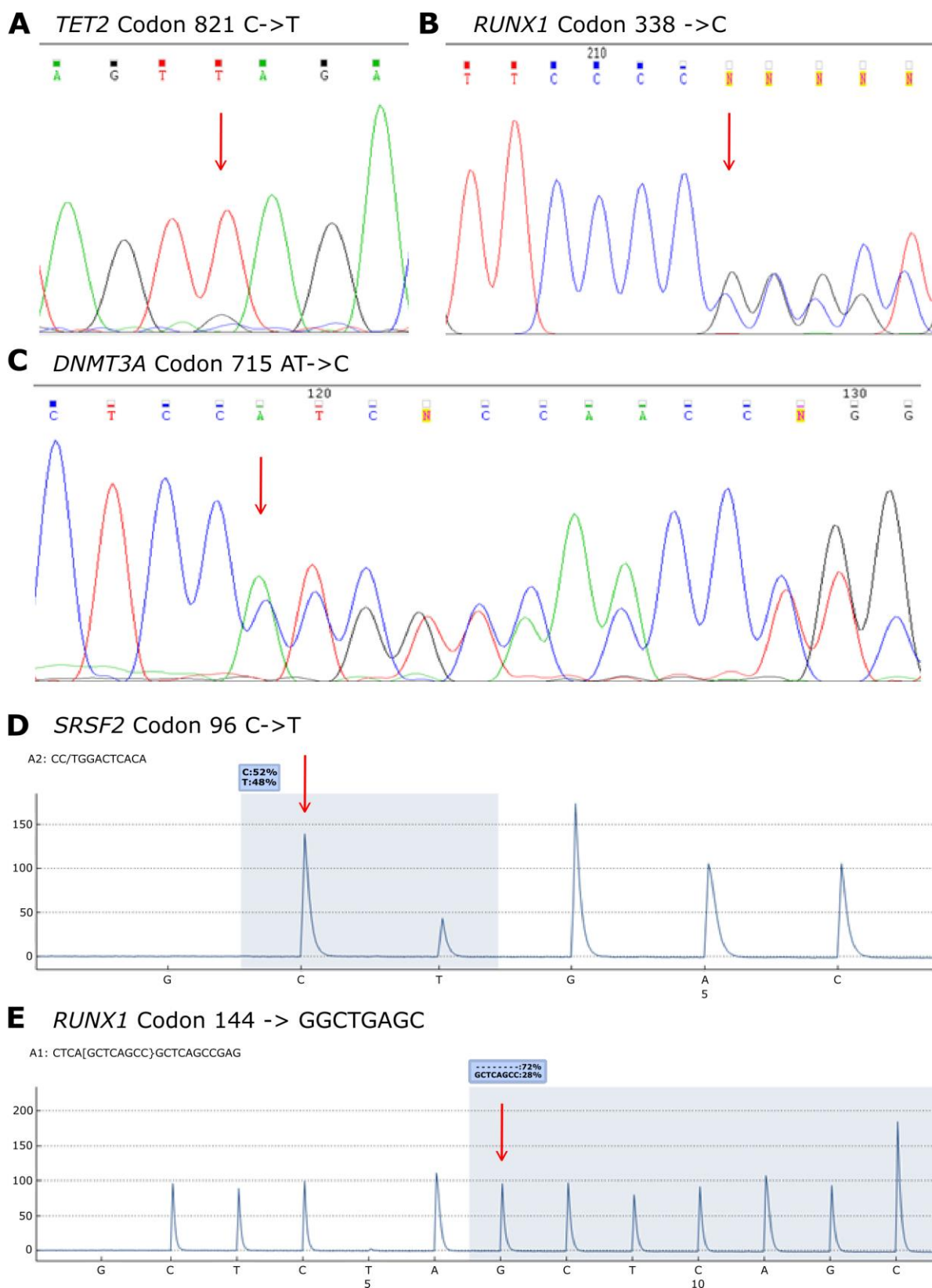
Figure 13. Confirmation of novel mutations.

(A) Sanger sequencing confirmation of the Gln821Ter (CAG→TAG) *TET2* mutation which results in termination of the protein. (B) Sanger sequencing confirmation of a

Ala338ArgfsTer262 mutation (C insertion) in *RUNX1* in a de novo AML patient which results in termination of the protein. (C) Sanger sequencing confirmation of a Ile715ProfsTer64 mutation (AT->C) in *DNMT3A* in a patient who developed AML after chemotherapy, which results in termination of the protein. (D) A Pro96Leu mutation (C->T) in *SRSF2* in a patient who developed AML after chemotherapy was confirmed by pyrosequencing showing a frequency of 48%. (E) A Leu144ArgfsTer4 (GGCTGAGC insertion) in *RUNX1* in a patient who developed AML after MDS was confirmed by pyrosequencing showing a frequency of 28%. An insertion of GCTCAGCC as reported on the reverse strand which corresponds to a GGCTGAGC insertion. Grey shaded regions for pyrosequencing results indicate the region where the presence of the wild-type and mutated change can be detected.

A *SETBP1* Codon 1276 TCAGA->TCGGG

B *SRSF2* Codon 47, 48 GGGAC->TGGAT/GGGCT

C *ASXL1* Codon 623 C->A

D *EZH2* Codon 290 T->A
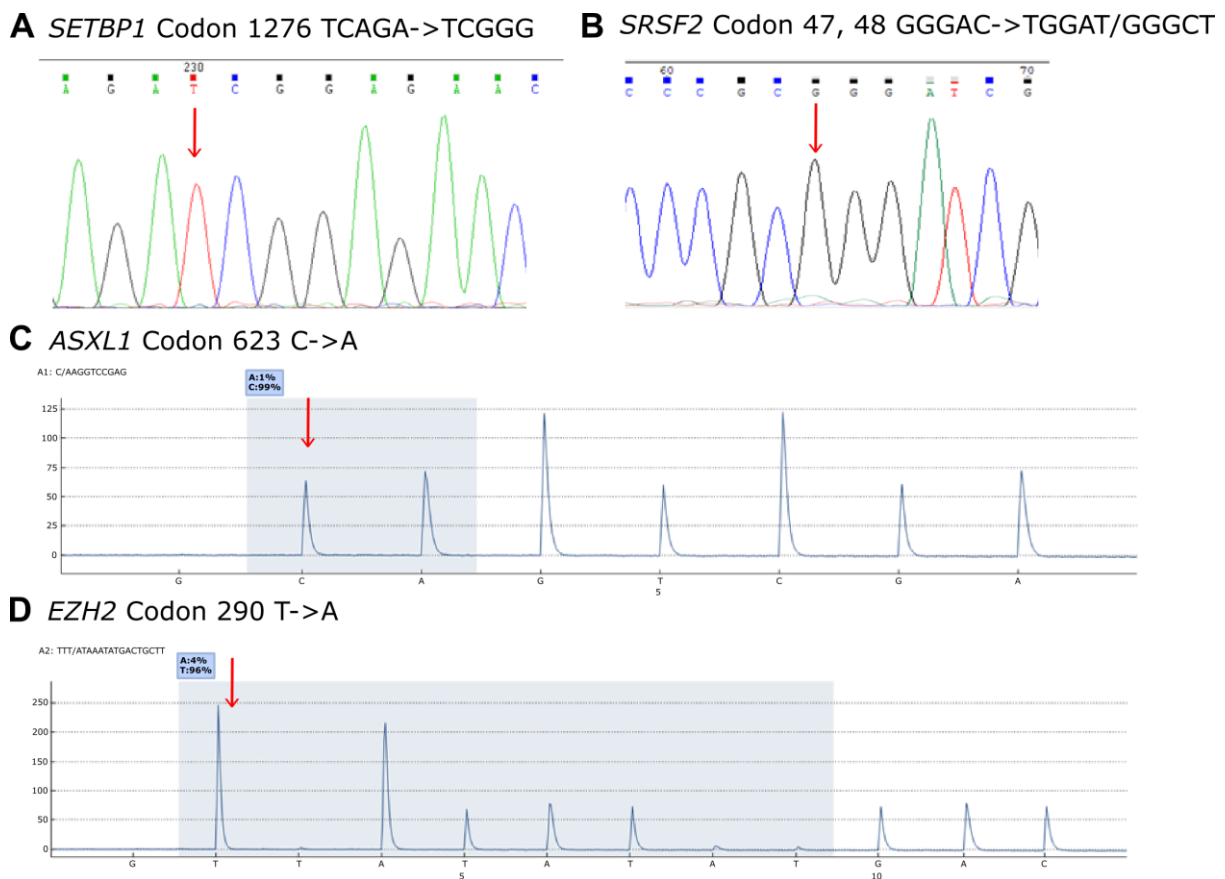
Figure 14. No confirmation of novel mutations.

(A) Sanger sequencing did not confirm a Glu1276Gly (TCAGA→TCGGG) mutation in *SETBP1* in patient 1. In these samples a SNP has been identified at codon 1275 with the sequence being TCG instead of TCA, which does not change the a.a. from serine.

(B) Sanger sequencing did not confirm a Arg47Leu (GGGAC→TGGAT), Asp48Ala (GGGAC→GGGCT/AC→GT/AC->CT) or Asp48Gly (GT→AC) mutation in *SRSF2* in patient 4. In these samples a SNP has been identified with the codon 48 being GAT instead of GAC, which does not change the a.a. from aspartate. (C) Pyrosequencing sequencing did not confirm a Gln623Lys (C→A) mutation in *ASXL1* in patient 5. (D) Pyrosequencing sequencing did not confirm a Phe290Leu (T→A) mutation in *EZH2* in patient 6. Grey shaded regions for pyrosequencing results indicate the region where the presence of the wild-type and mutated change can be detected.

	Patient 6	*Patient 15*	*Patient 19*	*Patient 26*
AML	De novo AML	AML after	AML after chemotherapy	s-AML
Previous disease	No	CLL	MDS	MDS
Age (y)	56	55	74	76
Gender	M	M	F	F
Karyotype	47,XY,+8[4]/46,XY[21]	46,XY[25]	47,XX,+del(22)(q11.2)[8]/47,XX,del(12(p11,2),+del(22)(q11.2)[10]/47,XX,del(12)(p11.2),del(20)(q11.2-13.1),+del(22)(q11.2)[5]	46,XX[30]
NPM1 mutation	Negative	Positive	Not available	Negative
FLT3 mutation	Negative	Negative	Not available	Negative
FAB diagnosis	M5	M5	Not available	M5
WBC (per mm3)	81860	Not available	Not available	32740
PLT (per mm3)	7000	Not available	Not available	198000
Hb (g/dL)	6.3	Not available	Not available	10.7
PB blasts %	24	Not available	Not available	26
BM blasts %	70	25	Not available	8
Treatment	Not available	Allogeneic bone marrow transplantation	Not available	Not available
Relapse	Yes	No	Not available	Yes
Survival	10 m / died of resistant disease	Still alive 84 m post diagnosis of AML	Not available	35 m / died of septic sock

Table 3. Clinical data of AML patients with novel mutations.

For *FLT3* mutations, samples were screened for *FLT3*-ITD (internal tandem duplication) and *FLT3*-TKD (tyrosine kinase domain) mutations. *NPM1* was screened for all mutations. WBC = white blood cell, PLT = platelet, Hb = haemoglobin, PB = peripheral blood, BM = bone marrow.

AML sub-type	ID	Mutations	
de novo AML	1	*TET2*	Pro363Leu (53%)
		TET2	Leu1721Trp (52%)
		SETBP1	His1100Arg (53%)
		CEBPA	Pro189del (9%)
	3	*FLT3*	Asp835Tyr (39%)
	4	*TET2*	Leu1721Trp (100%)
	5	*TET2*	Pro363Leu (50%)
		TET2	Leu1721Trp (49%)
	6	*TET2*	Gln821Ter (95%)
		RUNX1	Ala338ArgfsTer262 (47%)
		JAK2	Val617Phe (41%)
		ASXL1	Glu635ArgfsTer15 (18%)
	8	No mutation detected	
AML after chemotherapy	14	*TET2*	Pro29Arg (55%)
		TET2	Leu34Phe (49%)
		TET2	His1778Arg (46%)
	15	*TET2*	Leu1721Trp (48%)
		DNMT3A	Ile715ProfsTer64 (40%)
		IDH2	Arg140Gln (39%)
		NPM1	Trp288CysfsTer12 (20%)
	17	No mutation detected	
	19	*SRSF2*	Pro96Leu (23%)
	20	*TET2*	Leu1721Trp (53%)
		TET2	Pro363Leu (50%)
	21	*RUNX1*	Arg162Lys (73%)
		SRSF2	Pro95Arg (33%)
AML after radiotherapy	22	*DNMT3A*	Arg882His (43%)
		NPM1	Trp288CysfsTer12 (40%)
	23	*FLT3*	Asp835Val (15%)
	24	No mutation detected	
s-AML	26	*TET2*	Pro29Arg (50%)
		ASXL1	Gly646TrpfsTer12 (35%)
		RUNX1	Leu144ArgfsTer4 (22%)

Table 4. Total list of mutations in human AML patient samples.

All mutations, including previously known mutations and confirmed novel mutations in 6 de novo AML, 6 AML from chemotherapy, 3 AML from chemotherapy and 1 AML after MDS, detected using the Haloplex system. Mutated genes, codons affected and frequency are listed. Confirmed novel mutations are underlined.

3.1.3 Human AML *PU.1* SNPs

We investigated the presence of four SNPs in two regions of the URE, a distal regulatory unite (DRU) and proximal regulatory unit (PRU), of the gene *PU.1* by Sanger sequencing which have been linked to regulate *PU.1* expression in 23 human AML cases (Steidl et al. 2007; Bonadies, Pabst, and Mueller 2010). SNPs were detected in each AML patient sample (Figure 15) and normal donor (Figure 16). In each patient the SNPs detected were either heterozygous or homozygous, but no patient had these SNPs at every SNP site. The number of detected SNPs were similar across DRU, PRU 1, PRU 2 and PRU 3 sites with PRU 2 showing a slightly higher number of SNPs. Homozygosity was only detected in six AML patients for DRU, PRU 1 and PRU 2. These homozygous SNPs were in de novo AML, AML after chemotherapy and AML after radiotherapy patients. No homozygous SNPs were detected for PRU 3. In normal donors, a higher percentage of wild-type SNPs were reported, and homologous SNPs were reported at DRU 1 and PRU 2. The presence of these SNPs was detected in both human AML patient samples and normal donor samples, however, due to the lack of RNA, the effect of these SNPs on transcriptional expression could not be assessed.

DRU 1

TTA**T**AGC

TTA**N**AGC

TTA**G**AGC

43% WT (TT)

35% Het (GT)

22% Hom (GG)

PRU 1

TCT**G**ATT

TCT**N**ATT

TCT**A**ATT

43% WT (GG)

35% Het (AG)

22% Hom (AA)

PRU 2

GGC**C**GCG

GGC**N**GCG

GGC**T**GCG

39% WT (CC)

48% Het (CT)

13% Hom (TT)

PRU 3

CTC**G**GCA

CTC**N**GCA

70% WT (GG)

30% Het (GA)
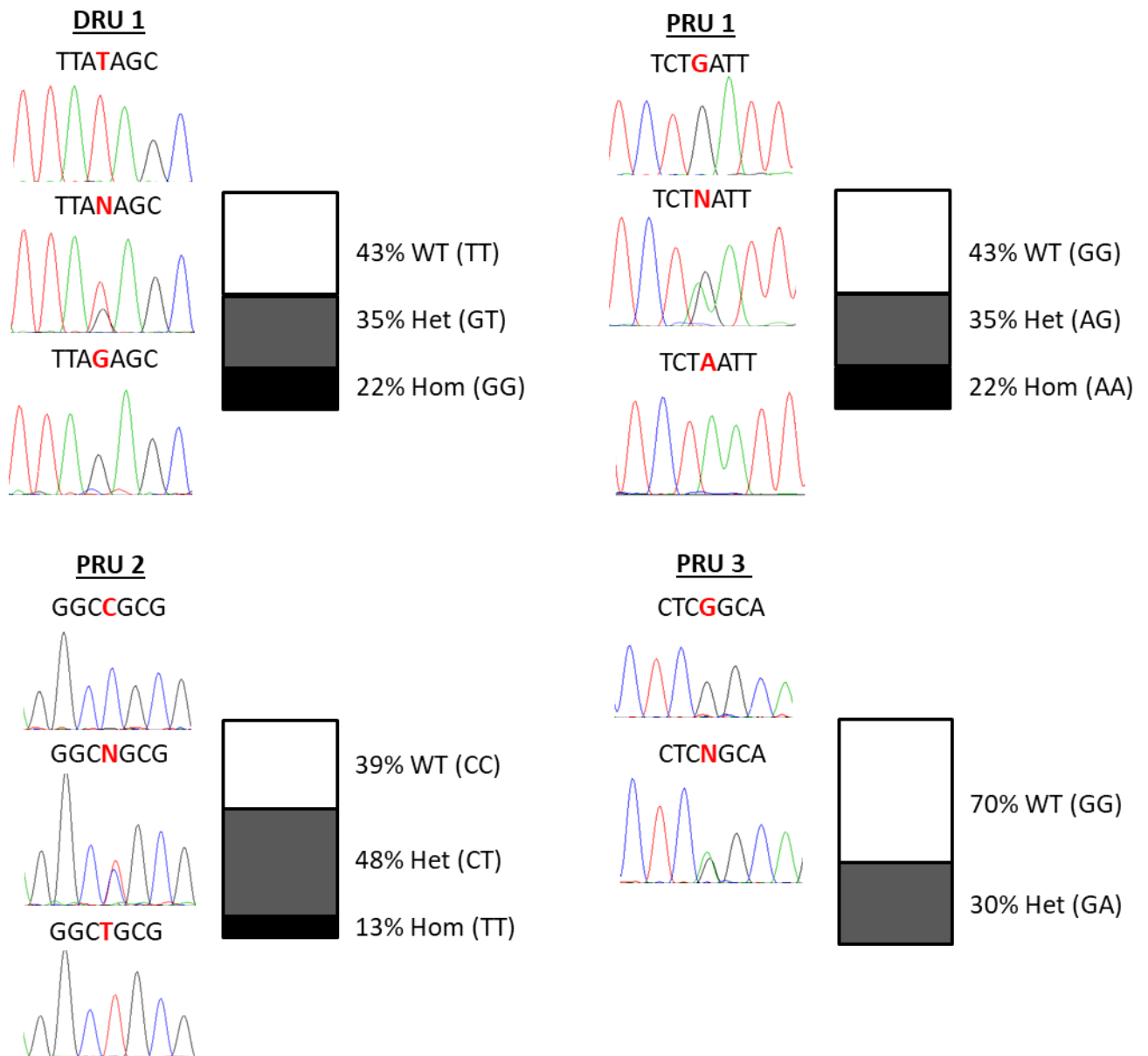
Figure 15. Sanger sequencing of *PU.1* SNPs in AML patients.

Sanger sequencing analysis of 4 SNPs located in the URE of *PU.1* in the bone marrow of 23 AML patient samples. SNPs were located in both the distal regulatory unit (DRU 1) and the proximal regulatory unit (PRU 1, 2, 3). Percentages of wild-type (WT), heterozygous (Het) and homozygous (Hom) SNPs are displayed.
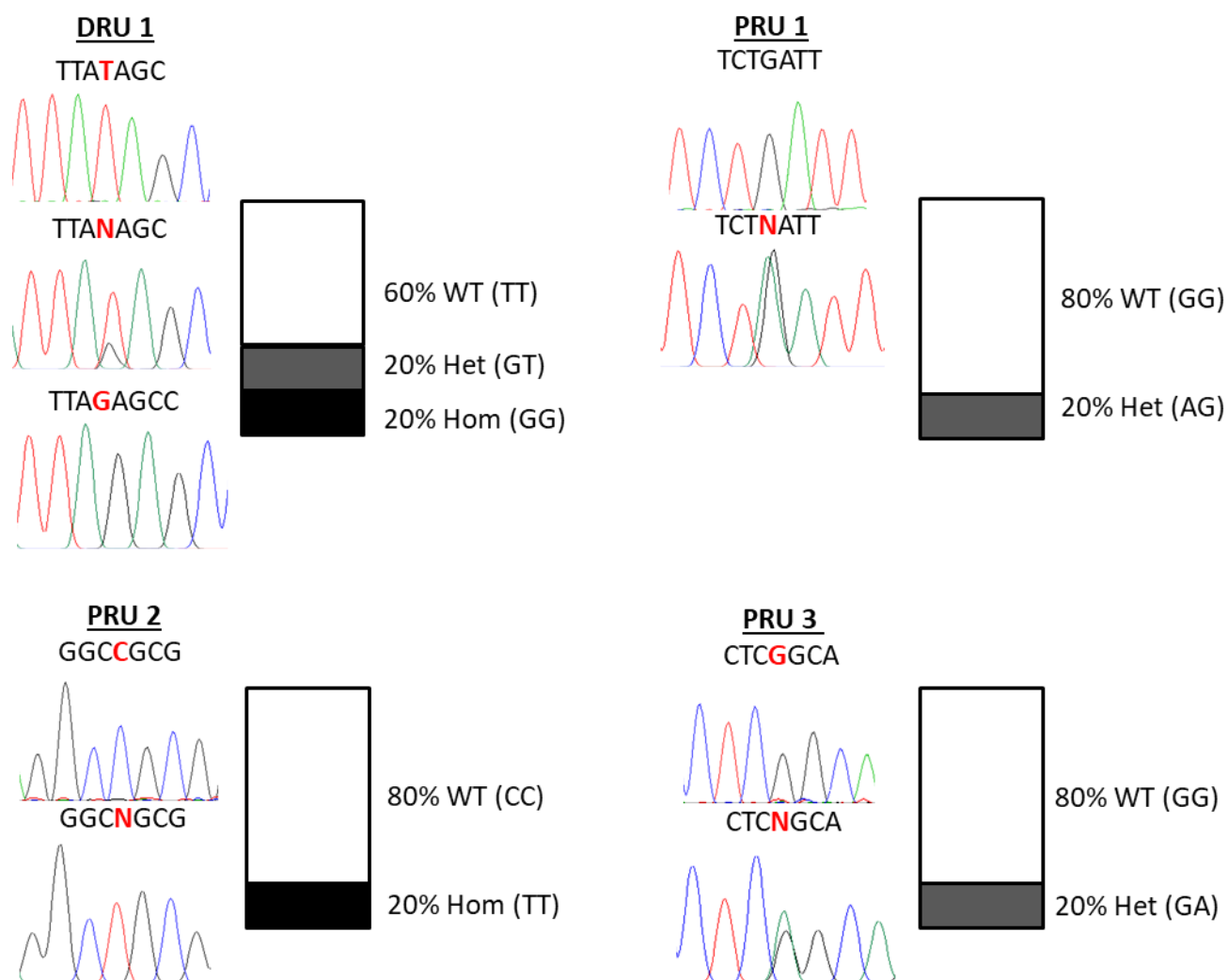
Figure 16. Sanger sequencing of *PU.1* SNPs in normal control donors.

Sanger sequencing analysis of 4 SNPs located in the URE of *PU.1* in the bone marrow of 5 control patient samples. SNPs were located in both the distal regulatory unit (DRU 1) and the proximal regulatory unit (PRU 1, 2, 3). Percentages of wild-type (WT), heterozygous (Het) and homozygous (Hom) SNPs are displayed.

3.1.4 Human AML *PU.1* promoter DNA methylation

DNA methylation analysis of the *PU.1* promoter in 5 normal bone marrow donors and 20 AML patients (10 de novo AML patient samples, 7 AML after chemotherapy patient samples, 2 AML after radiotherapy patient samples and 1 AML after MDS patient sample) was assessed by pyrosequencing. *PU.1* pyrosequencing primers were first designed to target the 4 CpG sites in the promoter previously shown to regulate *PU.1*

expression (Tatetsu et al. 2007). Validation of the primers by pyrosequencing with human methylation controls showed an average 2% level of DNA methylation in an unmethylated control and an average 82% level of DNA methylation in a methylated control (Figure 17).

AML patients had a slightly higher mean level of methylation for each CpG site, which was not significant, in comparison to normal donors. The range of methylation levels was much larger for all AML patients in comparison to normal donors with a portion of AML patients showing a greater than 60% methylation level (Figure 18). No significant difference in DNA methylation among the AML patient sub-groups (Figure 19) or among gender (Figure 20). This sub group with a high level of DNA methylation was composed of 3 male and 2 female AML patients of de novo AML, AML after chemotherapy and AML after radiotherapy groups. Age, however, may be a contributing factor as 5 out of the 6 samples with a > 59% methylation level were all aged 70 or older (Figure 21). For the 6th sample with a >59% methylation level, the age of the AML patient was unknown. In these AML patient samples, a sub-group of patients have a high level of DNA methylation of the *PU.1* promoter, indicating possible transcriptional repression.
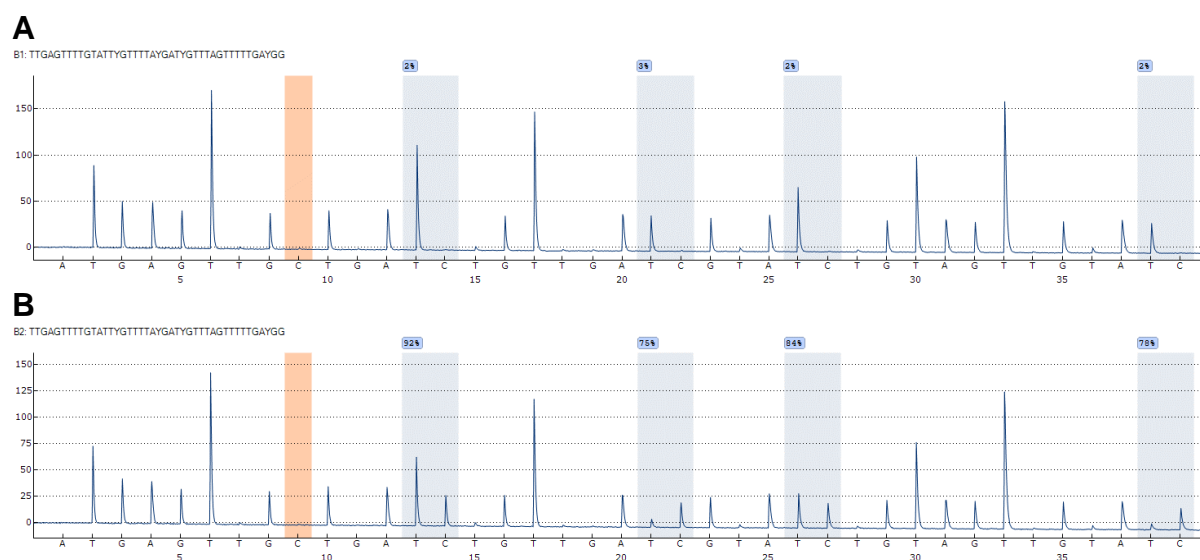
Figure 17. Pyrosequencing of *PU.1* CpGs.

Pyrosequencing pyrograms of CpG sites in the *PU.1* promoter in human bisulfite treated unmethylated negative control DNA (A) and methylated positive control DNA

(B). *PU.1* CpG regions are identified by a blue shading while a bisulfite control is identified by an orange shaded region. Percentage of methylation is indicated by a percentage above each CpG site.
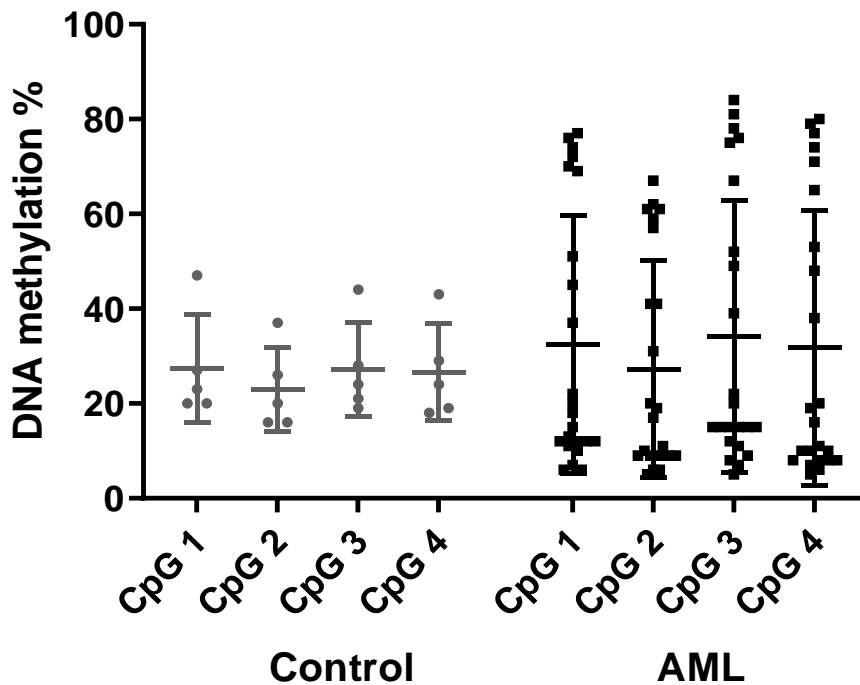
Figure 18. *PU.1* DNA methylation levels in control and AML samples.

DNA methylation was measured in bone marrow control samples from 5 control samples from normal donors and 20 AML patient samples by pyrosequencing at 4 CpG sites the promoter region of *PU.1*. Standard deviation and mean of the DNA methylation levels of each CpG are displayed.
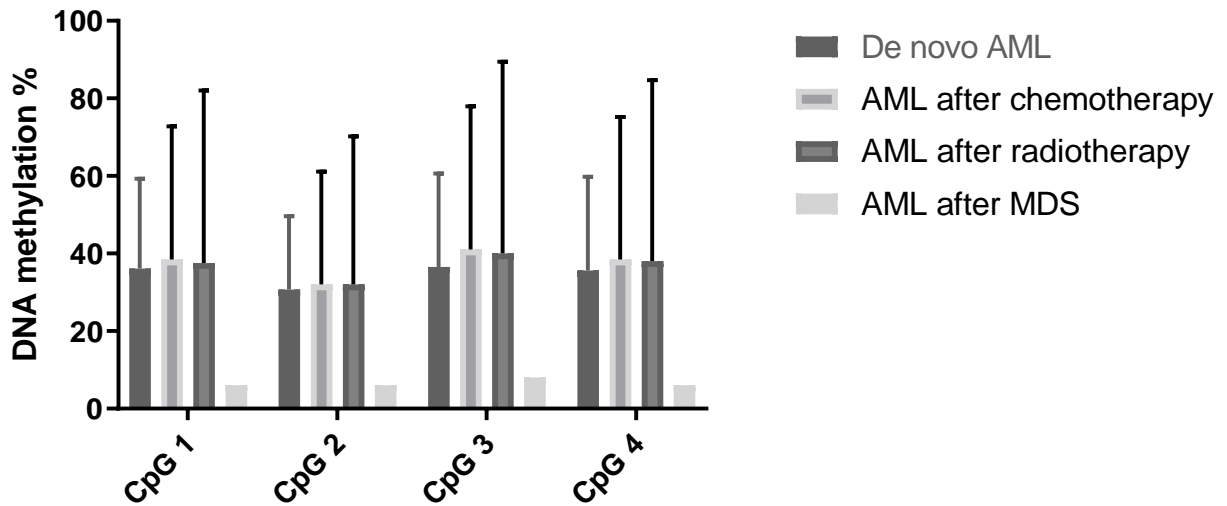
Figure 19. *PU.1* **DNA methylation levels grouped by treatment type.**

DNA methylation levels in bone marrow samples from 10 de novo AML patient samples, 7 AML after chemotherapy patient samples, 2 AML after radiotherapy patient samples and 1 AML after MDS patient sample. Standard deviation and mean of the DNA methylation levels of 4 CpG sites in the *PU.1* promoter is displayed.
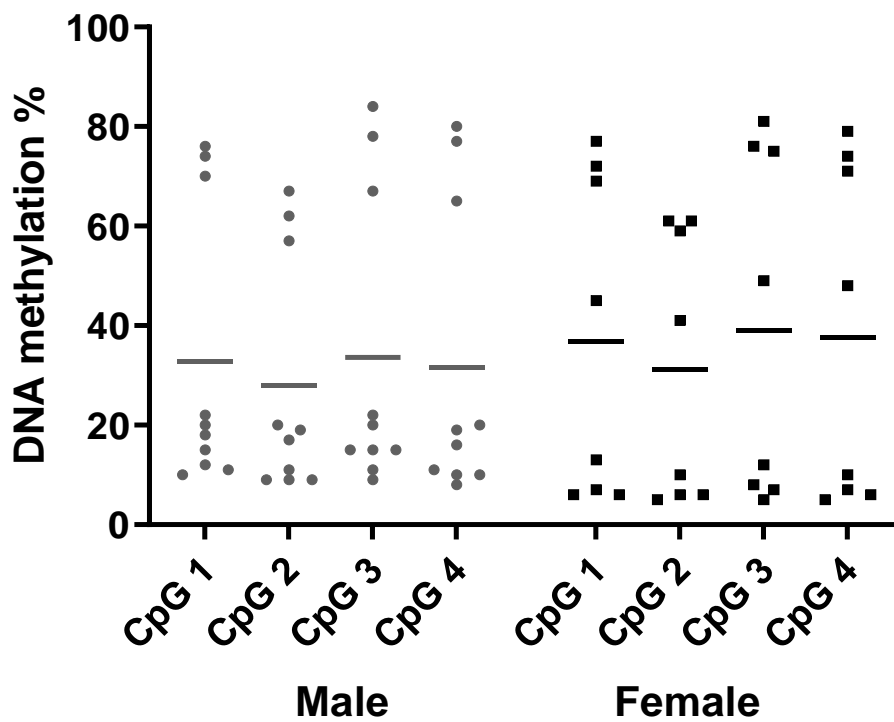
Figure 20. *PU.1* **DNA methylation levels grouped by gender.**

DNA methylation levels in bone marrow samples from 10 male and 8 female AML patient samples at the 4 CpG sites of the *PU.1* promoter. Mean of the DNA methylation levels of each CpG are displayed.
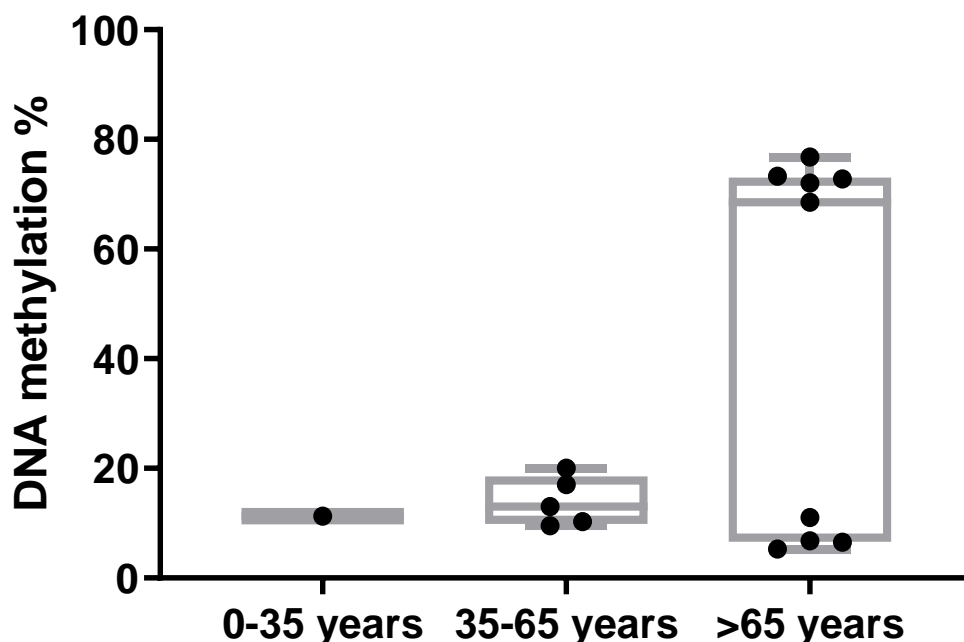
Figure 21. *PU.1* DNA methylation levels grouped by age.

DNA methylation levels in the *PU.1* promoter of AML patients separated in the age groups of 0-35 years old (n=1), age group 35-65 years (n=5) and the age group >65 years old (n=9).

3.1.5 Human AML *PU.1* transcriptional expression

Transcriptional expression was assessed in human AML patient samples where DNA methylation of the promoter was performed and where there was sufficient RNA for MQRT-PCR. Two AML patients (12, 13), which had no increase in DNA methylation, showed levels of *PU.1* expression similar to control bone marrow samples (Figure 22). *PU.1* expression was reduced in two AML patients (9, 11), one of which (patient 11) was reported in a previous figure to have a high level of *PU.1* promoter DNA methylation. This was a 3.5- fold reduction in patients 9 and 11 comparison to the mean of the control donor. This reduction is still a 2- fold reduction in transcription in

patients 9 and 11 in comparison to the lowest expressed control donor. Although more AML patient samples are required, this work indicates that *PU.1* expression could be lower in some AML patient samples.
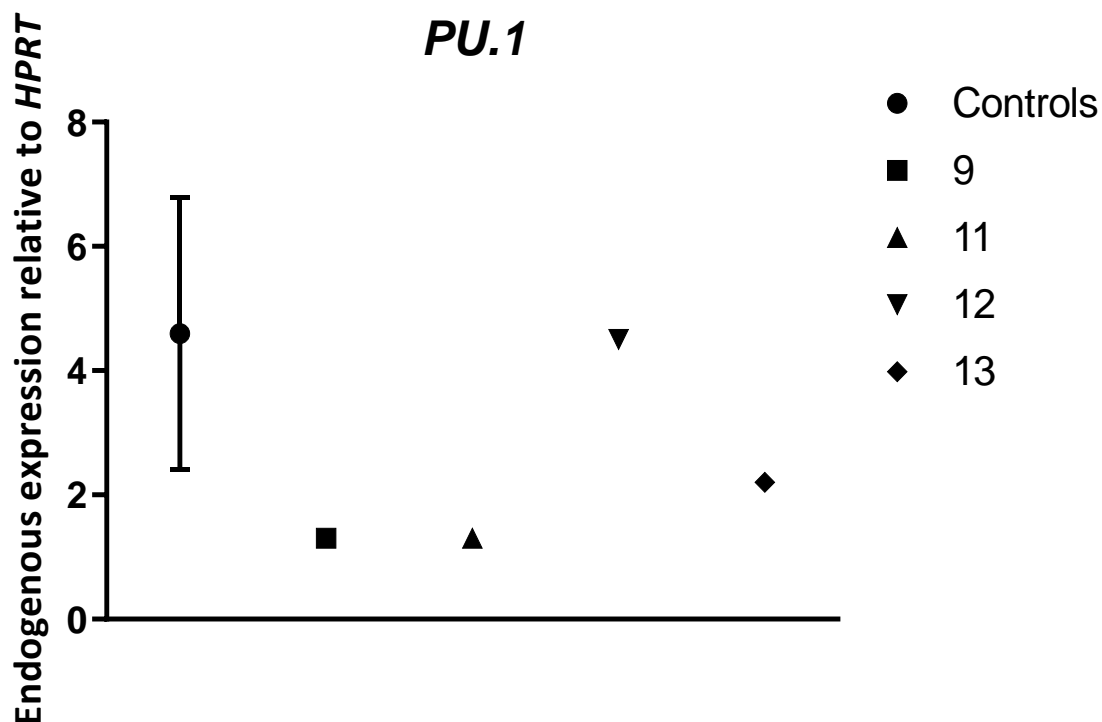
Figure 22. MQRT-PCR expression of *PU.1* in human AML.

PU.1 expression was analysed in 5 bone marrow donor control samples and 4 de novo AML patient samples (patient 9, patient 11, patient 12 and patient 13). Expression levels were normalised to *HPRT*. Error bars are displayed using the standard deviation.

Discussion and future directions

Overall this work identified novel AML mutations, investigated the presence of SNPs upstream of *PU.1/Sfpi1* and illustrated the epigenetic changes, particularly transcription expression and DNA methylation, in human AML cases, highlighting the role of *PU.1* in AML development.

The discovery of novel mutations is essential in the investigation of leukaemogenesis development and better understanding of the pathways of leukaemogenesis. Here we identified 5 novel mutations in the genes *TET2*, *DNMT3A*, *SRSF2* and *RUNX1*, all of which have been predicted to have a deleterious effect on the protein. *TET2* proteins function in DNA methylation by converting 5-methyl-cytosine to 5-hyroxymethylcytosine (5hmC) and may also result in DNA demethylation by removing the methyl group from the cytosine base (Feng et al. 2019). *TET2* mutations have been identified in a range of haematological malignancies such as AML, CML and MDS (Delhommeau et al. 2009). These mutations occur in 8-12% of de novo AML cases (Cancer Genome Atlas Research et al. 2013; Chou et al. 2011; Gaidzik et al. 2012). *TET2* mutations are located all along the gene, most of which are single nucleotide mutations, found to result in loss of function of generating 5hmC (Aslanyan et al. 2014). The novel Gln821Ter mutation found in patient 6 in our study was present in 95% of the bone marrow sample by Haloplex sequencing and, confirmed by Sanger sequencing, where a high mutational peak was observed (T) with a very small wild-type peak (G). This mutation was identified in a patient where 70% of bone marrow cells were blasts. This frequency suggests that *TET2* Gln821Ter probably occurred early in the process and is likely to be a driver mutation. The fact that this novel *TET2* mutation was found in 95% of the sample while the blasts of the patient was lower (70%) is in line with other studies which have reported that *TET2* mutations are associated with clonal haematopoiesis in healthy elderly people, are early events in leukaemogenesis and are frequently increased with age (Bullinger, Dohner, and Dohner 2017). This patient also had mutations in *RUNX1*, *janus kinase 2* (*JAK2)* and *ASXL1*. A pairwise mutation was therefore detected in patient 6 between *TET2* and *ASXL1*, which has been previously reported (Chou et al. 2011). This patient suffered a relapse and at 10 months died of resistant disease. The effect of a *TET2* mutation on survival from other studies is unclear. There are conflicting reports as in some studies it has been reported to significantly affect overall survival in AML patients (Aslanyan et al. 2014; Gaidzik et al. 2012), particularly in cases with an intermediate karyotype (Chou et al. 2011) while in other studies it has shown no significant difference in survival (Weissmann et al. 2012). In this case, interestingly the patient unfortunately died of resistant disease, possibly related to this new *TET2* mutation found in most if not all blast cells. This would need to be further investigated.

DNMT3A is a DNA methyltransferase responsible for the addition of a methyl group to cytosine, forming 5-methylcytosine which is essential for embryonic development (Okano et al. 1999). *DNMT3A* is commonly mutated in around 22% of human AML cases, with a lower frequency reported in Asian populations (Yamashita et al. 2010; Hou et al. 2012). *DNMT3A* mutations have been identified as preleukemic mutations that arise early in AML evolution and persist in time of remission (Medinger and Passweg 2017). Mutations are most frequently found in the methyltransferase (MTase) region of the gene, where R882 is located (Ley et al. 2010). The mutation R882 mutation is the most common mutation, accounting for 60% of all *DNMT3A* mutations (Ley et al. 2010; Hou et al. 2012). The R882H mutation was found to reduce the enzymatic activity of *DNMT3A* by 50% (Yamashita et al. 2010), and although no global DNA methylation changes could be detected (Ley et al. 2010), it is also associated with an up-regulation of *HOX* family members and *HOX* cofactor *MEIS1* (Yan et al. 2011; Ferreira et al. 2016). In this study two *DNMT3A* mutations were reported, a R882 mutation in patient 22 and a novel mutation in patient 15. In patient 15, this novel mutation in *DNMT3A* at codon 715 is located in the MTase region. It is present in 40% of the sample cells and is predicted to have a deleterious effect on the protein, as it results in its termination. Mutations in *DNMT3A* have been found to be associated with *FLT3*, *IDH1*, *IDH2* and *NPM1* (Ley et al. 2010; Hou et al. 2012; Cancer Genome Atlas Research et al. 2013). Importantly, this association is also seen in this study as patient 15 has *DNMT3A*, *TET2*, *IDH2* and *NPM1* mutations and patient 22 has *DNMT3A* and *NPM1* mutations. *DNMT3A* mutations, whether a R882 mutation or not, are associated with poor overall survival (Ley et al. 2010; Hou et al. 2012; Aslanyan et al. 2014). Interestingly, patients with *DNMT3A* mutations have been reported to respond well to the DNA methylation inhibitor decitabine with a recent study reporting 75% of *DNMT3A*-mutated patients achieving complete remission after treatment while 34% of wild type *DNMT3A* patients achieved complete remission, although the size of the study was small at just 46 patients (Metzeler et al. 2012). In this case however patient 15 was still alive at 84 months post diagnosis of AML. This surprisingly long survival result could be due to the patient also having an *NPM1* mutation which has been associated with favourable survival (Dohner et al. 2005).

The gene *SRSF2* is a splicing factor (Krainer, Conway, and Kozak 1990). *SRSF2* is most commonly mutated in chronic myelomonocytic leukaemia or MDS cases, rather

than AML cases (Yoshida et al. 2011). Spliceosome genes are found to be mutated in 5% of AML patients, with mutations in *SRSF2* found in 1% (Kihara et al. 2014; Yoshida et al. 2011). Frequent persistence of *SRSF2* mutations has been found in intensively treated AML patients in first complete remission (Rothenberg-Thurley et al. 2018). The most common point mutation is at codon Pro95 (Arbab Jafari et al. 2018), which is located in the RNA recognition motif forms a bond with RNA (Daubner et al. 2012). Two *SRSF2* mutations were reported in this study at codon Pro95 in patient 21 and a novel *SRSF2* mutation in patient 19. The novel mutation reported here in patient 19 in *SRSF2* occurs at Pro96 at a frequency of 23% with a predicted strong deleterious effect and so possibly interfering with RNA binding. Pyrosequencing analysis has found the mutation to occur at a higher level in 48% of the sample cells but unfortunately the number of blasts present in the bone marrow is not available for this patient. Of interest, both patients were diagnosed with AML following chemotherapy treatment for MDS; this is in line with the observation that *SRSF2* mutations are found especially in secondary AML patients (Bullinger, Dohner, and Dohner 2017). Unfortunately, we do not have further patient information on relapse and survival.

The next mutation found was in *RUNX1*, which is a transcription factor which activates genes involved in haematopoiesis. *RUNX1* is essential in early development with homozygous mutations in mice resulting in embryonic death (Okuda et al. 1996). *RUNX1* mutations have been frequently reported in MDS patients, and in AML cases. *RUNX1* mutations are included as a provisional category "AML with mutated *RUNX1*" in the 2017 recommendations from the European LeukemiaNet (Dohner et al. 2017). A higher frequency of *RUNX1* mutations was reported in cases which have developed from MDS (24%), rather than in de novo AML cases (9%) (Gaidzik et al. 2016). Two novel *RUNX1* mutations were found in patients 6 and 26 in this study and one previously reported *RUNX1* mutation in patient 21. In this study, the novel *RUNX1* mutation is present in exon 4, located in the Runt homology domain (RHD) which is responsible for DNA binding and one of the most commonly mutated exons in *RUNX1* (Gaidzik et al. 2016) and exon 8 in the transactivation domain (TAD) which binds with various growth factors, signalling molecules and transcription activators. The most common type of mutation is a frameshift mutation (Gaidzik et al. 2016; Schnittger et al. 2011), which we identified in 2 out of the 3 *RUNX1* mutations in this study. These novel mutations are also frameshift mutations, present in 22% and 47% of the samples

leading to termination of the protein. Pyrosequencing analysis of the *RUNX1* frameshift at codon 144 also found a similar frequency of the mutation at 28% of the sample cells. *RUNX1* mutations have shown a significant association with *ASXL1* mutations (Gaidzik et al. 2016), which is confirmed in both patient 6 and patient 26 in this study. Mutual exclusivity was also seen in patients that had *RUNX1* mutations and those that had *FLT3* and *NPM1* mutations, which has also been previously reported (Cancer Genome Atlas Research et al. 2013). Overall survival is significantly lower in patients with a *RUNX1* mutation (Gaidzik et al. 2016; Schnittger et al. 2011; Jalili et al. 2018) and in this study patients 6 and 26 with novel *RUNX1* mutations relapsed and died of resistant disease at 10 months and 35 months, respectively.

When looking at the mutations across all samples, the most commonly mutated gene was *TET2* which was mutated in 8 out of 16 patients screened and 3 of these patients had two *TET2* mutations, while one patient had three *TET2* mutations. *TET2* mutations were also present in a large proportion of the sample, usually 50% or more, which again illustrates its occurrence in early leukaemogenesis. As has been previously reported in the literature, mutations in this study involving genes functioning in DNA methylation (*DNMT3A*, *TET2*, *IDH2*) myeloid transcription factors (*RUNX1*, *CEBPA*), chromatin modifiers (*ASXL1*) and activated signalling (JAK2, *FLT3*) co-occurred across all AML sub-types. When comparing AML sub-types, AML patients after radiotherapy had the least mutations with a *DNMT3A* and *NPM1* mutation for patient 22, an *FLT3* mutation for patient 23 and no mutation for patient 24. Overall, however, no distinguishable feature could be detected which was specific for a particular AML sub-group. All AML sub-groups had varying amounts of chromosomal aberrations, point mutations, frame-shifts, *PU.1* polymorphisms and high *PU.1* promoter DNA methylation levels. This is likely due to the small number of samples in each sub-group. Also, it is important to note that after sequencing analysis, there were still 3 patients (patient 8, patient 17, patient 24) with no mutations detected and two of these patients had a normal karyotype. This highlights the need for further investigation into identifying mutations of interest.

Although *Sfpi.1* is commonly mutated in the CBA/Ca mouse model, in humans the gene, known as *PU.1*, is rarely mutated. A study by Mueller et al. (Mueller et al. 2002) reported *PU.1* mutations present in 7% of 126 cases of AML patients. This high occurrence however, has not been repeated with studies involving 77 AML cases

(Lamandin et al. 2002) and 60 AML cases (Vegesna et al. 2002) reporting no *PU.1* coding region mutations, this discrepancy is possibly due to the Japanese cohort used by Mueller et al. A study by Dohner et al involving 112 AML patients did find mutations in 2 AML patients, illustrating that *PU.1* mutations are not a common mechanism for AML induction. Other methods of *PU.1* disruption have since been investigated. Polymorphisms in the URE have been identified which leads to reduced *PU.1* expression with a deletion in this URE region identified in a remission case of an AML patient (Steidl et al. 2007; Bonadies, Pabst, and Mueller 2010). One of these SNPs (DRU) has been shown to reduce the enhancer activity of the URE and interfere with the binding of the *PU.1* transcription regulator, *special AT-rich sequence binding protein 1 (SATB1).*

In this study, these polymorphisms were detected in human AML patient samples at each SNP site. Unlike previous studies however, homozygosity for these SNPs was not detected for all SNPs in any patient (Bonadies, Pabst, and Mueller 2010). This is most likely due to the very low numbers of patient samples in this study. No homologous PRU 3 SNP was detected in any AML patient but it has been previously reported to occur at a low frequency (<5%) (Steidl et al. 2007). These SNPs, whether heterozygous or homozygous, also occur in normal human control samples and further investigation into the effect of these SNPs show that the SNP DRU1 reduces *PU.1* expression in specific cell types, such as GMPs and MEPS, rather than stem cells, and they appear to contribute to AML development rather than having a role in its initiation (Steidl et al. 2007). The data obtained on SNP analysis was not linked to transcriptional expression due to the fact that RNA was not available from the same patients. Although transcriptional analysis could not be performed and compared to SNPs, the potential effect of this SNP data on transcriptional changes could be further investigated. The URE and promoter region are locations where transcription factors bind to regulate gene expression and these SNPs could possibly occur in one of these binding sites. A number of transcription factors such as GATA-1 and RUNX1 have been reported as PU.1 transcription factors (Gupta et al. 2009). Future work could analyse the DNA sequence of this region for transcription factor recognition sequences. Transcription factors are commonly reported as mutated in AML cases, such as *RUNX1* (Cancer Genome Atlas Research et al. 2013), therefore detection of

transcription factor binding sites among the SNPs could indicate that their effect is lessened as a result.

DNA methylation levels among AML patients showed a range of methylation levels, which can be separated into those with a high or with a low *PU.1* methylation level. The difference in DNA methylation between these two groups was significant (Mann Whitney, p=0.019). This high level of *PU.1* promoter methylation was not due to classification of AML (de novo, AML after chemotherapy or AML after radiotherapy) or gender. Analysis of the age of the AML patients revealed that the patient samples with a high level of DNA methylation were all aged > 65 years. Within the > 65 year age group, there appears to be 2 sub-groups, one with a high level of DNA methylation and another with a low level of DNA methylation. Previous work has shown that transcriptional *PU.1* levels in 87 AML patients could also distinguish patients with a high *PU.1* methylation level from those with a low *PU.1* methylation level (Will et al. 2015). Gene expression profiles of those with a lower *PU.1* expression level sharing more similarly dysregulated genes to a mouse model $URE^{het}Msh^{-/-}$ designed to represent a phenotype representative of aging HSCs (Will et al. 2015), although the $URE^{het}Msh^{-/-}$ mouse model resulted in development of MDS-like disorder rather than AML.

In this study, the numbers of samples in these groups, however, are limited and so this work needs to be confirmed in a larger cohort. Unfortunately, for many of the AML samples, only DNA was available. For the samples where RNA was available, patient 11 showed a high level of *PU.1* methylation. Transcriptional analysis also showed a reduction of *PU.1* expression in patient 11 in comparison to the normal donor samples. This separation of AML patients with high or low *PU.1* DNA methylation levels could perhaps represent different pathways through which AML developed which, in cases with a higher DNA methylation level, involved repression of *PU.1* transcription with preference in an older age group. The high level of DNA methylation in these patients could be due to upstream mutations in epigenetic regulators, resulting in aberrant DNA methylation. DNA sequencing analysis was only performed in 2 patient samples with high DNA methylation levels (patient 20 and 23) and these patients also had mutations in the genes *FLT3* and *TET2* (Rasmussen and Helin 2016). Mutated or loss of *TET2* has been reported to result in decreased 5-hmC levels in the DNA of myeloid cancer patients (Ko et al. 2010) and hypermethylation of enhancers leading to

leukaemogenesis (Rasmussen et al. 2015). Therefore, in patient 20, mutation of *TET2* could have resulted in the hypermethylation of the *PU.1* URE due to low 5-hmC levels with reduction of *PU.1* transcriptional expression could possibly be the mechanism of blocking myeloid development in these samples in patient 20. The exact effect of *TET2* mutations, however, is currently not well understood. The *FLT3* Asp835 mutation in patient 23 is located within the kinase domain of the *FLT3* receptor and results in constitutively activation of the kinase causing proliferation of cells (Yamamoto et al. 2001). In this patient the high DNA methylation of the *PU.1* promoter could have produced leukemic blast cells with enhanced proliferation by the *FLT3* Asp835 mutation. A reduction of *PU.1* expression was also seen in a patient sample (patient 9) which had a low level of DNA methylation. This decrease in transcription does not appear to be caused by DNA methylation and may have another, as yet unidentified, mechanism of reducing transcription. It may be that the reduction of *PU.1* expression is caused by *miR-155*, a known negative regulator of *PU.1* (Vigorito et al. 2007) which has been previously reported to be upregulated in AML patient samples (Salemi et al. 2015). Even though this increase in *PU.1* promoter DNA methylation and decrease in *PU.1* transcription can only be demonstrated in one patient sample due to lack of RNA material, this illustrates the significance of *PU.1* in human AML development and the need to study this gene in more detail in human AML cases.

3.2 Mouse rAML sequencing and gene expression analysis

3.2.1 Introduction

The CBA/Ca mouse serves as an invaluable source of information in the investigation of development of radiation-induced AML. There are close histopathological similarities between humans and mice in the development of AML (Major 1979), however, there have been questions about the use of the mouse for investigation into human AML due to the well-established presence of the *Sfpi1* mutation in the mouse which has not been often reported in human cases of t-AML (Suraweera et al. 2005). A *PU.1* R235 mutation is a rare event in human AML with many studies unable to detect the presence of any *PU.1* coding mutations (Lamandin et al. 2002; Vegesna et al. 2002).

However, recent studies have reported a reduced expression in *PU.1* levels clearly evident in de novo AML cases (Basova et al. 2014; Steidl et al. 2006). Development of AML through the dysregulation of *PU.1* seems to be a dominant pathway in murine cases which needs further investigation in human cases (Verbiest et al. 2015). Here we aim to use both mouse and human AML samples induced by radiation and investigate an inter-species comparison.

Historical rAML samples available in the laboratory from different studies were combined. They are composed of spleen samples from CBA/H and F1 CBA/H x C57BL/Lia mice which were exposed to either X rays or neutrons but all of which were diagnosed with radiation-induced AML. The CBA mouse model is used to study radiation leukaemogenesis mechanisms with chromosome 2 *Sfpi1* deletion and point mutation already identified as driving events during AML development. In this study, our aim was to screen a total of 123 historical radiation-induced AML spleen samples for genes commonly mutated in human AML cases, to identify both genetic and epigenetic changes in the development of leukaemogenesis, aiming to better characterise the molecular mechanisms of rAML induction and to further establish if the CBA mouse is a good model of radiation-induced leukaemogenesis.

3.2.2 DNA mutations

In order to compare human AML with mouse AML, a literature review identified genes of interest which are well-known to contain hot spots of mutation in human AML such as *DNMT3A, IDH1*, *IDH2*, *FLT3*, *NPM1, KRAS*, *NRAS, CKIT* and *CEBPA* (Cancer

Genome Atlas Research et al. 2013; Fernandez-Mercado et al. 2012; Fried et al. 2012; Lindsley et al. 2015) (Table 5). The presence of a mutation in *Sfpi1* at codon R235 was included due to its common occurrence in the CBA mouse. PCR primers were designed to amplify the identified affected region for each gene in the mouse. A total of 123 historical CBA/H mouse rAML samples were sequenced by Sanger sequencing for the detection of these common mutations. CGH analysis was also performed on 116 samples to identify chromosome 2 deletions.

The most common alternation to occur in the rAML cases was the chromosome 2 interstitial deletion, always including *Sfpi1* in 94 samples (Table 6), of which 79 have been previously reported (Brown et al. 2015). A total of 3 genes were found to be mutated in the CBA/H spleen samples; mutations at codon R235 in the myeloid transcription factor *Sfpi1/PU.1*, insertions at exon 14 in the receptor tyrosine kinase *Flt3* and mutations at codon G12 in the signalling factor *Kras* with accompanying CGH data for chromosome 2. No mutations were found in the remaining locations sequenced; *Npm1* exon 12 mutation, *Idh1* R132, *Idh2* R140, *Dnmt3a* R882, *Nras* G12, *C Kit* exon 17 and *Cebpa* exon 1 frame shift. Full details for each murine case are detailed in Table 7.

Gene	Mutation	Reference
FLT3	Exon 14	Fernandez-Mercado et al. 2012
DNMT3A	R882	Fried et al.2012
IDH1	R132	Ley et al. 2013
IDH2	R140	Ley et al. 2013
NRAS	G12	Fernandez-Mercado et al. 2012
KRAS	G12	Fernandez-Mercado et al. 2012
C-KIT	Exon 17	Lindsley et al. 2015
CEBPA	Exon 1	Lindsley et al. 2015
NPM1	Exon 12	Fernandez-Mercado et al. 2012

Table 5. Commonly mutated human AML genes.

Genes were identified by a literature search and references of the mutations listed.

Sfpi1/Pu.1 R235	Amino Acid Change	No. of Samples (%)	Chr 2 Del
CGC>TGC	Arg-->Cys	40 (32)	Del 36, Retained 2, Unknown 2
CGC>CAC	Arg-->His	29 (24)	Del 26, Retained 2, Unknown 1
CGC>AGC	Arg-->Ser	11 (9)	Del 10, Unknown 1
CGC>CTC	Arg-->Leu	3 (2)	Del 3, Retained 0
CTGCGC>CTATGC	Leu, Arg->Leu, Cys	1 (1)	Unknown 1
None (CGC)	None	39 (32)	Del 19, Retained 18, Unknown 2
Flt3 Exon 14	**Insertion**	**No. of Samples (%)**	
Insertions	DFYVDFKDY* HFYVDFRDYEY* VKMLKE** FYVDFRDY NFRDYEYDLKW	5 (4)	Retained 5
No insertion	-	119 (96)	Del 95, Retained 17, Unknown 7
Kras G12	**Amino Acid Change**	**No. of Samples (%)**	
GGT>GAT	Gly>Asp	2 (2)	Deletion 2
GGT>CGT	Gly>Arg	1 (1)	Retained 1
None (GGT)	None	120 (97)	Del 93, Retained 20, Unknown 7

Table 6. Mutations in *Sfpi1* R235, *Flt3*-ITD and *Kras* G12 in murine rAML samples.

List of the mutations and associated protein changes found by Sanger sequencing in the DNA of 123 mice diagnosed with rAML, the number of mice the mutations were found to be present in and also including chromosome 2 deletion information. *Previously published in Finnon et al. 2012. ** Previously published in Finnon et al. 2012 and only *Flt3*-ITD information included for this sample. Percentages for *Flt3*-ITD are calculated from 124 mice.

Strain	Dose	Gender	Chr 2	*Sfpi1*	*Flt3*	*Kras*	Diagnosis
CBA[gfp/gpf]	3 Gy Xrays	F	No Del	R235	No ITD	G12	10 Months
	3 Gy Xrays	F	No Del	R235	ITD	G12	13 Months
	3 Gy Xrays	F	No Del	R235	ITD	G12	13 Months
	3 Gy Xrays	F	No Del	R235	No ITD	G12	23 Months
	3 Gy Xrays	F	Del	R235	No ITD	G12	14 Months
	3 Gy Xrays	M	Del	R235H	No ITD	G12	15 Months

	3 Gy Xrays	M	Del	R235H	No ITD	G12	17 Months
	3 Gy Xrays	M	Del	R235C	No ITD	G12	19 Months
	3 Gy Xrays	M	Del	R235H	No ITD	G12	20 Months
	3 Gy Xrays	M	Del	R235H	No ITD	G12	14 Months
	3 Gy Xrays	M	Del	R235S	No ITD	G12	15 Months
	3 Gy Xrays	M	Del	R235C	No ITD	G12	15 Months
	3 Gy Xrays	M	Del	R235	No ITD	G12	14 Months
	3 Gy Xrays	M	Del	R235H	No ITD	G12	19 Months
	3 Gy Xrays	M	Del	R235H	No ITD	G12	9 Months
	3 Gy Xrays	M	Del	R235C	No ITD	G12	16 Months
	3 Gy Xrays	M	Del	R235H	No ITD	G12	16 Months
	3 Gy Xrays	M	Del	R235	No ITD	G12	11 Months
	3 Gy Xrays	M	Del	R235	No ITD	G12	13 Months
	3 Gy Xrays	M	No Del	R235C	No ITD	G12	22 Months
	3 Gy Xrays	M	Del	R235	No ITD	G12	15 Months
CBA/H	3 Gy Xrays	M	Del	R235C	No ITD	G12	14 Months
	3 Gy Xrays	M	Del	R235H	No ITD	G12	15 Months
	3 Gy Xrays	M	Del	R235C	No ITD	G12	15 Months
	3 Gy Xrays	M	Del	R235L	No ITD	G12	16 Months
	3 Gy Xrays	M	Del	R235H	No ITD	G12	16 Months
	3 Gy Xrays	M	Del	R235H	No ITD	G12	18 Months
	3 Gy Xrays	M	Del	R235C	No ITD	G12	19 Months
	3 Gy Xrays	M	Del	R235C	No ITD	G12	15 Months
	3 Gy Xrays	M	Del	R235C	No ITD	G12D (Case 1)	22 Months
	4.5 Gy Xray	M	Unknown	R235	No ITD	G12	12 Months
	4.5 Gy Xray	M	No Del	R235	No ITD	G12	13 Months
	4.5 Gy Xray	M	No Del	R235	No ITD	G12	14 Months
CBA/H	1 Gy N	M	Unknown	R235C	No ITD	G12	Unknown
	1 Gy N	M	Del	R235C	No ITD	G12	Unknown
	1 Gy N	M	Del	R235H	No ITD	G12	Unknown
	1 Gy N	F	Del	R235	No ITD	G12	Unknown
	1 Gy N	M	Del	R235H	No ITD	G12	Unknown

1 Gy N	F	Del	R235H	No ITD	G12	Unknown
1 Gy N	M	Del	R235H	No ITD	G12	Unknown
1 Gy N	M	Del	R235S	No ITD	G12	Unknown
1 Gy N	M	Del	R235	No ITD	G12	Unknown
1 Gy N	F	Del	R235C	No ITD	G12	Unknown
1 Gy N	Unknown	No Del	R235C	No ITD	G12	Unknown
0.5 Gy N	M	Del	R235	No ITD	G12	Unknown
0.5 Gy N	Unknown	Del	R235	No ITD	G12	Unknown
0.5 Gy N	F	Del	R235H	No ITD	G12	Unknown
0.5 Gy N	Unknown	No Del	R235	No ITD	G12	Unknown
1 Gy N	Unknown	No Del	R235	No ITD	G12	Unknown
1 Gy N	M	Del	R235C	No ITD	G12	Unknown
1 Gy N	M	Del	R235H	No ITD	G12	Unknown
1 Gy N	F	Del	R235C	No ITD	G12D (Case 2)	Unknown
1 Gy N	F	Del	R235L	No ITD	G12	Unknown
1 Gy N	M	Del	R235C	No ITD	G12	Unknown
1 Gy N	M	Del	R235H	No ITD	G12	Unknown
1 Gy N	Unknown	Del	R235L	No ITD	G12	Unknown
0.5Gy +1Gy N	F	Del	R235S	No ITD	G12	Unknown
0.5Gy +1Gy N	Unknown	No Del	R235	No ITD	G12	Unknown
0.5Gy +1Gy N	M	Del	R235H	No ITD	G12	Unknown
0.5Gy +1Gy N	Unknown	Del	R235	No ITD	G12	Unknown
0.5Gy +1Gy N	M	Unknown	R235S	No ITD	G12	Unknown
0.5Gy +1Gy N	M	No Del	R235H	No ITD	G12	Unknown
0.5Gy +1Gy N	F	Del	R235	No ITD	G12	Unknown
0.1 + 1 Gy N	Unknown	Del	R235	No ITD	G12	Unknown
0.1 + 1 Gy N	F	Del	R235	No ITD	G12	Unknown
0.1 + 1 Gy N	Unknown	No Del	R235	No ITD	G12	Unknown
0.1 + 1 Gy N	F	No Del	R235H	No ITD	G12	Unknown
0.1 Gy + 0 N	F	No Del	R235	No ITD	G12	Unknown
0.1 Gy + 0 N	M	Del	R235C	No ITD	G12	Unknown
0.1Gy + 1Gy N	M	Del	R235C	No ITD	G12	Unknown
0.1Gy + 1Gy N	M	Del	R235H	No ITD	G12	Unknown
0.1Gy + 1Gy N	F	Del	R235S	No ITD	G12	Unknown
0.1Gy + 1Gy N	M	Del	R235S	No ITD	G12	Unknown
0.1Gy + 1Gy N	M	Del	R235C	No ITD	G12	Unknown
0.25Gy + 1Gy N	F	Del	R235H	No ITD	G12	Unknown
0.25Gy + 1Gy N	M	Del	R235S	No ITD	G12	Unknown
0.25Gy + 1Gy N	F	Del	R235C	No ITD	G12	Unknown
0.5+ 1 Gy N	F	Del	R235S	No ITD	G12	Unknown
0.5+ 1 Gy N	Unknown	No Del	R235	No ITD	G12	Unknown

	1Gy + 0 Gy N	F	Del	R235C	No ITD	G12	Unknown
	1Gy + 0 Gy N	M	Del	R235C	No ITD	G12	Unknown
	0.5Gy + 0 N	F	Del	R235C	No ITD	G12	Unknown
	0.5Gy + 0 N	F	Del	R235	No ITD	G12	Unknown
	0.5Gy + 0 N	Unknown	No Del	R235	No ITD	G12	Unknown
	1Gy + 1Gy N	M	Unknown	ATGC	No ITD	G12	Unknown
	0.25 + 0Gy N	M	Unknown	R235C	No ITD	G12	Unknown
	0.25 + 0Gy N	M	Del	R235H	No ITD	G12	Unknown
	0.25 + 0Gy N	Unknown	Del	R235	No ITD	G12	Unknown
	0.25 + 0Gy N	F	No Del	R235	No ITD	G12	Unknown
	1Gy + 1 Gy N	M	Del	R235H	No ITD	G12	Unknown
	1Gy + 1 Gy N	F	Del	R235C	No ITD	G12	Unknown
	0.25 + 1Gy N	F	Unknown	R235	No ITD	G12	Unknown
	0.25 + 0Gy N	M	Unknown	R235H	No ITD	G12	Unknown
	0.25 + 1Gy N	M	Unknown	R235C	No ITD	G12	Unknown
	0.5 + 0Gy N	M	Del	R235H	No ITD	G12	Unknown
F1 CBA/H	3 Gy Xrays	F	No Del	R235	ITD	G12	Unknown
x	3 Gy Xrays	F	No Del	R235	No ITD	G12R (Case 3)	Unknown
C57BL/Lia	3 Gy Xrays	F	Del	R235C	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235C	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235C	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235C	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235C	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235S	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235C	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235S	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235C	No ITD	G12	Unknown
	3 Gy Xrays	F	Del	R235C	No ITD	G12	Unknown
	3 Gy Xrays	Unknown	Del	R235C	No ITD	G12	Unknown
	3 Gy Xrays	F	Del	R235C	No ITD	G12	Unknown
	3 Gy Xrays	F	No Del	R235	ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235	No ITD	G12	Unknown
	3 Gy Xrays	F	Del	R235	No ITD	G12	Unknown
	3 Gy Xrays	F	Del	R235	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235C	No ITD	G12	Unknown
	3 Gy Xrays	F	Del	R235C	No ITD	G12	Unknown
	3 Gy Xrays	F	Del	R235H	No ITD	G12	Unknown
	3 Gy Xrays	F	Del	R235H	No ITD	G12	Unknown
	3 Gy Xrays	F	Del	R235H	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235C	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235S	No ITD	G12	Unknown
	3 Gy Xrays	M	Del	R235C	No ITD	G12	Unknown

Table 7. Details of strain, radiation dose, gender, chromosome 2 deletion, mutation and date of AML diagnosis for all rAML cases.

List of 123 rAML murine cases with strain, dose, gender, chromosome 2 deletion information from CGH arrays and *Sfpi1*, *Flt3* and *Kras* mutation details. Mutations are colour coded depending on the amino acid change. N = neutrons, Del = deletion.

As expected, *Sfpi1* was found to be the most commonly mutated gene with both codon R235 mutations and chromosome 2 deletions frequently identified. A total of 4 different types of missense mutations affecting codon R235 and one silent mutation affecting codon L134 were identified, affecting 68% of all cases in total. Mutations at codon R235 result in a conversion of arginine to either cysteine (R235C), histidine (R235H), serine (R235S) or leucine (R235L) with cysteine being the most common change at 32.5%. A high level of *Sfpi1* chromosome 2 deletions has previously been reported to occur in 90% of rAML cases with R235 mutations in 70% of these (Silver et al. 1999; Cook et al. 2004). Here, in the largest study to be reported so far, we report a lower level of chromosome 2 deletions, 81% in samples analysed by CGH, with 89% of them also containing a R235 mutation.

Flt3 was the next most commonly mutated gene with 5 different insertions in 5 CBA/H mice with these insertions in three mice previously reported (Finnon et al. 2012). An ITD was detected by the presence of two bands after PCR amplification of *Flt3* with one band amplifying the 333bp product and another larger band amplifying the amplicon with an insertion. Insertions in two of these mice (AML C, AML D) have been previously reported, while a fifth case was only reported in Finnon et al. (Finnon et al. 2012) and here we expanded on this work by reporting two new additional cases (AML A, AML B) (Figure 23 A). DNA Sanger sequencing revealed these insertions to be 24bp (AML A) and 33bp (AML B) in length with both being heterozygous for the mutation (Figure 23 B). These insertions are predicted to lead to a 6 and 11 amino acid insertion in the juxta membrane domain of *Flt3* (Figure 23 C), as is seen in human AML (Small 2006). *Flt3*-ITDs occurred in 4% of all cases with all four cases being female with the further fifth case from Finnon et al. also being female. Using Fisher's exact test for proportions, the presence of the *Flt3*-ITDs significantly occurred in 5 female mice with p=0.0045 when comparing its presence in 39 female with 72 male mice without *Flt3-*

ITDs being detected. Importantly, these cases, although small in number, indicate a female specific pathway with the presence of an *Flt3*-ITD.

A

Sample: CBA C57 Brain | AML A | AML B | AML C | AML D | AML E | AML F | AML G

333bp

Flt3-ITD: -   -   -   +   +   +   +   -   -   -
      del2: -   -   -   -   -   -   -   +   +   -

B

AML A

nt 2017 CTTCTACGTTGACTTCAGGGACTA

TTGACTTCAGGGACTA TGAATATGACCTTAAGTGGGAGTTCCCGAGAG
nt 2002   nt 2049

AML B

nt 2035 AACTTCAGGGACTATGAATATGACCTTAAGTGG

TTGACTTCAGGGACTATGAATATGACCTTAAGTGG GAGTTCCCGAGAG
nt 2002   nt 2049

C

AML A aa.598
FYVDFRDY

EYFYVDFRDY EYDLKW EFPRENLEF
aa.589

NFRDYEYDLKW
AML B aa.604

Figure 23. Analysis of *Flt3*-ITD in murine rAMLs.

(A) Agarose gel electrophoresis of *Flt3*-ITD PCR of a panel of murine rAMLs on a 2% gel. A normal amplicon is represented by a single band of 333 bp while and *Flt3*-ITD has an additional larger band. The gel is loaded as indicated in the image. "CBA" and "C57" refer to normal spleen DNA from CBA/H and C57BL6, respectively; "Brain" refers to brain tissue from the animal in which AML A developed and represents a normal tissue control; and AMLs E-G refer to independent AML samples. *Flt3*-ITDS in AMLs C and D have been previously analysed in Finnon et al. 2012. The presence

or absence of an *Flt3*-ITD and chromosome 2 deletion are stated below the gel for each sample (*Flt3*-ITD: + presence and – absence; del2: + presence and – absence. (B) CBA/H sequence of exon 14 on chromosome 5 showing *Flt3*-ITDs in AML A at nucleotide 2017 (24bp) and in AML B at nucleotide 2035 (33bp). (C) Predicted CBA/H *Flt3* protein sequence with ITDs of AML A (aa.598) and AML B (aa.604).

Three *Kras* G12 mutations, two G12D and one G12R, were also identified occurring in 2% of all samples overall. The two cases with *Kras* G12D mutations also carried *Sfpi1* R235C while the *Kras* G12R mutation case did not have any co-occurring *Sfpi1* R235 mutation. To our knowledge, this is the first time a Gly→Arg mutation at codon 12 in *Kras* has been reported in the CBA mouse model. To assess the influence of this mutation on the protein function, these amino acid changes in *Kras* were analysed using both PolyPhen2 and PredictSNP algorithms (Table 8). Data mining assessment of these mutations reported a predicted deleterious impact for protein function using both PolyPhen2 (possibly damaging/Sens=0.75/Spec=0.87) and PredictSNP (deleterious: 87%) algorithms. PolyPhen-2 calculated 75% sensitivity and 87% specificity of the prediction in gaining the deleterious effect (for both amino acid changes). While PolyPhen-2 tested the amino acid conservation in the human organism, PredictSNP was run to verify those findings for the mouse model. The PredictSNP algorithm confirmed PolyPhen-2 predictions with 87% confidence. Thus, using two independent bioinformatic methods we expect the deleterious impact on protein function from the observed mutations.

Gene	Chromosome	Ref->Seq	AA change	a.a. prediction	
				PolyPhen - 2	PredictSNP
Kras	6	GGT-->GAT	Gly-->Asp (G12D)	posDEM/S=0.740/Sens=0.85/Spec=0.92	D:87%
		GGT-->CGT	Gly-->Arg (G12R)	posDEM/S=0.884/Sens=0.82/Spec=0.94	D:87%

Table 8. PolyPhen2 and PredictSNP analysis of *Kras* in murine AML.

Kras mutations in CBA/H mice with AML were analysed with predicted amino acid change effect.

The mutations affecting codon G12 in the gene *Kras* and codon R135 in the gene *Sfpi1* detected by Sanger sequencing were confirmed by pyrosequencing using the PyroMark Q48 (Figure 24). The 3 mouse samples with *Kras* mutations were analysed and the percentage of each mutation was calculated as illustrated in Table 9. The frequency of the *Sfpi1* mutation is high, present in 69-70% of the sample, indicating that it is a driver mutation. The frequency of the *Kras* mutations is lower at 30-46%, indicating that the *Kras* G12 mutation is a secondary mutation, similar to previously suggested reports that it acts as a co-operating mutation, not capable of initiating AML by itself (Chang et al. 2015; Nakagawa et al. 1992; Zhao et al. 2014). The sequence of mutational events in case 1 and case 2 illustrate a model of clonal expansion in rAML. The novel Gly-->Arg mutation has been confirmed in the CBA/H mouse by two techniques, Sanger sequencing and pyrosequencing, at a frequency of 38-39% and is present in the absence of a *Sfpi1* mutation. This suggests that there is another factor driving AML development which currently has not been identified. Overall, this work shows that in murine rAML cases the most common genetic alteration is a chromosome 2 deletion with a *Sfpi1* point mutation, with minor cases with *Flt3*-ITDs and *Kras* mutations.
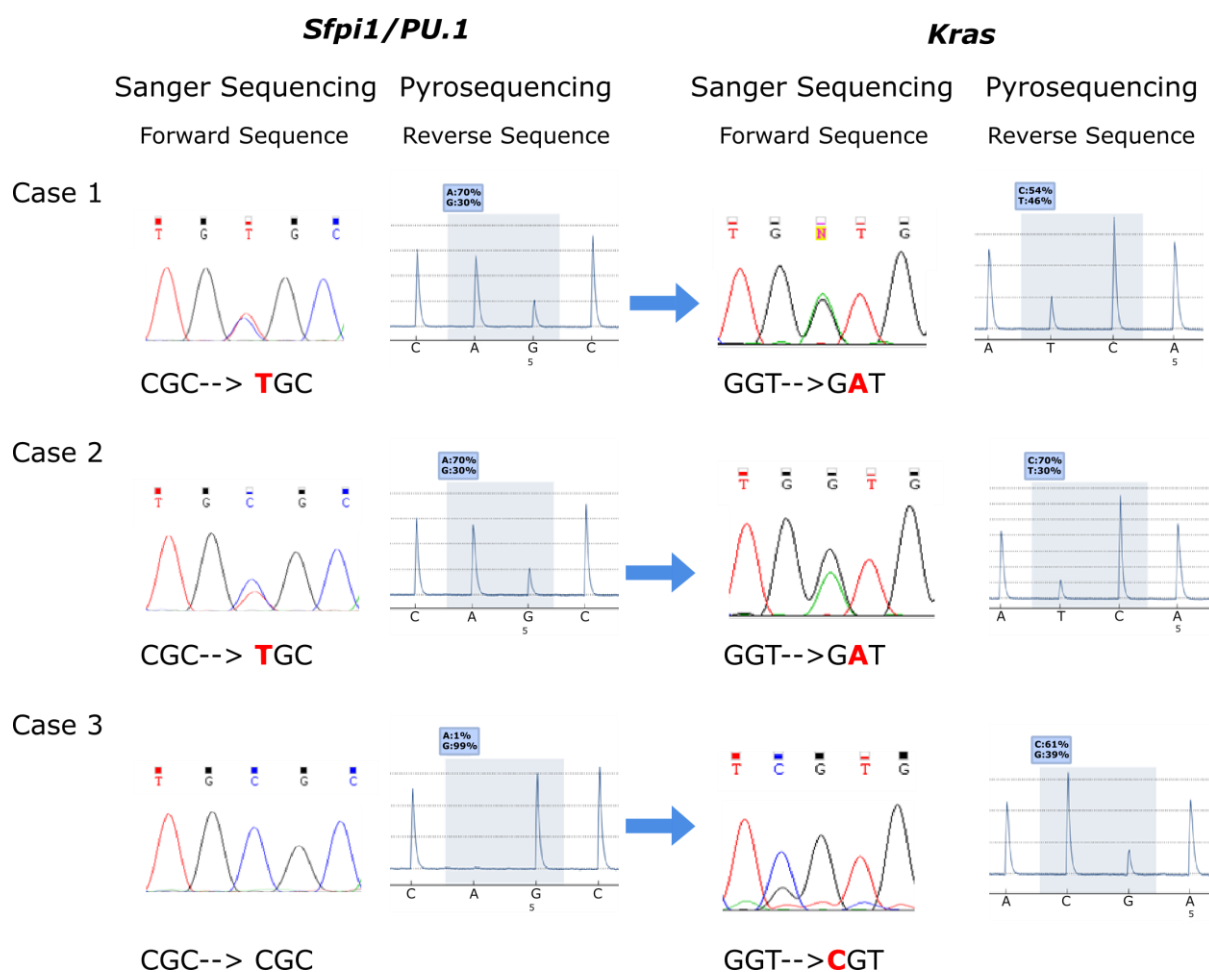
Figure 24. Analysis of *Sfpi1/PU.1* R235 and *Kras* G12 codons by Sanger sequencing and pyrosequencing in three rAML cases.

Case 1 and case 2 have *Sfpi1* R235C mutations of CGC→TGC and *Kras* G12D mutations of GGT→GAT, while case 3 was unmutated for *Sfpi1* but has a *Kras* G12R mutation of GGT→CGT.

Sample	*Kras* G12 mutation	*Sfpi1* R235 mutation	% *Kras* mutation	% *Sfpi1* mutation
Case 1	GGT>GAT	CGC>TGC	45-46	70
Case 2	GGT>GAT	CGC>TGC	30-31	69-70
Case 3	GGT>CGT	CGC	38-39	0
Control	GGT	CGC	0	0

Table 9. Analysis of *Kras* and *Sfpi1* mutations in murine AML.

The percentage of these mutations in 3 CBA/H mice with AML and 1 CBA/H without AML was quantified by pyrosequencing.

3.2.3 mRNA expression

Transcriptional expression of *Sfpi1* in rAML mouse spleen samples was measured to analyse the level of expression associated with the presence or absence of a R235 mutation. MQRT-PCR was performed on 69 rAML spleen samples with a mutation in codon R235, 35 rAML spleen samples without a mutation in codon R235 and on three bone marrow control samples from control CBA/H mice. Importantly, samples without a mutation in codon R235 had a significantly lower expression level of *Sfpi1* when compared to control bone marrow samples and AML samples with a R235 mutation (Figure 25). In the 3 samples which had a *Kras* mutation, the two which had a *Sfpi1* R235 mutation had a high level of *Sfpi1* expression (case 1 and case 2), while the sample which did not have a *Sfpi1* R235 mutation (case 3) had a very low level of *Sfpi1* expression at 0.13. Hence, the repression of *Sfpi1* could therefore be driving AML development for all cases in the absence of a *Sfpi1* mutation, but specifically for case 3.

For samples that had no *Sfpi1* R235 mutation, there was a significantly lower level of *Sfpi1* expression, however a few samples showed a high level of expression. These included samples with an *Flt3*-ITD. Out of the four *Flt3*-ITD cases, two cases had a low level of *Sfpi1* expression, while the other two cases had the highest levels of *Sfpi1* expression. Gene expression analysis of *Flt3* expression itself in all samples did show a significant increase in expression in samples with an ITD (Figure 26). Although an inverse relationship between *PU.1* and *FLT3* expression has previously been reported (Inomata et al. 2006), a negative correlation was not seen in this study (Figure 27). Overall, murine rAML samples with a *Sfpi1* R235 mutation had a higher level of expression and cases that had no *Sfpi1* R235 mutation had a lower level of expression.
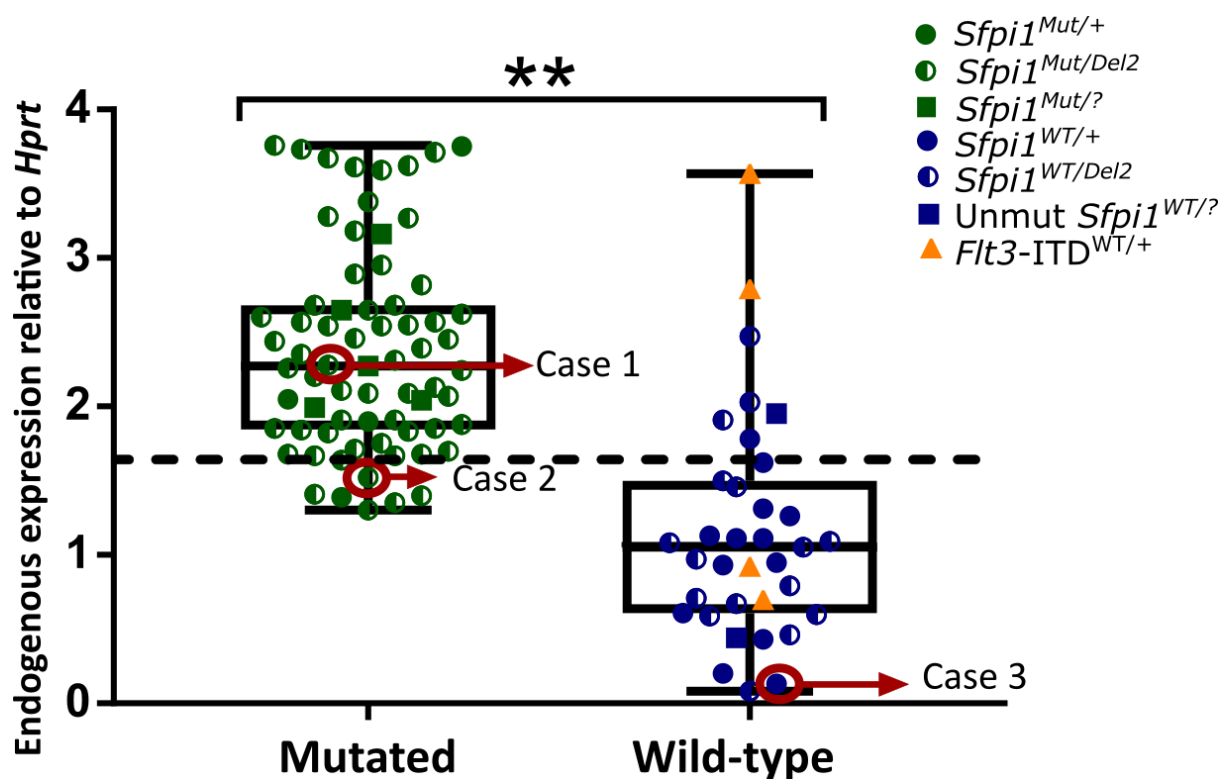
Figure 25. **MQRT-PCR expression of *Sfpi1* in murine rAML samples.**

Expression was analysed in 3 CBA/H bone marrow control samples (dashed line represents the median of expression level), 69 samples mutated in codon R235 (green circles) and 35 wild type samples (blue circles) including 4 samples with an *Flt3*-ITD (orange triangles). Expression of *Sfpi1* in case 1, 2 and 3, as depicted in Figure 2, are circled in red. Chromosome 2 deletion information is also represented by a half full circle while samples with unknown chromosome 2 information are represented by a square box. Expression levels were normalised to *Hprt*. Significance (p≤0.001) was calculated by performing a Mann Whitney test on gene expression data and indicated with asterisks (**).
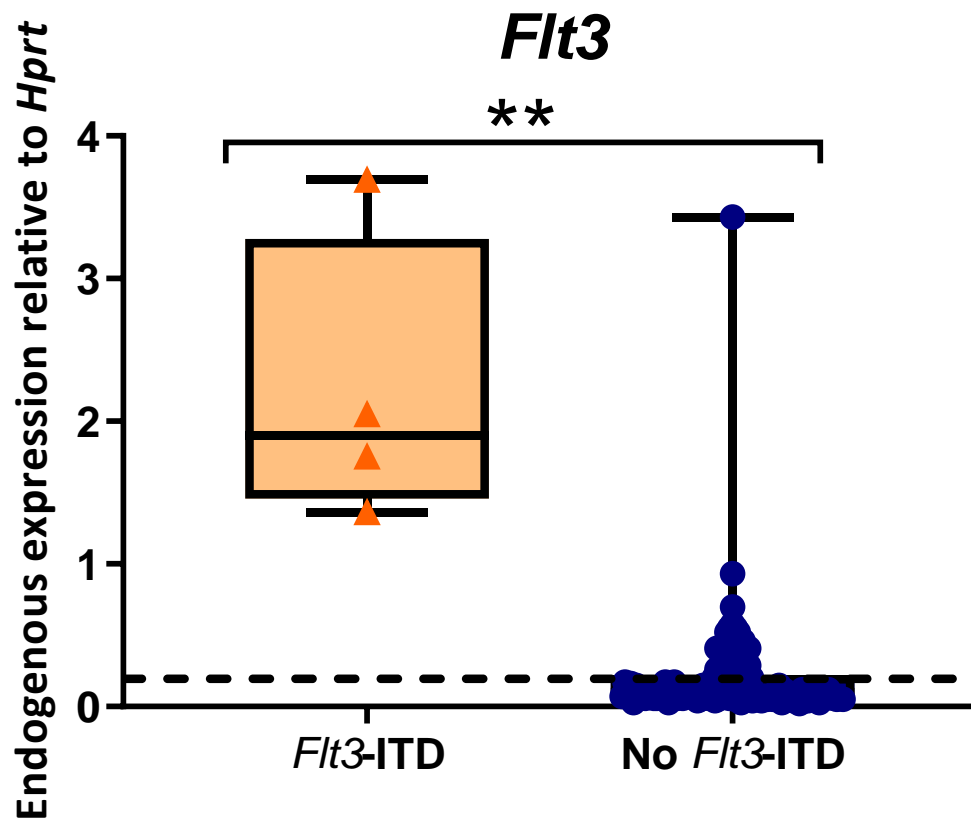
Figure 26. MQRT-PCR expression of *Flt3* in murine rAML samples.

Expression was analysed in 3 CBA/H bone marrow control samples (dashed line median expression), 4 samples with an *Flt3*-ITD (orange triangles) and 100 samples without an *Flt3*-ITD (blue circles). Expression levels were normalised to *Hprt*. Significance (p≤0.001) was calculated by performing a Mann Whitney test on gene expression data and indicated with asterisks (**).
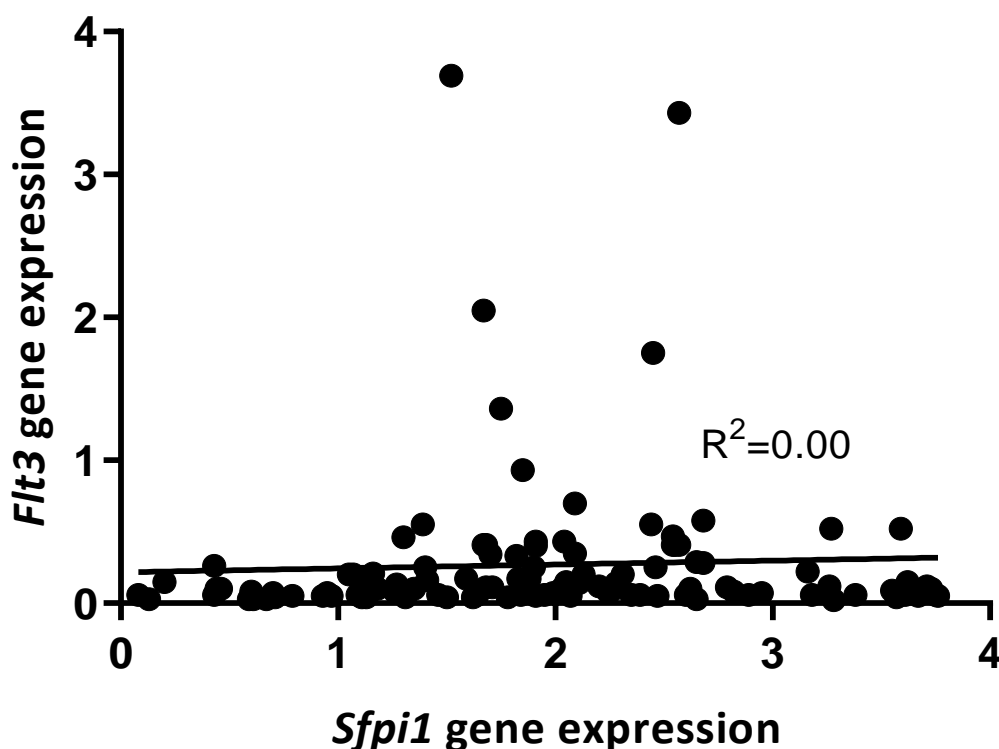
Figure 27. Correlation analysis of *Sfpi1* against *Flt3* transcriptional expression.

Correlation analysis was performed using linear regression analysis in 104 rAML cases. Correlation analysis was performed using GraphPad Prism 7 with R^2 values displayed.

3.2.4 miRNA expression

To assess if specific epigenetic modifications could be responsible for the overall lower expression of *Sfpi1* in samples without a R235 mutation, miRNA expression analysis was performed. miRNA expression was assessed by the nCounter miRNA Expression Panel by Nanostring. Analysis of the data by BRBArrayTools software identified significantly expressed miRNA of interest between groups with the use of a class comparison analysis. Using a stringent significance threshold of p≤0.001 and p≤0.05, miRNAs of interest were identified (Figure 28). Comparison of samples with and without a *Sfpi1* mutation showed a significantly higher level of *miR-1983* and *miR-582-5p* expression (p≤0.05) in samples with the mutation (Figure 28 A). Samples with an

Flt3-ITD showed a significantly lower level of *miR-582-5p* and *miR-467c* (p≤0.001) in comparison to samples with a *Sfpi1* mutation (Figure 28 B).
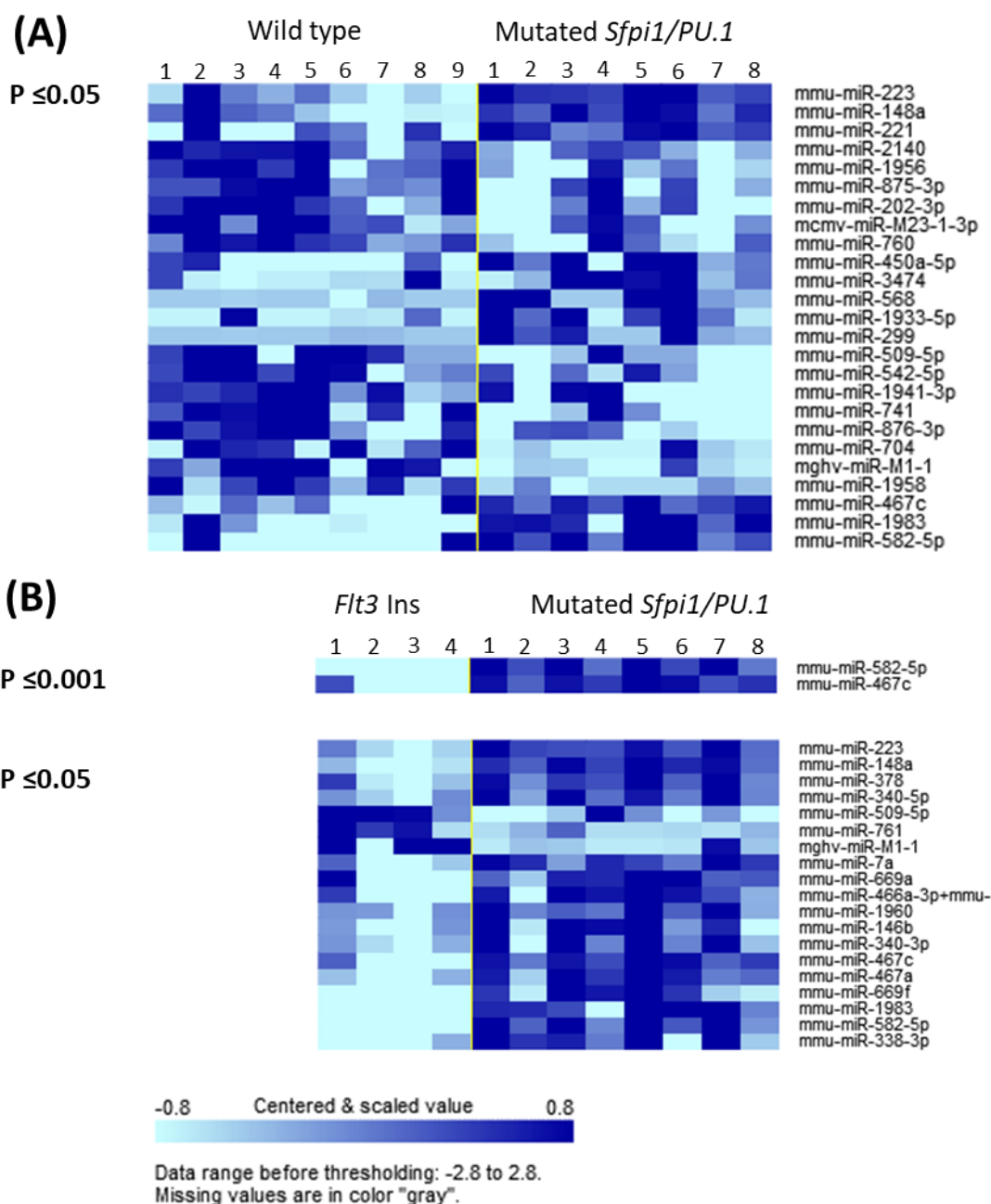
Figure 28. nCounter system miRNA expression in rAML samples.

A class comparison analysis was performed on samples using BRB-ArrayTools developed by Dr. Richard Simon and the BRB-ArrayTools Development Team. (A) miRNA expression in 9 unmutated *Sfpi1* R235 and 8 mutated *Sfpi1* R235 samples

where p≤0.05. (B) miRNA expression in 4 samples with an *Flt3*-ITD and 8 mutated R235 *Sfpi1* samples where p≤0.001 and p≤0.05.

Housekeeping genes are constitutively expressed in all cells and so are used in QPCR as controls to normalise the data. They correct for differences in cDNA quantities among the samples caused by technical errors such as pipetting and so play an important part in a study design. However, the consistency of housekeeping genes can vary across tissue types so it is important to reassess for each new study design (Bustin and Nolan 2004). To identify a suitable housekeeping gene for normalisation, the expression of *RNU6*, *SNORD47* and *SNORD66* was analysed by QPCR. NormFinder was then used to establish which gene was the most stable (Figure 29). The NormFinder algorithm analyses QPCR data to look at the variability across the data and identifies the optimal normalisation gene among a set of candidates by ranking them with the lowest value indicating the most stable. *RNU6* was identified as the most stable with the lowest value and so used for QPCR analysis of the miRNA of interest.
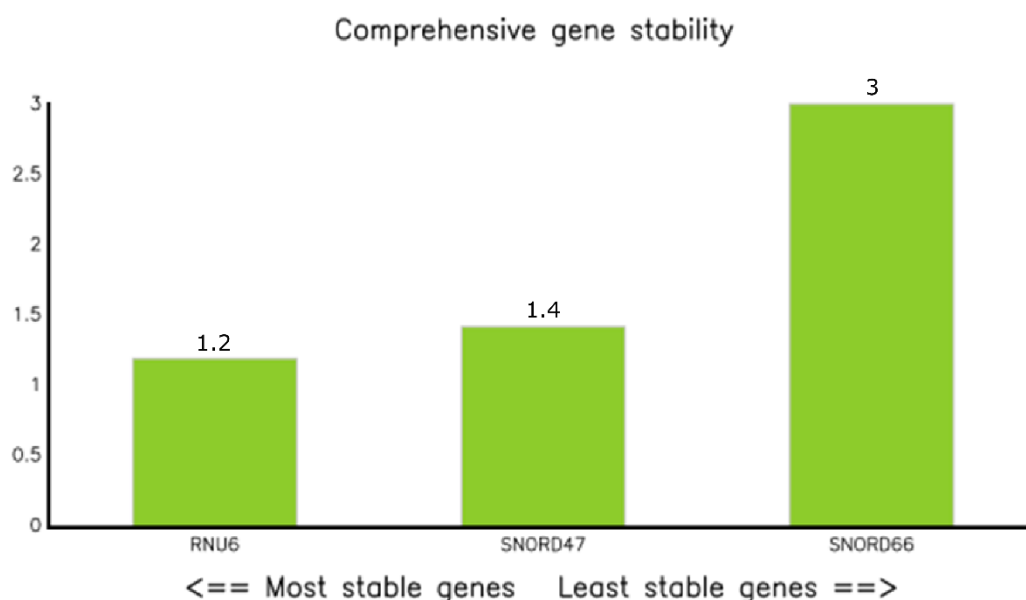
Figure 29. NormFinder gene stability results.

These results show the stability rating of *RNU6*, *SNORD47* and *SNORD66* miRNA expression in rAML samples.

The expression of *miR-1983*, *miR-582-5p*, and *miR-155* in 104 rAML samples (69 *Sfpi1* mutated samples and 35 *Sfpi1* unmutated samples) was confirmed by QPCR (Figure 30). *miR-1983* and *miR-582-5p* showed a significant upregulation in *Sfpi1* R235 mutation cases in comparison to control bone marrow samples (dashed line) (Figure 30 A and B). This upregulation in cases with a *Sfpi1* R235 mutation was also significant in comparison to *Sfpi1* wildtype cases. *miR-155*, a regulator of *Sfpi1* expression, was significantly upregulated in all rAML samples in comparison to control bone marrow samples (Figure 30 C). This upregulation, however, was not significantly higher in samples that were unmutated for *Sfpi1* in comparison to *Sfpi1* wildtype cases. The expression of *miR-582-5p*, and *miR-467c* in 73 rAML samples (69 *Sfpi1* mutated samples and 4 *Flt3*-ITD samples) was also compared (Figure 31). Both miRNA show a significant upregulation in *Sfpi1* R235 mutated samples and a significant down-regulation in *Flt3*-ITD samples in comparison to control bone marrow. This difference in expression between *Sfpi1* R235 mutated samples and *Flt3*-ITD samples was significant for both *miR-582-5p* (Figure 31 A), and *miR-467c* (Figure 31 B). This work identified the miRNA, *miR-1983*, *miR-582-5p* and *miR-467c*, which may be involved in the pathway of cases with a *Sfpi1* R235 mutation.
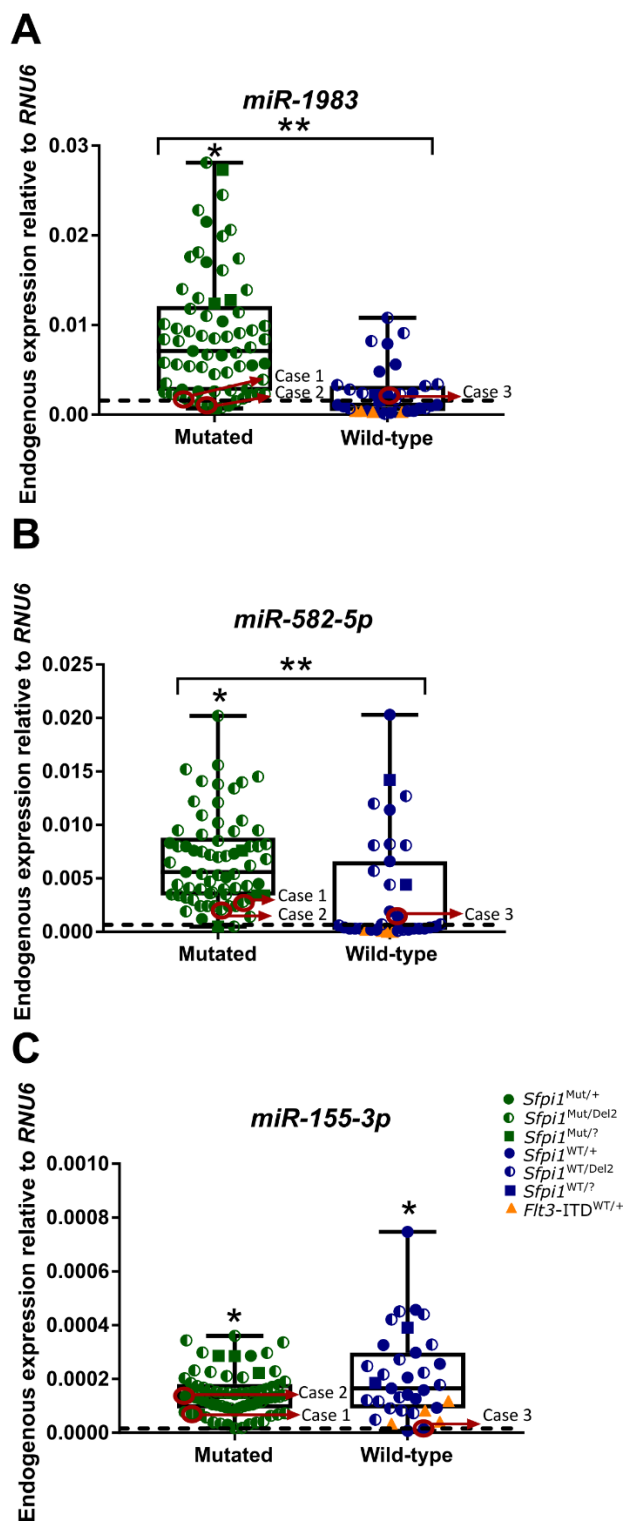
Figure 30. QRT-PCR expression of (A) *miR-1983*, **(B)** *miR-582-5p* **and (C)** *miR-155-3p* **in murine rAML samples.**

Expression was analysed in 3 CBA/H bone marrow control samples (dashed line median expression), 69 samples mutated in codon R235 (green circles) and 35 wild

type samples (blue circles) including 4 samples with an *Flt3*-ITD (orange triangles). Expression of *Sfpi1* in case 1, 2 and 3, as depicted in Figure 1, are circled in red. Expression levels were normalised to *RNU6* gene. Significance was calculated by performing a Mann Whitney test on gene expression data and indicated with an asterisk (*=p ≤0.05, ** p=≤0.001).
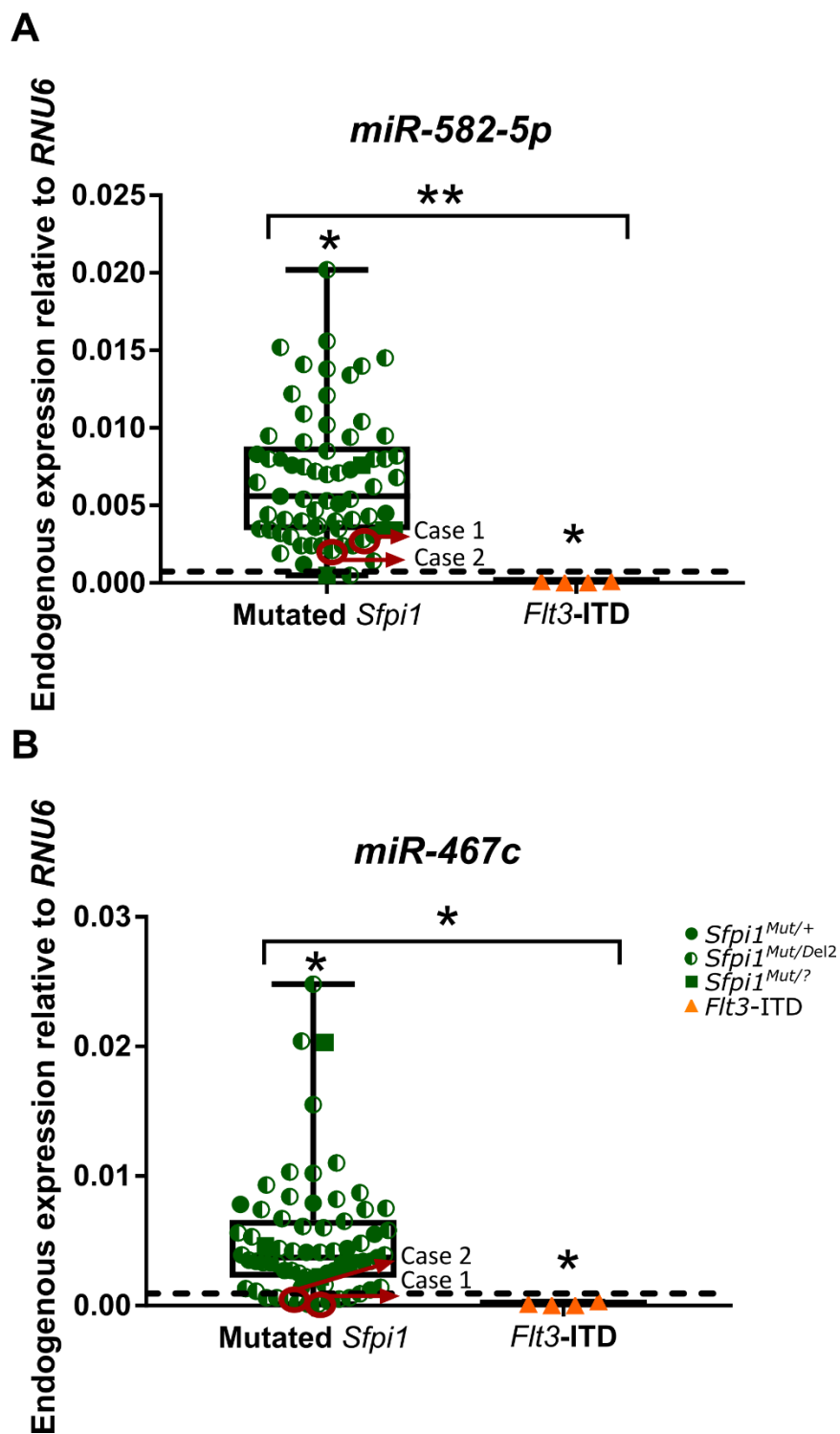
Figure 31. QRT-PCR expression of (A) *miR-582-5p* **and (B)** *miR-467c* **in murine rAML samples.**

Expression was analysed in 3 CBA/H bone marrow control samples (dashed line median expression), 69 samples mutated in codon R235 (green circles) and 4 samples with an *Flt3*-ITD (orange triangles). Expression of *Sfpi1* in case 1 and 2, as depicted

in Figure 1, are circled in red. Expression levels were normalised to *RNU6* gene. Significance was calculated by performing a Mann Whitney test on gene expression data and indicated with an asterisk (*=p ≤0.05, ** p=≤0.001).

3.2.5 DNA methylation

To further investigate the repression of *Sfpi1* in unmutated R235 samples, DNA methylation was next analysed. Here, DNA methylation primers were designed to analyse 4 CpG sites in the promoter region and 5 CpG sites in the URE by pyrosequencing (Figure 32), which have been shown to regulate Sfpi1 transcriptional expression (Chen, Ray-Gallet, et al. 1995; Li et al. 2001). These primer designs, two primer sets for the URE region and one primer set for the promoter region, were first validated by use of a commercial positive DNA methylated mouse control, with DNA methylation levels of 90% to 100% reported (Figure 33).
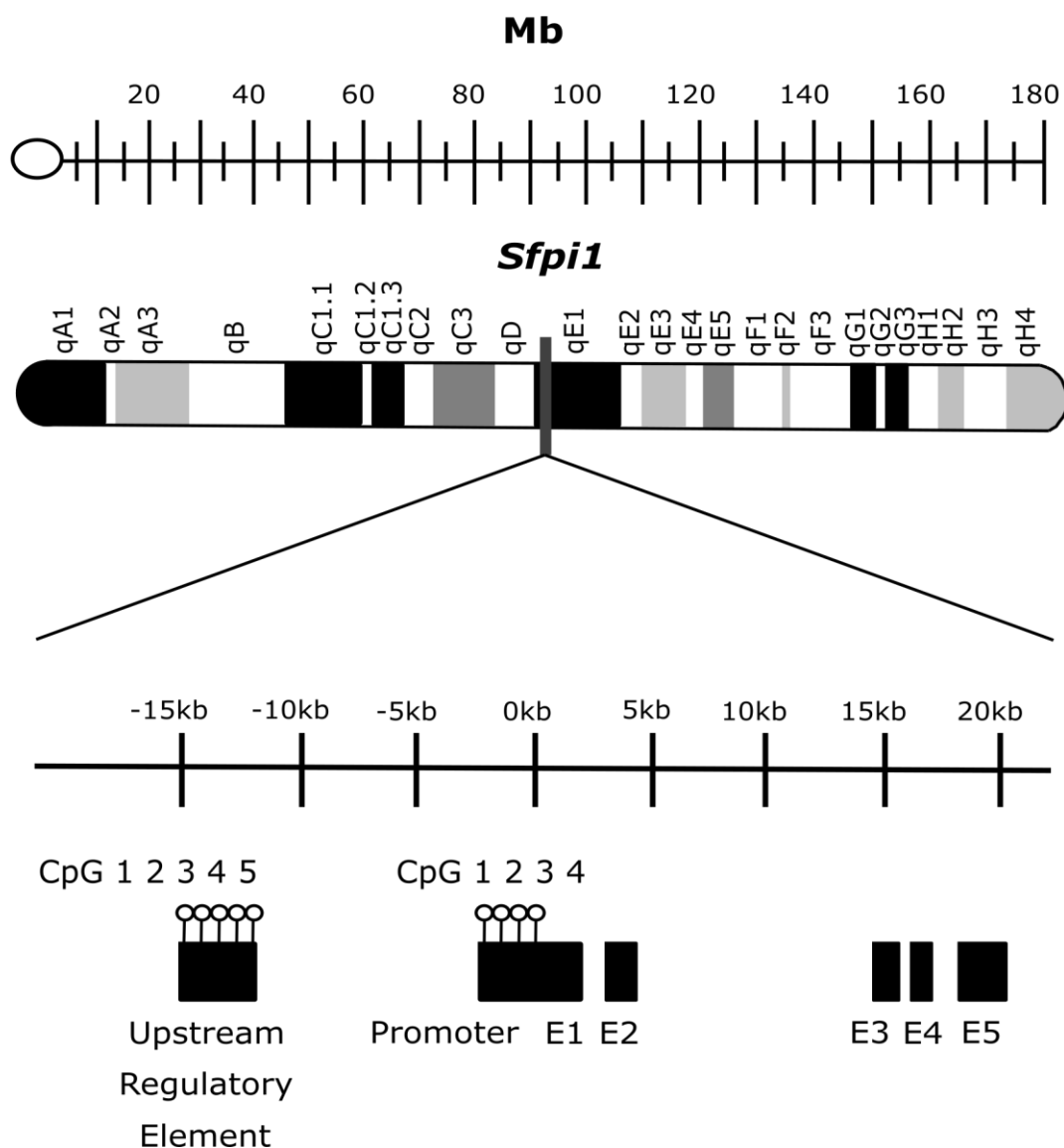
Figure 32. Genomic location of methylated CpG sites in *Sfpi1* in the mouse.

(A) *Sfpi1* is located on chromosome 2 (2qE1). The gene *Sfpi1* consists of five exons (E1 – E5) and two upstream regions which control its expression, an upstream regulatory element which is located 14kb upstream of the transcription start site and a promoter region. Five CpG sites were analysed in the upstream regulatory element while 4 CpG sites were analysed in the promoter.

A

B

C
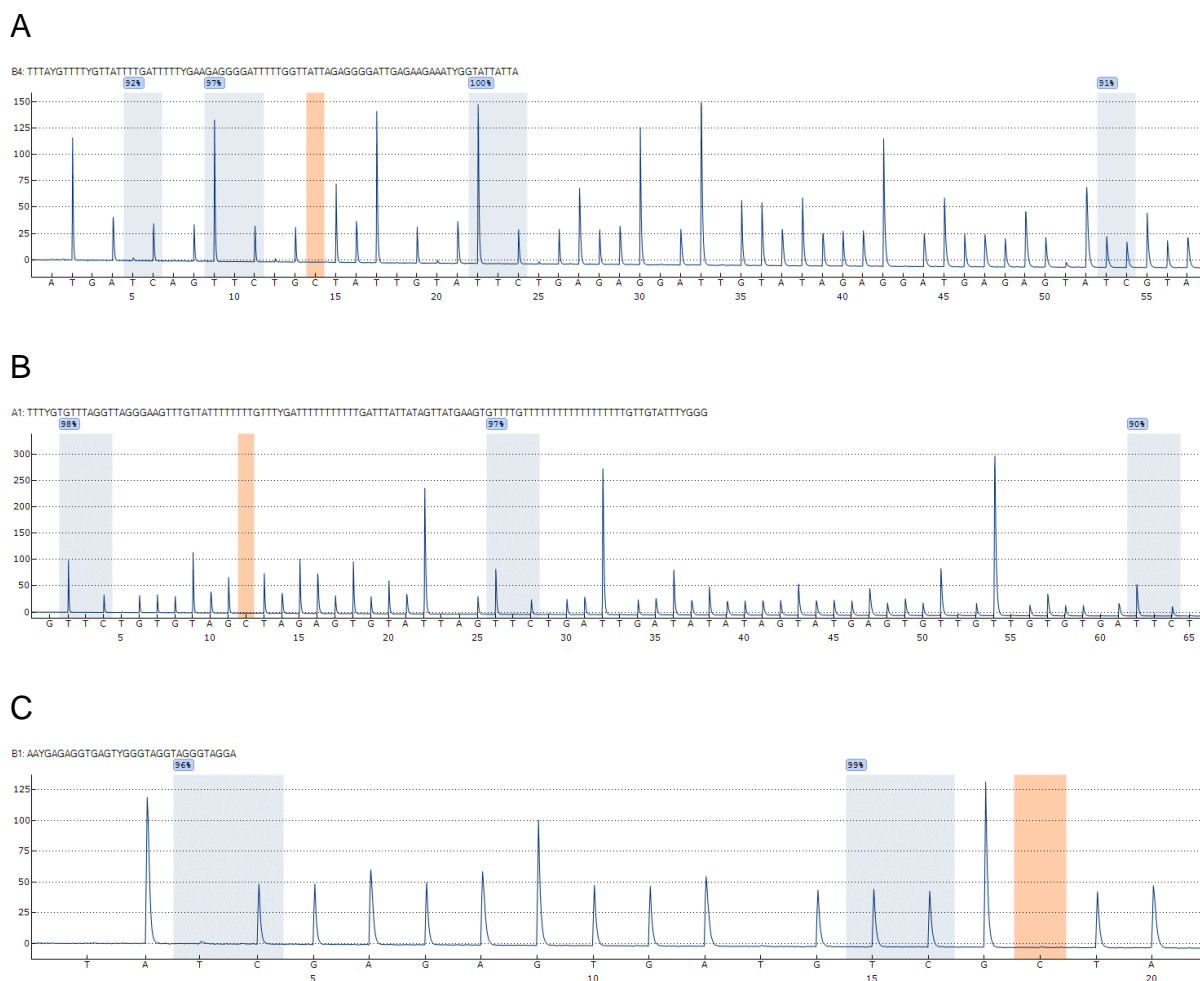
Figure 33. Pyrosequencing of *Sfpi1* in the mouse using a positive control.

Pyrosequencing pyrograms of methylated mouse control DNA in the promoter region (A), URE region 1 (B) and URE region 2 (C) of *Sfpi1*. CpG regions are identified by a blue shading. A bisulfite control, which checks the efficiency of the bisulfite conversion of unmethylated cytosine to thymine by the insertion of a cytosine into the sequence before a thymine, is identified by an orange shaded region. Percentage of methylation is indicated by a percentage above each CpG site.

DNA methylation levels in the samples mutated for *Sfpi1* R235 showed a low level of DNA methylation, averaging 15% across all cases (Figure 34). In samples which were unmutated for *Sfpi1* R235, a higher level of DNA methylation was observed. This is particularly evident for case 3 which has a very high level of DNA methylation averaging 76% and reaching 91% at one CpG site. For case 3, DNA methylation may well be the factor causing repression of *Sfpi1* gene expression. Overall, DNA

methylation levels were significantly higher at each of the 5 CpGs in the URE and the 4 CpGs in the promoter region in samples without a *Sfpi1* R235 mutation when compared to samples with a *Sfpi1* R235 mutation and control bone marrow (Figure 35).
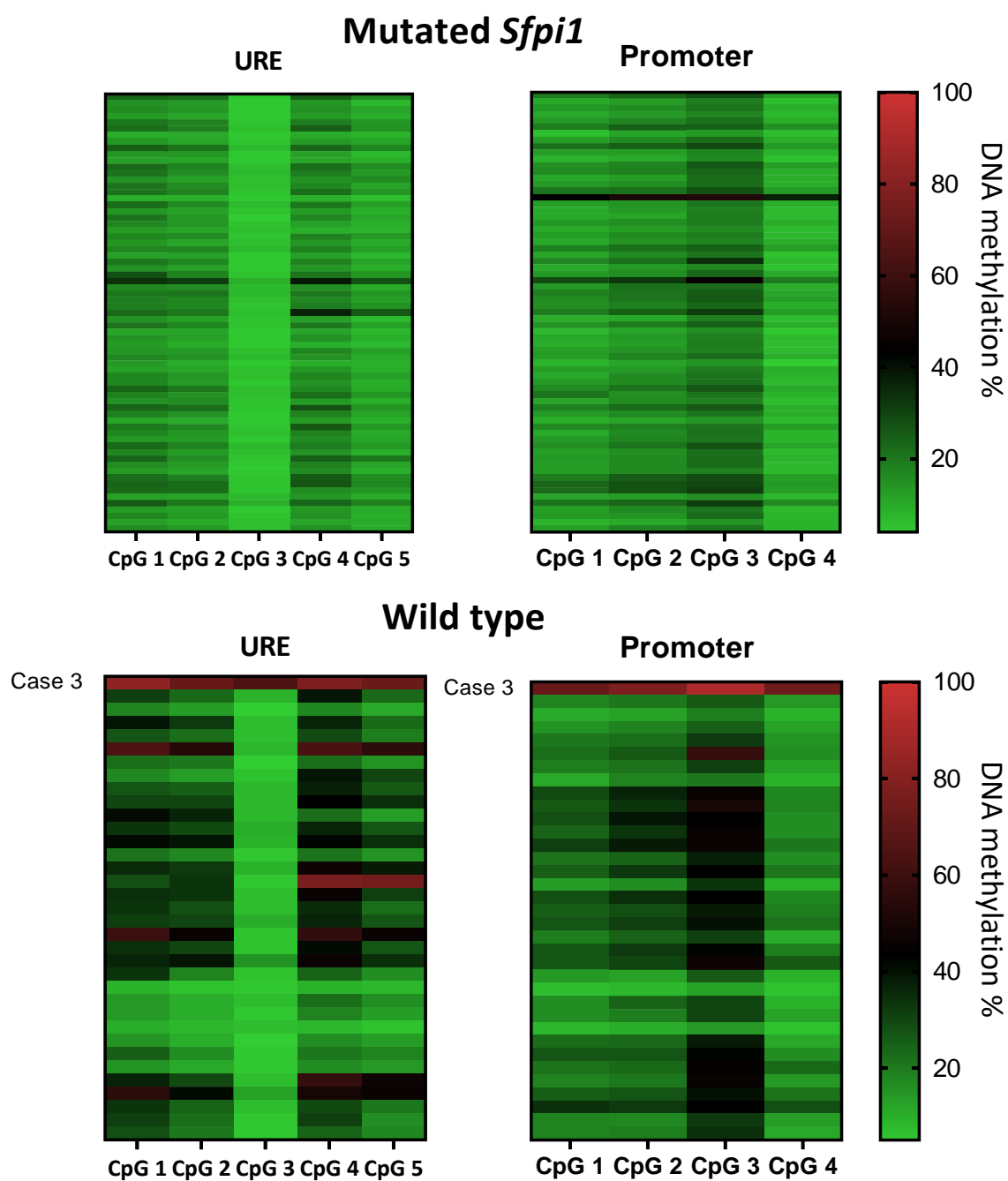
Figure 34. Heatmap analysis of DNA methylation in rAML samples.

69 samples mutated in codon R235 (top) and 45 samples unmutated in codon R235 (bottom) of *Sfpi1* were analysed by pyrosequencing at 5 CpG sites in the upstream regulatory element (URE) and 4 CpG sites in the promoter region. DNA methylation of case 3, as depicted in Figure 1, is identified.
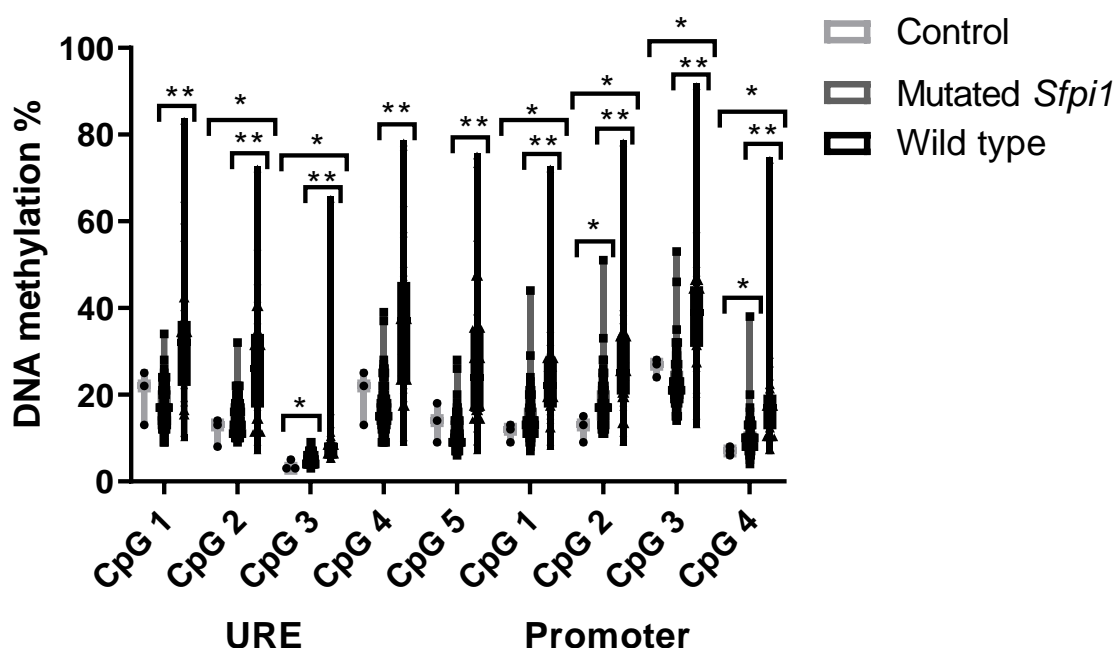
Figure 35. DNA methylation levels in rAML mouse samples.

DNA methylation was measured in 3 CBA/H bone marrow control samples, 69 samples mutated in codon R235 and 35 samples unmutated in codon R235 by pyrosequencing at 5 CpG sites in the upstream regulatory element (URE) and 4 CpG sites in the promoter region. Significance was calculated by performing a Mann Whitney test on gene expression data and indicated with an asterisk (*=p ≤0.05, ** p=≤0.001).

To investigate if the change in DNA methylation levels is linked to the change in *Sfpi1* transcriptional levels, correlation analysis was performed (Figure 36). An inverse correlation between DNA methylation levels and *Sfpi1* transcriptional expression was

evident, which showed a strong correlation for the URE CpG 1 (A), URE CpG 2 (B), URE CpG 4 (D), URE CpG 5 (E) and promoter CpG 3 (H) in particular of R^2=0.28, R^2= 0.24, R^2= 0.25, R^2= 0.21 and R^2= 0.21 respectively. DNA methylation has been widely studied as a factor of aging with studies reporting global hypomethylation in CD4+ T cells with increasing age (Heyn et al. 2012) and also local hypermethylation at specific sites (Beerman et al. 2013). To investigate if age at diagnosis was causing this inverse relationship, correlation analysis was performed on samples whose date of sacrifice was known, however no correlation was seen (Figure 37). Samples with known age at diagnosis however, were limited at just 33 mice and most were of similar age at sacrifice of 15 months, only 4 mice were sacrificed at a younger age (i.e. less than 12 months) and only 4 mice sacrificed at greater than 20 months. Overall, DNA methylation levels were significantly higher in rAML cases without a *Sfpi1* R235 mutation when compared to cases with a mutation.
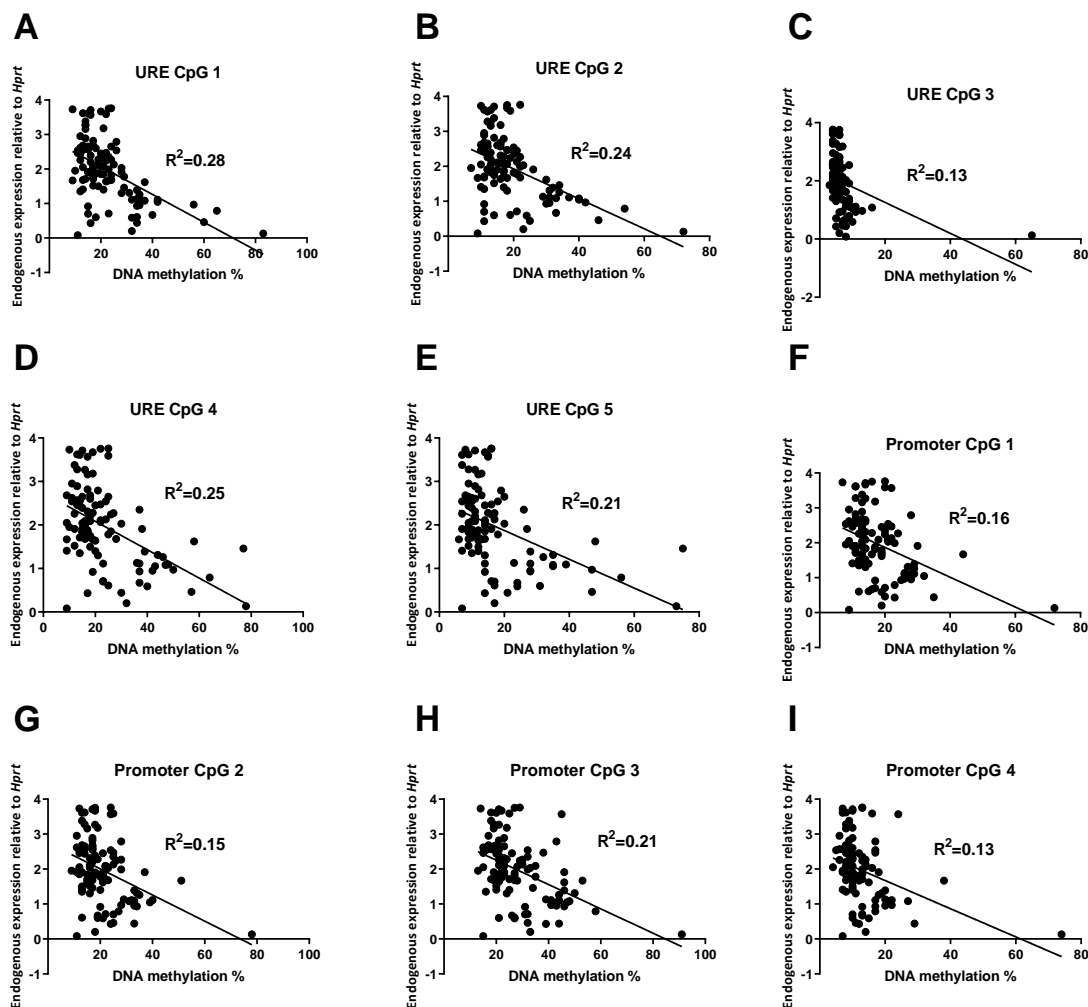
Figure 36. Correlation analysis of *Sfpi1* mRNA expression against DNA methylation.

Correlation analysis was performed on 104 rAML cases for URE CpG 1 (A), 2 (B), 3 (C), 4 (D) and 5 (E) and promoter CpG 1 (F), 2 (G), 3 (H) and 4 (I). Correlation analysis using linear regression shows a linear relationship between mRNA expression and DNA methylation in particular for CpG 4 and 5 in the URE and CpG 3 in the promoter of *Sfpi1*. Correlation analysis was performed using GraphPad Prism 7 with R^2 values displayed.
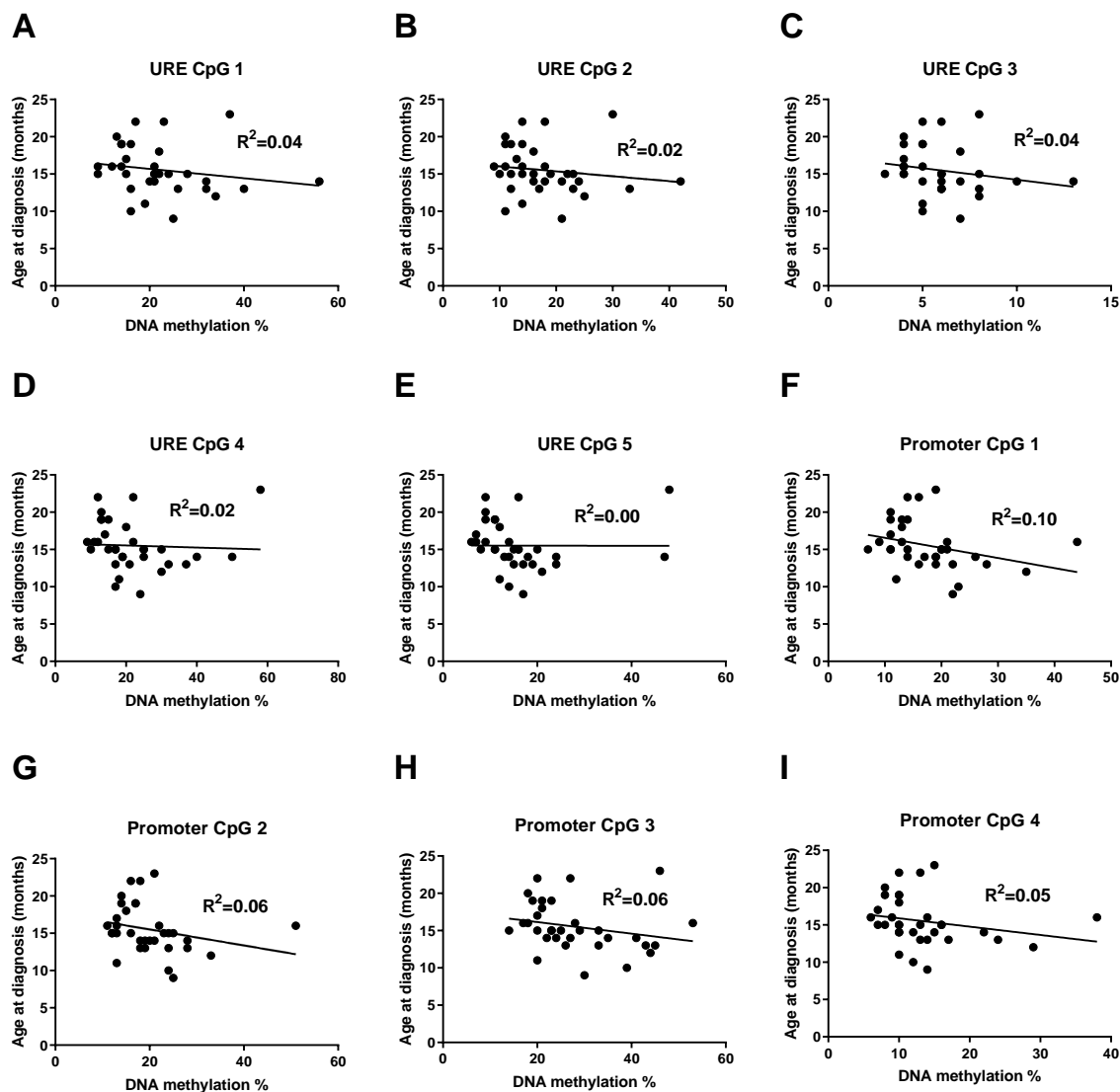
Figure 37. Correlation analysis of age at rAML diagnosis against DNA methylation.

Correlation analysis was performed for URE CpG 1 (A), 2 (B), 3 (C), 4 (D) and 5 (E) and promoter CpG 1 (F), 2 (G), 3 (H) and 4 (I) in 33 rAML samples. Correlation analysis using linear regression shows no relationship between age at diagnosis and DNA methylation. Correlation analysis was performed using GraphPad Prism 7 with R^2 values displayed.

Discussion and future directions

In this study, we have screened a large cohort of mouse rAML samples for commonly mutated genes in human AML identifying 3 mutated genes, *Sfpi1*, *Flt3* and *Kras*, further validating the use of the CBA mouse for radiation-induced AML studies. This work illustrates the central role *Sfpi1* dysregulation plays in rAML development in the CBA mouse with *Sfpi1* being affected by with chromosome 2 deletion and point mutation or else DNA methylation of the *Sfpi1* promoter with a decrease in transcription expression.

Sfpi1 chromosome 2 deletions has previously been reported to occur in 90% of rAML cases, with R235 mutations occurring in 70% of cases with a chromosome 2 deletion (Azumi and Sachs 1977; Silver et al. 1999) (Cook et al. 2004). Here in this large study a slightly lower frequency of chromosome 2 deletions is reported in 81% of samples analysed by CGH with 89% also containing a R235 mutation. Intriguingly, cases with a chromosome 2 deletion appear to consist mainly of male mice. When comparing the presence of a chromosome 2 deletion relative to gender, deletions were identified in 61 out of 65 male mice and in 27 out of 37 female mice. A two-sample t-test between proportions showed a significant difference in the presence of chromosome 2 deletions between gender ($p=0.006$). It is worth considering that this significance may be influenced by the 4 female mice with *Flt3*-ITDs, all of which have no chromosome 2 deletions and when removed from the study, show no significant difference in chromosome 2 deletions between gender. Previous studies analysing chromosome 2 deletions were mainly performed only using male mice. One study using both male and female mice reported no difference in the occurrence of chromosome 2 deletions based on gender (Clark et al. 1996), however, the numbers of rAML mice in this study was small at just 15 mice. There were no differences in the presence of chromosome 2 deletions by radiation type. This was expected as chromosome 2 deletions has previously been assessed in a sub-set of these cases and also found no difference between radiation type (Brown et al. 2015).

As previously reported in the CBA/H or CBA/Ca mouse model (Cook et al. 2004), a *Sfpi1* R235 mutation was the most commonly reported mutation occurring in 68% of all cases analysed. A cytosine to thymine mutation was the most frequent mutation at codon R235. This mutation has commonly been seen in many other cancers in both

mouse (Cook et al. 2004) and human cases (Alexandrov et al. 2015; Welch et al. 2012). It is likely due to the presence of a CpG site in the codon. The first cytosine in the codon which when methylated as 5-methylcytosine, can spontaneously deaminate to form thymine (Shen, Rideout, and Jones 1994). If this deamination remains unrepaired and cell replication occurs, this C>T mutation will remain in the genome.

FLT3 mutations are the most common mutation in human de novo AML and *FLT3*-ITDs are commonly present in around 20% of patients (Thiede et al. 2002; Cancer Genome Atlas Research et al. 2013; Kihara et al. 2014). In this study, *Flt3*-ITDs were reported to occur in 4% of samples and importantly appears to be a significantly female specific pathway. Sex differences have long been investigated with a recent investigation in preleukaemic cells revealing a sex imbalance in the target cell of mutation. CBA/Ca mice were screened for chromosome 2 deletions after radiation exposure with a myeloid and lymphoid phenotype more common in males compared to females (36% and 22%, respectively) while a lymphoid phenotype was more common in females compared to males (38% and 16%, respectively) (Verbiest et al. 2018). This work suggested the target cell of AML development to be a naïve hematopoietic stem cell or progenitor in male mice, giving rise to myeloid leukaemia and preferentially a common lymphoid progenitor in female mice. The female sex preference for *Flt3*-ITDs reported in this study, however, is not reflected in human de novo AML cases (Shen et al. 2011) and so appears to be a specific pathway present in the CBA/H mouse or possibly a specificity of radiation leukaemogenesis, very difficult to confirm in humans.

Comparison of this study with t-AML patients, however, shows a similar outcome. In a study of 140 t-AML/t-MDS patients, *FLT3*-ITDs were found in 7% of all cases, 8 out of 10 were females, and *FLT3* mutations were significantly associated with previous radiotherapy only and a normal karyotype (Christiansen et al. 2005). The incidence of *Flt3*-ITD in female t-AML patients and female rAML mice therefore may be a consequence of radiation exposure. Sex-specific gene expression changes in genes coding for growth factors, protein kinases, nuclear DNA-binding proteins, Wnt signalling pathway members have been identified after X-ray irradiation in mice (Kovalchuk et al. 2004; Besplug et al. 2005). Spleen samples from C57BL/6 mice showed a significant upregulation of *tumour necrosis factor receptor superfamily member 5* (*TNFRSF5*), which plays important roles in apoptosis and proliferation, in

female mice and a downregulation in male mice after acute 0.5 Gy irradiation (Besplug et al. 2005). Sex-specific dysregulation of pathways by radiation could result in the selective advantage of certain sub-populations, possibly by sex hormones. The incidence of rAML in mice has previously been seen to occur at a higher rate in males in comparison to females, with the removal of the ovaries increasing the occurrence of rAML and the removal of the gonads slightly decreasing rAML, although these changes are not significant (Upton et al. 1958). Sex differences in regulation of hematopoietic stem cell cycle regulation have been identified with estrogen reportedly increases hematopoietic stem cell self-renewal in female mice (Nakada et al. 2014). Also, as previously reported in human cases, these *Flt3*-ITDs did not occur with *Sfpi1* mutations (Inomata et al. 2006), confirming that both seem to be exclusive to each other and this further establishes the mechanism of AML development between human and mouse AML.

In this study, three *Kras* mutations, consisting of two G12D mutations and one G12R mutation, were identified occurring in 2% of all cases which is within range of the rates in human AML of 1.5%-9% (Tyner et al. 2009) (Stirewalt et al. 2001; Neubauer et al. 1994; Illmer et al. 2005). Two cases had a *Kras* G12D mutation which has been previously reported in human AML (Neubauer et al. 1994; Tyner et al. 2009) and mouse lymphomas (Guerrero et al. 1984). $Kras^{G12D}$ mice have been shown to develop a lethal haematopoietic disease characterised by leucocytosis, splenomegaly and increased leukaemic blasts in the peripheral blood and bone marrow (Kelly et al. 2019). A novel amino acid change of Gly→Arg was also found in one mouse. Similarly, in humans a G12D mutation has been the most commonly reported (Tyner et al. 2009; Neubauer et al. 1994), with a G12R mutation being less frequently mutated (Stirewalt et al. 2001). To our knowledge, this is the first time a *Kras* G12R mutation, previously reported in human AML (Bolouri et al. 2018), has been reported in the mouse.

The frequency at which these mutations occurred, as detected by pyrosequencing, gives us an insight into the development of rAML. The *Sfpi1* mutation at codon R235 occurred at a high frequency in the sample and so appears to be one of the initiating mutations. This codon is located in the DNA binding domain and it is assumed that a mutation, or deletion, results in a block in myeloid development producing immature blast cells, thought of as the first hit in AML. The *Kras* G12 mutation occurs at a lower frequency and so is predicted to occur later in leukaemogenesis. This mutation is an

activating mutation and so gives the immature blast cells a growth advantage. The RAS proto-oncogene belongs to a small family of GTPases which function in signal transduction pathways. Mutations in codons 12, 13 and 61 can result in an amino acid change which leads to a resistance of the protein to GTPase-activating proteins, leaving it bound to GTP and so remaining in an active state. Mutations in the G12 codon up-regulate mRNA and protein expression in the bone marrow (Zhao et al. 2014;Guerrero et al, 1984) with a glycine to aspartic acid (G→D) mutation being the most common (Stirewalt et al. 2001; Neubauer et al. 1994). *RAS* mutations are not considered to be the initiating event in radiation leukaemogenesis, as mice expressing oncogenic RAS rarely develop AML spontaneously (Li et al. 2011) and mutations have only been found in murine leukaemic cases with an overt phenotype, not in cases with limited accumulation of leukaemic blast cells in tissue (Rithidech et al. 1996). *Kras*^{G12D} has been shown to co-operate with other mutations, such as loss of *Dnmt3a* to progress leukaemia development (Chang et al. 2015). Here, we illustrate the clonal expansion of AML through acquisition of a *Sfpi1* R235 driver mutation followed by a secondary *Kras* G12 mutation which probably contribute to leukaemia progression.

Previous work has shown that reduced expression of *Sfpi1* in HSCs and progenitors down to 20% of wild type levels will result in the development of AML within 3-8 months (Rosenbauer et al. 2004). Transcriptional analysis of rAML samples showed a significant reduction of *Sfpi1* expression specifically in AML cases with an absence of a R235 mutation. Transcriptional reduction of *Sfpi1* expression, a necessary step in this model, could therefore be driving AML development in these cases without a R235 mutation. PU.1/Sfpi1 is a myeloid transcription factor responsible for development of the myeloid lineage and its expression is tightly regulated during differentiation. PU.1 functions as a master regulator of hematopoiesis, controlling HSC levels through the transcription of multiple cell-cycle regulators (Staber et al. 2013). PU.1 is expressed at a low level in HSCs but its expression increases during myeloid differentiation, particularly in granulocytes (Cheng et al. 1996; Chen, Zhang, et al. 1995). PU.1 binds to the granulocyte-macrophage colony-stimulating factor (GM-CSF) receptor α promoter and, together with C/EBP α, regulates the expression of GM-CSF in myeloid development (Hohaus et al. 1995). Pu.1 and C/EBP α also bind to the granulocyte colony-stimulating factor (G-CSF) receptor (Smith et al. 1996) and the macrophage colony-stimulating factor (M-CSF) (Zhang et al. 1994). PU.1 also interacts with other

co-factors such as Satb1 and Runx1 to regulate early T cell development (Hosokawa et al. 2018). A decrease in PU.1 levels, by deletion of the URE, has been shown to lead to malignant myeloid transformation (Rosenbauer et al. 2004). Transcription factor levels is therefore thought to determine cell fate, with normal levels leading to differentiation of cells and reduced levels leading to a preleukemic phase of poor differentiation before developing into AML (Rosenbauer et al. 2005).

A low level of PU.1 is thought to be required to maintain the basic stem cell functions of HSC such as self-renewal and differentiation. In a RUNX1/Eto9a-dependent leukemia mouse model PU.1 levels were decreased which resulted in a delayed onset of leukemia (Staber et al. 2014). Also, in mixed lineage leukemia (MLL), PU.1 is involved in crosstalk with MEIS1 and HOXA9 to drive an aggressive form of leukemia (Zhou et al. 2014). In cells from MOZ-TIF2-induced leukemia, PU1. Is essential in maintaining the AML phenotype by interacting with MOZ-TIF2 to stimulate expression of macrophage colony-stimulating factor receptor (CSF1R) (Aikawa et al. 2010). A reduced level of PU.1 expression, therefore, appears to play an important role in leukemogenesis. A significant down-regulation of *Sfpi1* has also been reported by Salemi et al. 2015 in *Flt3* mutated cases of human AML (Salemi et al. 2015). In our *Flt3*-ITD samples, 2 cases had a low expression of *Sfpi1* while 2 cases had a high expression of *Sfpi1* in comparison to wild-type levels. This difference in transcriptional expression may be due to the CBA/H and F1 CBA/H x C57BL/Lia strain differences between samples or perhaps be due to a targeting of different downstream pathways by *Flt3* involving yet more pathways of development. For samples that do not have a reduced expression of *Sfpi1*, including two *Flt3*-ITD samples, it is possible that myeloid development is blocked by affecting a different myeloid factor, although in this mouse model it would be rare. In these *Flt3*-ITD samples, *Flt3* may not function to decrease *Sfpi1* but instead target signal transducer and activator of transcription 5 (*Stat5)* or *Ras* to produce abnormal cell growth. Although human AML may have a low rate of *PU.1* mutations, the gene can be targeted for disruption by an alternative route such as *miR-155*, *SPI1* URE methylation (Verbiest et al. 2015), and therefore, as illustrated in this study, further establish the link between human and mouse AML. An increase in *FLT3* gene expression has been previously reported in *FLT3* mutated human AML samples (Carow et al. 1996; Ozeki et al. 2004). The presence of the ITDs disrupt the negative regulatory function of the juxta membrane region, resulting in phosphorylation

of the receiver and an increase in tyrosine kinase activity (Kiyoi et al. 1998). Here, as in human AML, we also report and confirm an increase in *Flt3* expression in all 4 rAML samples with an ITD. *FLT3* has many downstream targets, including *PU.1,* and there is work to indicate an inverse relationship between them, however it is unclear in this study (Inomata et al. 2006).

Several epigenetic mechanisms, such as the role of miRNA, were investigated in this study in order to identify miRNA of interest in rAML development. We identified *miR-1983*, *miR-582-5p* and *miR-467c* which all show different levels of expression among the rAML mutated sub groups. *miR-1983* showed a significant up regulation in *Sfpi1* R235 mutated samples. This increase, however, may be due to hypokalaemia, which is a result of the bloods potassium levels being too low. Hypokalaemia is frequently seen in human AML cases and can result in a decrease in the hormone alderstone, a regulator of *miR-1983* (Edinger et al. 2014). Both up and down-regulation of *miR-582-5p* has been reported in a number of human cancers, however the information regarding *miR-582-5p* in leukaemia is limited with a study by Schotte et al. (Schotte et al. 2009) reporting up-regulation of *miR-582-5p* in ALL while a study by Zhang et al. (Zhang et al. 2009) reporting downregulation of *miR-582-5p* in ALL and AML cases. *miR-582* is reportedly up-regulated in colorectal cancer (Li and Ma 2018) and promotes cancer progression by decreasing expression of the tumour-suppressor *phosphatase and tensin homolog* (*PTEN*) (Song et al. 2017). miR-582-3p has been shown to reduce mRNA and protein levels of negative regulators of the Wnt/β-catenin signalling pathway such as *axin 2* (*AXIN2*), *dickkopf WNT signaling pathway inhibitor* (*DKK)* and *secreted frizzled related protein 1* (*SFRP1)* in human lung cancer cells resulting in the up-regulation of Wnt pathway downstream targets (Fang et al. 2015). The high expression of miR-582-5p in samples with a *Sfpi1* R235 mutation may indicate that the Wnt pathway is activated in these AML cases. For AML development, the link of 582-3p and 582-5p to the Wnt/β-catenin signalling pathway in human lung cancer (Jin et al. 2017) is of particular interest as activation of the Wnt/β-catenin signalling pathway is essential in the development of leukaemic cells (Staal et al. 2016; Wang et al. 2010).

miRNA analysis of *miR-155* was also analysed as it has been previously reported to directly target the gene *Sfpi1/Pu.1* (Vigorito et al. 2007). It was suggested to have an oncogenic role being upregulated in B cell lymphoma (Eis et al. 2005), Hodgkin

lymphoma (Kluiver et al. 2005) and AML carrying *FLT3*-ITD mutations (Faraoni et al. 2012) but more recently, evidence of an anti-leukaemic role has been also identified by induction of apoptosis and differentiation (Palma et al. 2014). An increased expression of *miR-155* has been reported in human AML samples with an *FLT3*-ITD (Salemi et al. 2015) and *FLT3*-ITDs have been shown to upregulate *miR-155* expression by approximately 10-fold in a murine bone marrow cell line (Gerloff et al. 2015). However, an increase of *miR-155* in *Flt3*-ITD samples was not observed here. An up-regulation in *miR-155* expression, although not significant, was seen in *Sfpi1* R235 mutated samples in comparison to *Sfpi1* wildtype samples. Therefore, as a negative regulator of *Sfpi1* expression, this microRNA may have a role in the down regulation of *Sfpi1* expression in R235 mutated samples in this study, however, it does not appear to be the main cause.

When looking in more detail at global miRNA expression among *Sfpi1* mutated and *Flt3*-ITD AML groups, the expression of two microRNA *582-5p* and *467c* showed distinct patterns among the two groups. Both microRNA showed significant down regulation in samples with an *Flt3*-ITD and significant up-regulation in samples with a *Sfpi1* R235 mutation in comparison to the control (Figure 31). These microRNA appear to be dysregulated in rAML but by different methods. *miR-582-5p* has been reportedly both up and down regulated in human leukaemia cases which may be representative of different pathways of development.

As previously mentioned, the high expression of *miR-582-5p* in *Sfpi1* R235 cases could be activating the Wnt pathway through repression of Wnt negative regulators. Although surprising, the low expression of *miR-582-5p* in rAML cases with an Flt3-ITD cases could be acting in a different pathway. Forkhead box c1 (*Foxc1)* is a target of *miR-582-5p* and so a downregulation of *miR-582-5p* could cause an increase in expression and activity of *Foxc1*. Upregulation of *FOXC1* has been reported in human AML cases, specifically in those with an *Flt3*-ITD, and has been shown to collaborate with *Hoxa9* to block macrophage differentiation and enhance clonogenic potential (Somerville et al. 2015). As of yet, little has been reported on *miR-467c* expression in cancer, with studies reporting an increase in *miR-467c* expression in tuberous sclerosis (Cai et al. 2017) and a link to apoptosis (Druz et al. 2013), however studies are limited.

To assess if other epigenetic modifications could be responsible for the overall lower expression of *Sfpi1* in samples without a R235 mutation, DNA methylation analysis was also performed as previous studies have demonstrated a link between DNA methylation and *PU.1* transcriptional expression. Treatment of a T-cell line, which does not express *PU.1*, with the demethylating agent 5-azacytidine resulted in the RT-PCR expression of *PU.1* (Amaravadi and Klemsz 1999). An increase in expression of PU.1 was also seen after treatment with 5-Aza-2'-deoxycytidine in murine erythroleukaemia (MEL) cells (Fernandez-Nestosa et al. 2013) and in human myeloma cells (Tatetsu et al. 2007). These studies suggest that DNA methylation is controlling the expression of *PU.1* and may be playing a part in the down-regulation of *PU.*1 in oncogenesis. Expression of *Sfpi1/Pu.1* is maintained by a promoter and by a highly conserved upstream regulatory element (URE), a kb -14 site in mice (Li et al. 2001) and a kb-17 site in humans (Tatetsu et al. 2007). Binding sites for *PU.1* have been demonstrated in both the *PU.1* promoter and URE, illustrating a possible autoregulation loop (Okuno et al. 2005; Chen, Ray-Gallet, et al. 1995). Epigenetic analysis has revealed the promoter and -17 kb upstream regulatory element of the *PU.1* gene to be highly methylated in human classical Hodgkin lymphoma cells (Yuki et al. 2013), human myeloma cells (Tatetsu et al. 2007) and MEL cells, resulting in a block in cell differentiation (Shearstone et al. 2011; Fernandez-Nestosa et al. 2013).

Here, DNA methylation analysis of the 4 CpG sites in the promoter region and 5 CpG sites in the URE also showed up-regulation in all rAML CpG sites in comparison to the bone marrow control. A significant up-regulation in DNA methylation was evident in all CpG sites in rAML samples unmutated for *Sfpi1* R235 in comparison to R235 mutated samples. A correlation between DNA methylation and *Sfpi1* expression was strongest at URE CpG 1, CpG 2, CpG 3, CpG 4 and promoter CpG 3. These CpG sites are possibly the most important if methylated as they correlate strongest to transcriptional expression. The URE is possibly more important than the promoter with 4 of its CpG sites showing a strong correlation to gene expression in comparison to one CpG in the promoter. *In vitro* work has shown the URE to increase *PU.1* promoter activity specifically in stably transfected human and murine myeloid cell lines (Li et al. 2001). The URE may be specifically targeted for DNA methylation during AML development as the URE increases promoter activity in myeloid cells. The URE is necessary for *PU.1* expression with a previous study showing that disruption of the URE leads to an

80% decrease in PU.1 gene expression in murine bone marrow (Rosenbauer et al. 2004). In human AML erythroleukaemia cell lines, recent work has highlighted the importance of the URE as a binding site for *GATA-1* which recruits *DNMT1* to regulate *PU.1* expression (Burda et al. 2016). This mechanism, however, was not seen in murine erythroleukaemia cases. Analysis of human myeloma cell lines revealed three out of four cell lines with transcriptional down-regulation of *PU.1* also had very highly methylated 17-kb upstream regions, while a cell lines with up-regulation of *PU.1* had no DNA methylation. In contrast, the promoter region had varying levels of DNA methylation across all cell lines (Tatetsu et al. 2007). The 17-kb upstream region, therefore, shows a strong correlation to *PU.1* transcriptional expression and may have a greater influence than the promoter.

Importantly, in case 3, the methylation level was very high at all CpG sites, reaching 91%. Correlation analysis revealed an inverse relationship between DNA methylation and *Sfpi1* expression in a few of the CpG sites. It seems that the high level of DNA methylation, particularly in case 3, repressed *Sfpi1* expression, showing for the first time that a promoter methylation can be driving AML development, specifically in the absence of a *Sfpi1* R235 mutation. Although a rare event, the development of AML in case 3 clearly seems to involve *Sfpi1* disruption through promoter DNA methylation and transcriptional repression followed by a *Kras* G12R mutation, reported for the first time in the CBA/H mouse (Figure 38).

To summarize, we confirmed the main pathway and identified new pathways of radiation leukaemogenesis (Figure 39). The major pathway consists of a previously described chromosome 2 deletion with a *Sfpi1* R235 mutation. A sub-pathway reveals that AML can develop where there is a chromosome 2 deletion and significant increase in *Sfpi1* DNA methylation with reduction of *Sfpi1* transcriptional expression in the absence of a R235 mutation on the second *Sfpi1* allele. A newly identified minor pathway neither requires chromosome 2 deletion nor *Sfpi1* R235 mutation, but rather a significant increase in *Sfpi1* DNA methylation with reduction of *Sfpi1* transcriptional expression and this occurs predominantly in female cases. Minor pathways include *Flt3*-ITD, Sfpi1 R235 or *Kras* G12 mutations. The cases of *Flt3*-ITD clearly appear to have a sex bias with all cases so far, 4 seen here a fifth case reported in Finnon et al. but not analysed here, presenting in females. The minor pathway presenting in case 3 represents a novel pathway consisting of site specific DNA methylation associated

with reduced *Sfpi1* transcription and a *Kras* G12 mutation. For the first time we identified a case where an epigenetic mechanism can be the first driving modification followed by a *Kras* G12 mutation. Overall this work provides new insight into the pathways leading to rAML development, involving genetic mutations as well as epigenetic changes and, for the first time, specific gene DNA methylation.
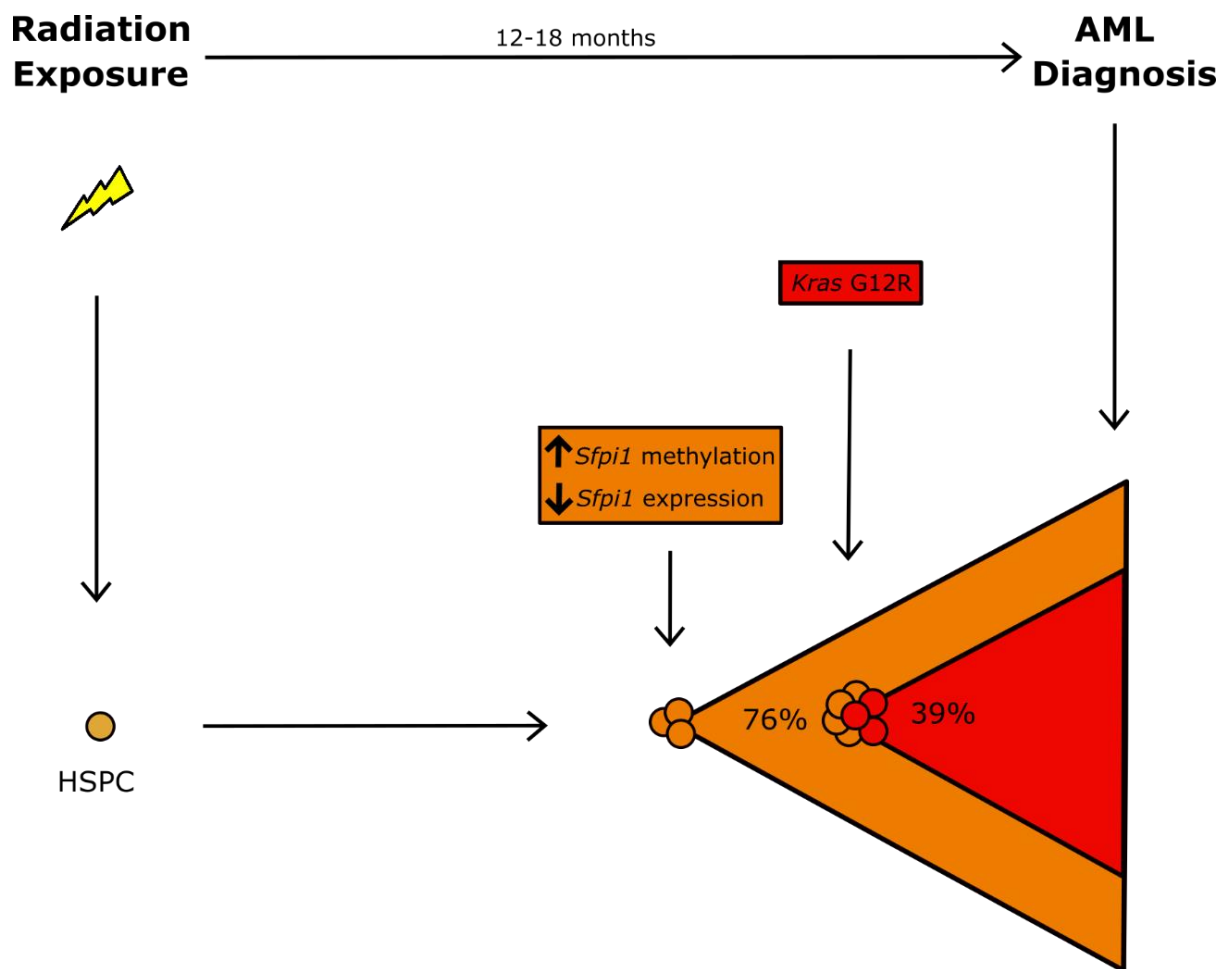
Figure 38. Proposed model of leukaemogenesis for case 3.

After exposure to irradiation in a hematopoietic stem or progenitor cell, an increase in *Sfpi1* URE and promoter methylation occurred in 76% of the sample, leading to a reduced expression of transcriptional *Sfpi1*, producing immature blast cells. At a later stage a G12R mutation in *Kras* occurred, present in 39% of the sample, giving the immature blast cells a growth advantage leading to full AML development.
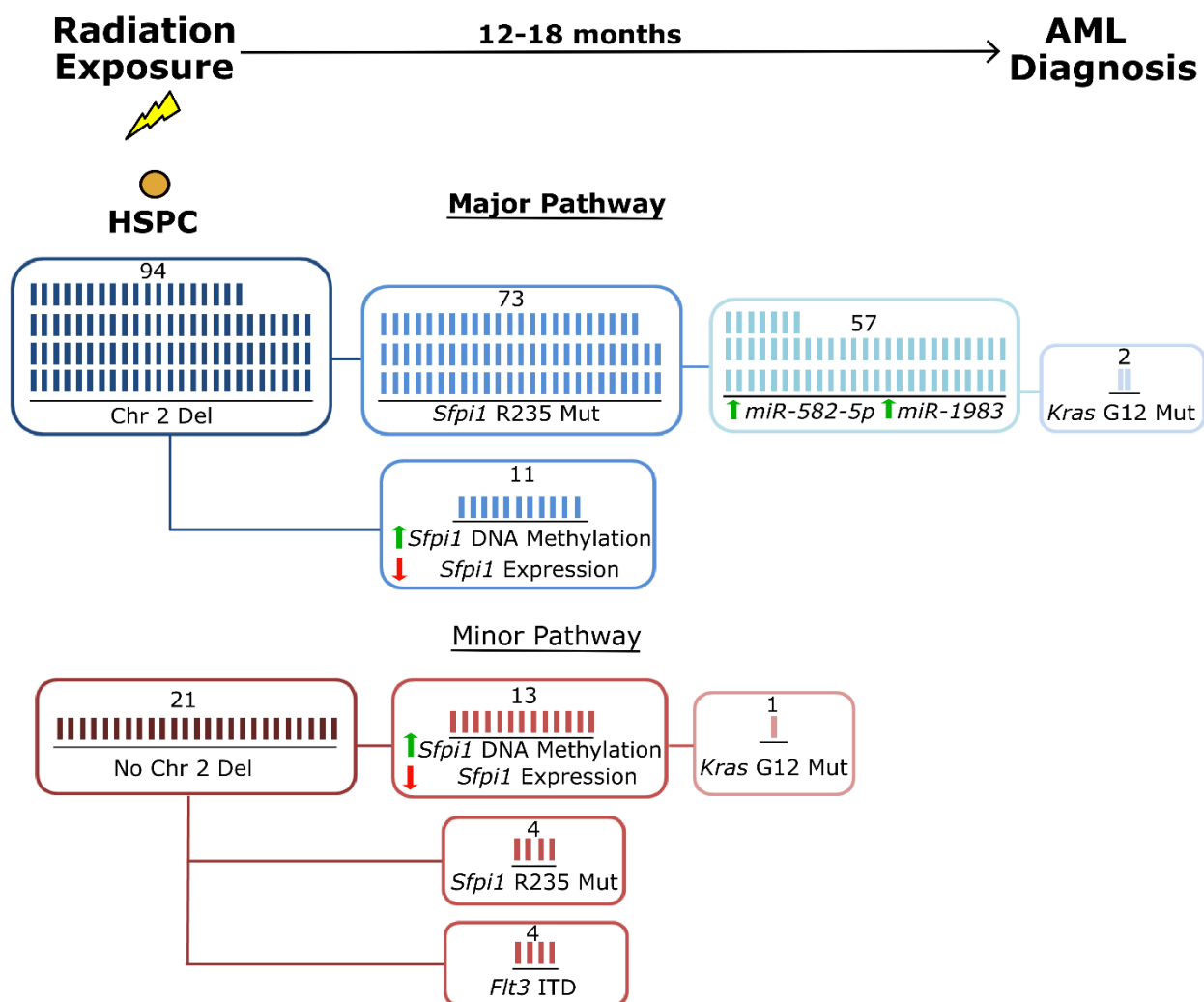
Figure 39. Genetic and epigenetic pathways of radiation-induced AML in CBA mice.

After irradiation exposure, there appears to be two main pathways of rAML development. The major pathway consists of a chromosome 2 deletion with a *Sfpi1* R235 mutation while a minor pathway consists of no chromosome 2 deletion and no *Sfpi1* R235 mutation but instead a significant increase in *Sfpi1* DNA methylation leading to a reduction of *Sfpi1* transcriptional expression. Percentages for chromosome 2 deletion gender data are calculated from 88 mice with a deletion and 14 mice with no deletion where gender was known.

3.3 Mouse HSPC transcriptional analysis

3.3.1 Introduction

Hematopoietic stem cells have been widely studied and characterised over the past few decades and cell surface markers established for isolation of HSC and MPP populations. The expression profiles of HSPCs have been investigated in recent years to further characterise these cells to identify regulatory pathways that control HSC self-renewal, dormancy and differentiation. The expression of genes such as HoxB4 (Antonchuk, Sauvageau, and Humphries 2001), BMI1 proto-Oncogene, polycomb ring finger (*Bmi1*) (Park et al. 2003) and Wnt (Reya et al. 2003) and Notch pathways (Varnum-Finney et al. 1998) have been associated with the renewal of HSCs. Transcriptional profiles in HSPC sup-populations has identified genes associated with quiescence, adhesion and cytoprotection as important for HSCs while genes associated with differentiation, proliferation and chemotaxis are expressed among MPP populations (Forsberg et al. 2005). Previous studies have shown that long-term HSC reside in the BM CD34- fraction (Osawa et al. 1996). Transcriptional profiling of the CD34- and CD34+ fractions of the LSK population has revealed 210 differentially expressed genes between HSC subsets (Zhong et al. 2005). Very few differentially expressed genes were in common with previous transcriptional studies of HSC however this could be due to different cell surface markers used, highlighting the heterogeneity among populations and need for further characterisation.

A large study analysing the proteome, transcriptome and methylome of murine HSCs and MPPs was performed by Cabezas-Wallscheid et al. alongside in vivo reconstitution experiments. Using cell surface markers previously identifying the populations as HSC and MPP1-MPP4 (Wilson et al. 2008), gene ontology analysis revealed transcriptional signatures such as oxidation reduction and response to hypoxia were enriched in HSCs while DNA replication and cell proliferation were overrepresented in MPP1 (Cabezas-Wallscheid et al. 2014). This work illustrates the processes keeping the HSC in a quiescent state while the MPPs proliferative and enter the cell cycle ready for differentiation. Transcriptional expression will provide further information on the regulation of the hematopoietic process but also can identify transcriptional changes in cancer cells, developing a pre-leukemic transcriptional profile.

Gene expression signatures have shown to have prognostic features. Transcriptional analysis of human AML leukemic stem cells and non-leukemic HSCs identified a core transcriptional program in both which associated with clinical outcome (Eppert et al. 2011). Particularly, in cytogenetically normal AML cases high expression of these core genes negatively correlated with complete remission. This is particularly important for CN-AML as they lack cytogenetic prognostic markers and so are classified as intermediate risk even though variations in remission among cases have been seen. This work was further developed by Metzeler et al. in a large study involving 364 cytogenetically normal AML patient samples and combined gene mutations, gene expression and miRNA analysis to develop a signature (Metzeler et al. 2013). A signature including mutations such as *FLT3*-ITD, *WTI* and *RUNX1* mutations, wild-type *CEBPA* and *TET2* and high erythroblast transformation-specific transcription factor ERG (*ERG)*, BAALC binder of MAP3K1 and KLF4 (*BAALC)* and *MIR-155* expression was associated with a low complete remission rate and shorter disease-free survival (Metzeler et al. 2013). Gene expression profiling of therapy-related AML patient samples has identified subtypes within this AML patient group. CD34+ hematopoietic stem cells were analyzed by Affymetrix arrays using probes for 12,600 genes and identified 61 genes which could separate the patients into 2 major groups based on gene expression levels (Qian et al. 2002). These groups also had different chromosomal aberrations, with chromosome 7 deletions in group A and complex karyotypes in group B, identifying potentially separate pathways of AML development characterised by mutations affecting different genes functioning differentiation and proliferation. Patient prognosis was not investigated in this study. Recent work has tried to establish a genetic profile between AML subgroups. A custom AML-array was designed to try to distinguish transcriptional profiles between AML subgroups M1 and M2 (Handschuh et al. 2018). Although 83 genes were differentially expressed between the AML patients and healthy normal patients, a signature was not identified between AML sub-types, possibly due to the pre-selection of genes on the custom set and the fact that these sub-types are closely related.

Despite progress, isolation of a sufficient amount of murine hematopoietic stem cells for molecular studies remains a challenge due to the low abundance of these cells, estimated at just 5000 cells per mouse (Mayle et al. 2013). Isolation of HSPC sub-populations can be achieved through the use of cell surface markers in flow cytometry,

first by lineage depletion of differentiated cells and then by isolation using markers known to be expressed on HSPCs such as c-Kit+ and Sca-1+ (Spangrude, Heimfeld, and Weissman 1988). Other cell surface markers can also be included, such as SLAM family receptors, which have been shown to further separate hematopoietic stem cells from progenitor cells (Kiel et al. 2005). Studies involving the isolation of HSCs using these markers have mainly been carried out using C57BL/6 mice. However, studies using the CBA mouse model are even more challenging due to the weak expression of Sca-1 surface marker in this mouse strain allowing isolation of only very small numbers of HSCs (Spangrude and Brooks 1993).

In order to avoid this problem, the use of alternative cell surface markers CD27 and CD201 have been investigated. These markers allow the isolation of a much larger number of HSCs, demonstrated from transplantation assays into lethally irradiated recipients and these markers are expressed even following hematopoietic injury (Vazquez, Inlay, and Serwold 2015). Here the first aim was to develop a method of analysing transcriptional changes in HSPCs at different time-points after irradiation. Once a method was established, the second aim was to analyse the transcriptional changes after irradiation in the radiation sensitive AML mouse model CBA/Ca, and the non-radiation sensitive C57BL/6 mouse, to identify long-term transcriptional markers of radiation exposure and strain differences.

3.3.2 Hematopoietic stem and progenitor cell numbers

The number of cells in HSPC populations using the markers CD27 and CD201 were first determined. CBA/Ca mice were irradiated with 0 and 1 Gy and sacrificed at 1 week. This time-point was chosen as it is well known that irradiation results in bone marrow depletion in the days following radiation for high doses and by 1 week bone marrow numbers are starting to return to original levels after high doses of radiation (Oben et al. 2017). A dose of 1 Gy was chosen as it is a dose high enough to cause AML development in CBA mice while sparing enough cells allowing stem cell studies after irradiation. Male mice were used for all HSPC cell sorting experiments.

Using the cell surface markers CD27 and CD201, the cell numbers for CD27-, CD201-, CD27+, CD201-, CD27-, CD201+ and CD27+, CD201+ populations were determined

from approximately 25 x 10^6 lineage depleted bone marrow cells from a combination of 3 CBA/Ca mice (Figure 40). Each population had different numbers of cells. The largest cell population sorted, at approximately 250,000 cells was for the CD27-CD201- population, followed by the CD27+CD201- population which isolated around 140,000 cells. The lowest numbers of cells obtained was from the CD201+CD27- population, amounting to 3,000 cells per mouse while for the CD201+CD27+ population, thought to contain the long-term HSCs, the cell number was higher at approximately 30,000 cells. After radiation exposure however, this number had decreased to approximately 7,000 cells. The cell numbers are below the recommended cell number for many RNA extraction kits so testing of RNA extraction kits specific for low cell numbers and pre-amplification kits is therefore required.
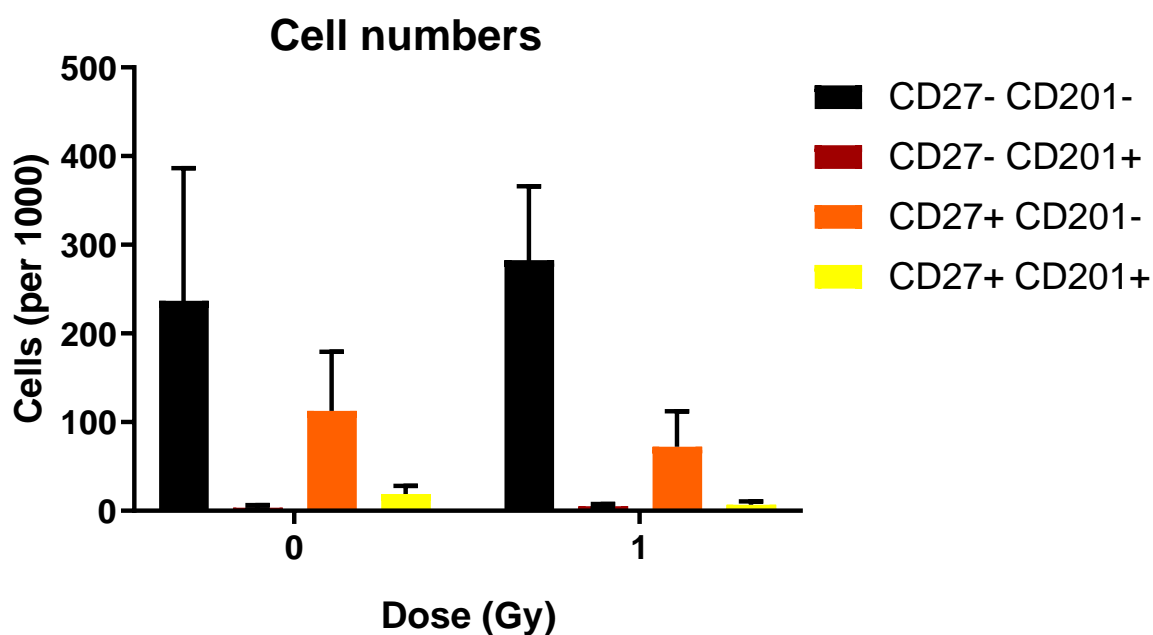
Figure 40. Sorted cell numbers of lineage depleted bone marrow in CBA mice using the cell surface markers CD201 and CD27.

CBA/Ca mice were irradiated in vivo with doses of 0 Gy and 1 Gy and sacrificed 7 days after IR with 1 mouse per dose. Error bars represent the standard deviation of cell numbers from 4 experiments.

3.3.3 RNA extraction

Due to the low numbers of cells isolated using the CD27 and CD201 cell surface markers and, in order to isolate the maximum amount of RNA for several downstream applications, different RNA extraction kits were tested. The miRNeasy Kit from Qiagen was tested for the isolation of both mRNA and miRNA from the same sample. Two extraction kits, Single Cell RNA Purification Kit from Norgen Biotek Corp and ReliaPrep^{TM} RNA Miniprep Systems from Promega, were tested for RNA extraction from very small numbers of cells. The miRNeasy kit states that the minimum amount of starting material is 100 cells depending on the cell type but advises to start with 3-4 x10^6 cells. The ReliaPrep^{TM} RNA Miniprep Systems from Promega state that it extracts mRNA from 100 to 5 x10^6 cultured cells, however, this does not include miRNA. The Single Cell RNA Purification kit states that it extracts total RNA including miRNA, from 1 to 200,000 cells.

miRNeasy RNA Extraction

RNA was extracted from 4 cell quantities 1,000,000, 500,000, 50,000 and 10,000 cells using the miRNeasy RNA Extraction Kit (Table 10). The kit produced good quality RNA with RIN values of 9.8 to 10 for all samples. The quality could only be assessed for samples that had a high concentration within range of the Tapestation (25-100ng/µl) (Figure 41). At 50,000 cells or below, the RNA was not at the concentration required for reliable reading on the Tapestation or on the Nanodrop. The total amount of RNA extracted of 50 ng from 10,000 cells was also at the limit of reverse transcription for many kits. The High Capacity Reverse Transcription Kit claims to be able to perform RT down to 20 ng, however low amounts of mRNA starting material can lead to high Ct values and a lack of reproducibility (Stahlberg et al. 2004).

With a minimum expected yield of a few thousand cells per mouse per population using the CD27 and CD201 cell surface markers, the RNA yield using the miRNeasy RNA Extraction kit would be insufficient. Therefore, pooling of a larger number of mice would be necessary to use this kit.

Cell No.	Total RNA (ng)	RIN	Tapestation Lane
1,000,000	985	9.2	A1
500,000	540	9.5	B1
50,000	192	9.5 !	C1
10,000	50	9.7 !	D1

Table 10. RNA quantity from different HSPCs cell numbers extracted using the Qiagen miRNeasy RNA Extraction kit.

RIN values and corresponding lane numbers after Tapestation analysis are also provided. Exclamation marks denote lanes where the sample concentration is below the recommended range for the Tapestation.
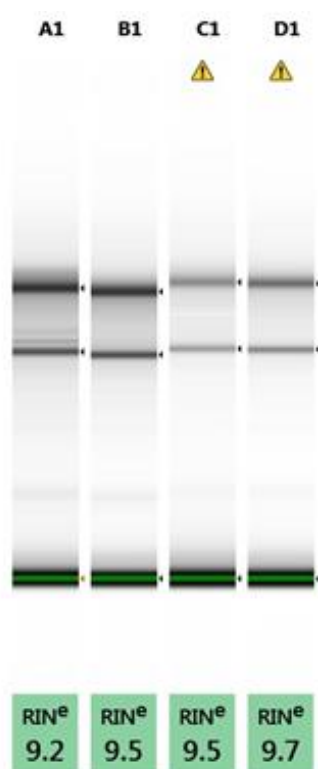
Figure 41. RNA screentape analysis by the Tapestation of samples of different cell numbers extracted using the miRNeasy kit.

Lane sample details are given in Table 10. The yellow triangle with exclamation mark indicates the sample concentration is below the recommended range for the Tapestation.

ReliaPrep^{TM} RNA Miniprep Systems

The ReliaPrep kit from Promega was tested alongside the Single Cell RNA Purification Kit from Norgen as kits for extraction of RNA from lower numbers of cells. The RNA quality again showed high RIN values which was out of range of the Tapestation and the quantity of RNA extracted was still at the limit for downstream applications (Table 11).

Cell No.	Total RNA (ng)	RIN	Tapestation Lane
20,000	72	9.9 !	A1
10,000	51	9.8 !	B1

Table 11. RNA quantity from different HSPCs cell numbers extracted using the ReliaPrep^{TM} RNA Miniprep Systems.

RIN values after Tapestation analysis are also provided. Exclamation marks denote lanes where the sample concentration is below the recommended range for the Tapestation.

Single Cell RNA Purification Kit

Extraction of RNA using a specified kit for low cell numbers, the Single Cell RNA Purification Kit, did give a slightly higher yield of RNA and the highest yield overall (Table 12). These samples had high RIN values of 9.9 to 10 showing that this kit produces good quality RNA, although it was again out of the optimum range of the Tapestation system.

Cell No.	Total RNA (ng)	RIN	Tapestation Lane
20,000	89	10 !	A1
10,000	65	9.9 !	B1

Table 12. RNA quantity from different HSPCs cell numbers extracted using the Single Cell RNA Purification Kit.

RIN values after Tapestation analysis are also provided. Exclamation marks denote lanes where the sample concentration is below the recommended range for the Tapestation.

As the Single Cell RNA Purification Kit produced the highest RNA yield with good quality, it was therefore used to extract RNA from all cell populations using the markers CD27 and CD201 (Figure 42). The Single Cell RNA Purification Kit extracted around 200 – 400 ng RNA from the largest cell populations CD27-CD201- and CD27+CD201-. For the smaller cell populations, CD27-CD201+ and CD27+CD201+, the kit extracted less than 100 ng per experiment, with 46 ng being the lowest amount of RNA extracted from a CD27-CD201+ 0Gy sample. Although this approach would avoid the bias introduced with using an amplification kit, the RNA extracted using this kit could result in insufficient material for the project. In order to obtain sufficient material more mice would be required, or experiments would need to be repeated. Use of amplification kits would have to be investigated which could provide information using fewer cells and allow rarer HSC populations using the widely studied SLAM markers to be investigated.
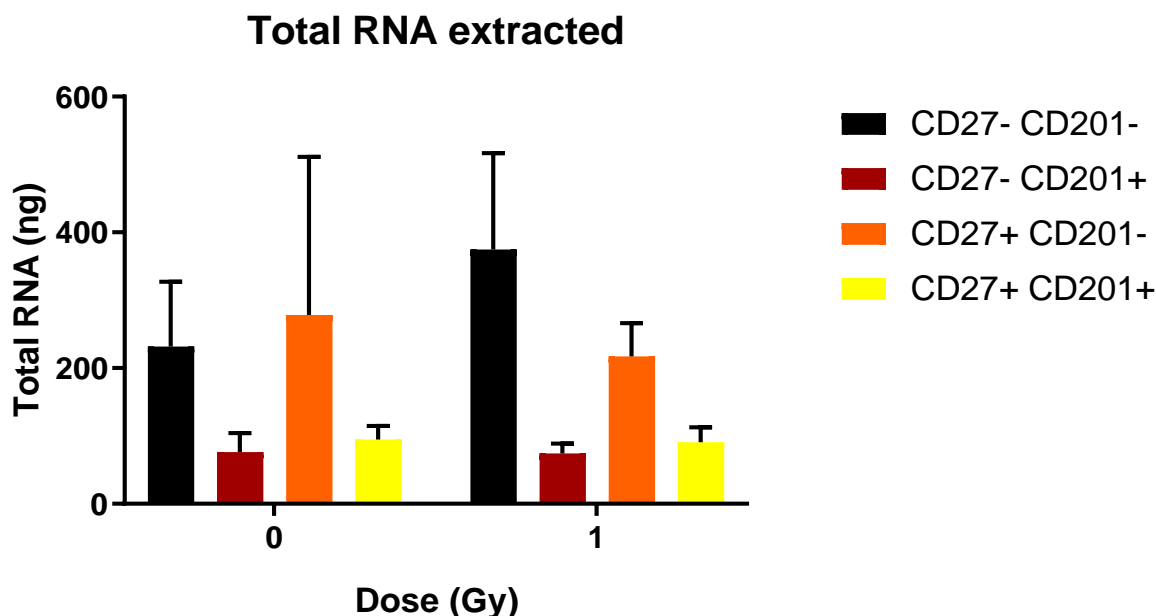
Figure 42. Total RNA extracted from CBA HSPC sub-populations by the Single Cell RNA Purification Kit.

CBA/Ca mice were irradiated with doses of 0 Gy and 1 Gy and sacrificed 1 week after irradiation with one mouse per dose. HSPC sub-populations were sorted using the cell surface markers CD27 and CD201. The data represent the mean standard deviation of four independent biological replicates.

3.3.4 Amplification kits

Due to the low RNA yield obtained using traditional RNA extraction kits, the use of three pre-amplification kits were analysed. The kits included the REPLI-g® WGA & WTA Kit, which can amplify cDNA and DNA from a minimum of 25 cells for use in DNA sequencing and QPCR, the REPLI-g® WTA Kit, which can amplify cDNA from a single cell, and the CellsDirect One Step qRT-PCR Kit, which can also amplify cDNA from a single cell. A previous publication by Moignard et al. has demonstrated amplification of cDNA from a single cell using a modified protocol of the CellsDirect One Step qRT-PCR Kit including a pre-amplification step (Moignard et al. 2013).

The REPLI-g® WGA & WTA Kit was tested with 25 and 100 lineage-depleted bone marrow cells (Figure 43). A positive control sample was included which were cells reverse transcribed into cDNA without pre-amplification. The stated minimum number of cells produced detectable up-regulation, however it varied from 28 - 38 Ct. A larger number of cells, 100 cells, produced a higher Ct value of 26 Ct. Therefore, a larger number of cells than the minimum 25 would be required in each sample to obtain consistent amplification, which would not be feasible for long-term HSPC populations.

To further test the limit of such amplification kits, the REPLI-G® WTA Single Cell Kit, which can amplify cDNA from 1 cell, was tested. MQRT-PCR was performed on these samples with 1 and 10 cells using the housekeeping gene *Hprt*. Amplification was detected from the sample containing 10 cells at 29 Ct and for the single cell samples amplification was detected at 34 – 39 Ct (Figure 44). Samples containing at least 10 cells would be required for use with this kit.

The CellsDirect One Step qRT-PCR Kit was also tested using a modified protocol from Moignard et al. Amplification of single cells was detected in 11 out of 15 single cell samples at 17 – 25 Ct (Figure 45). This kit produced the lowest Ct value out of the 3 kits and the protocol included a stopping point after cell collection, shortening the

length of the total cell sorting protocol to 11 hours in comparison to 15 hours using the REPLI-G® kits. The CellsDirect One Step qRT-PCR Kit using the modified protocol from Moignard et al was therefore selected due to it allowing single cell studies, the flexibility of the stopping point and the shorter cell collection time.
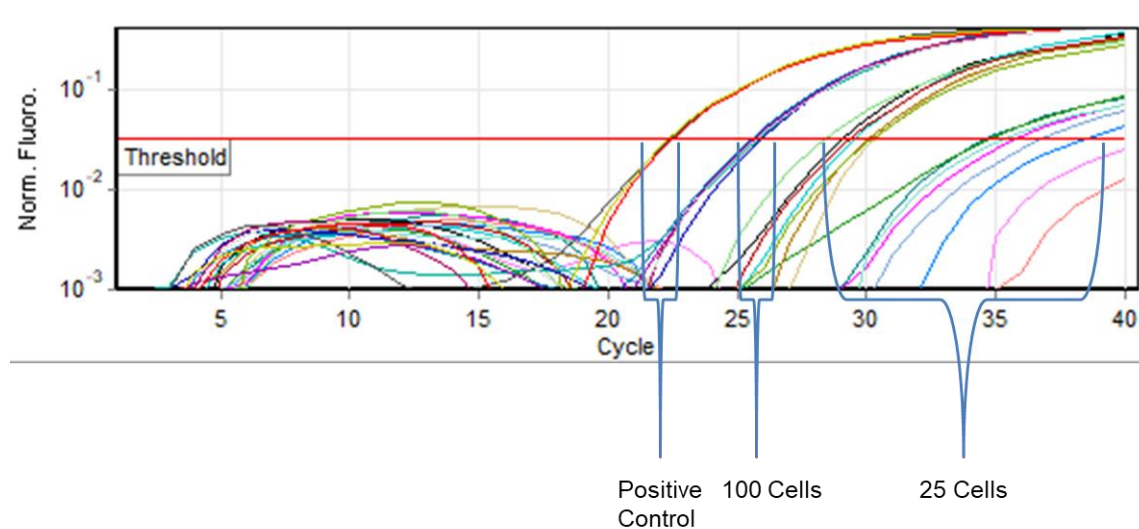
Figure 43. Gene expression of *Hprt* in cells amplified using the REPLI-g WGA & WTA kit.

QRT-PCR amplification plot of CBA/Ca cDNA from 15 samples with 25 cells and 5 samples with 100 cells amplified with the REPLI-g WGA & WTA kit. A positive control sample consisting of RNA extracted with the miRNeasy kit and reverse transcribed using the High Capacity Reverse Transcription Kit is labelled.
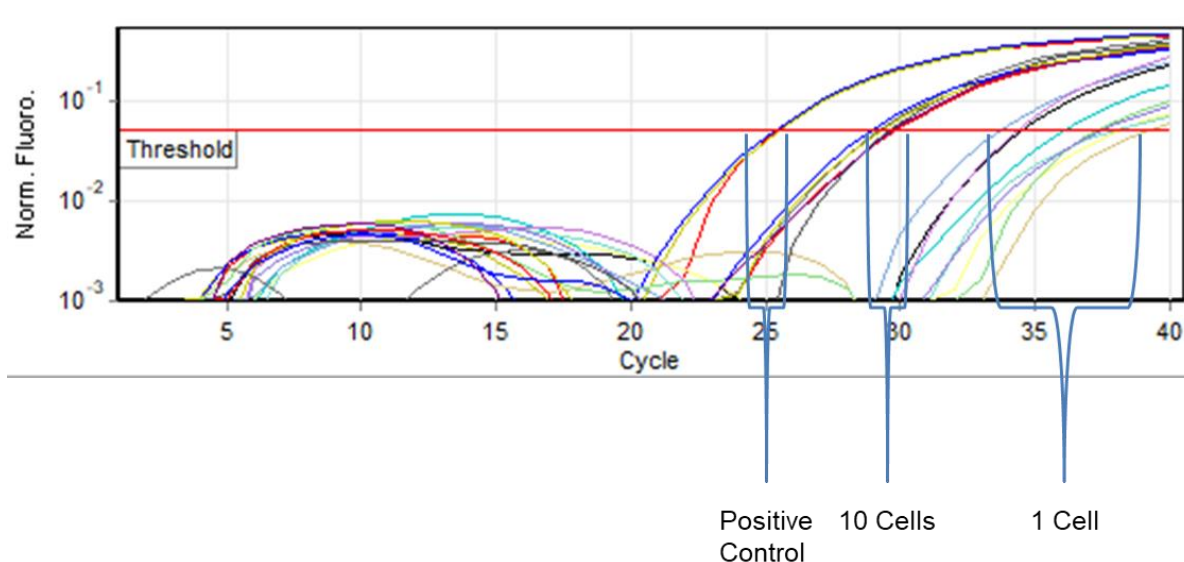
Figure 44. **Gene expression of *Hprt* in cells amplified using the REPLI-g WGA & WTA kit.**

QRT-PCR amplification plot of CBA/Ca cDNA from 5 samples with 10 cells and 10 samples with 1 cell amplified with the REPLI-g WTA kit for the gene *Hprt*. A positive control sample consisting of RNA extracted with the miRNeasy kit and reverse transcribed using the High Capacity Reverse Transcription Kit is labelled.
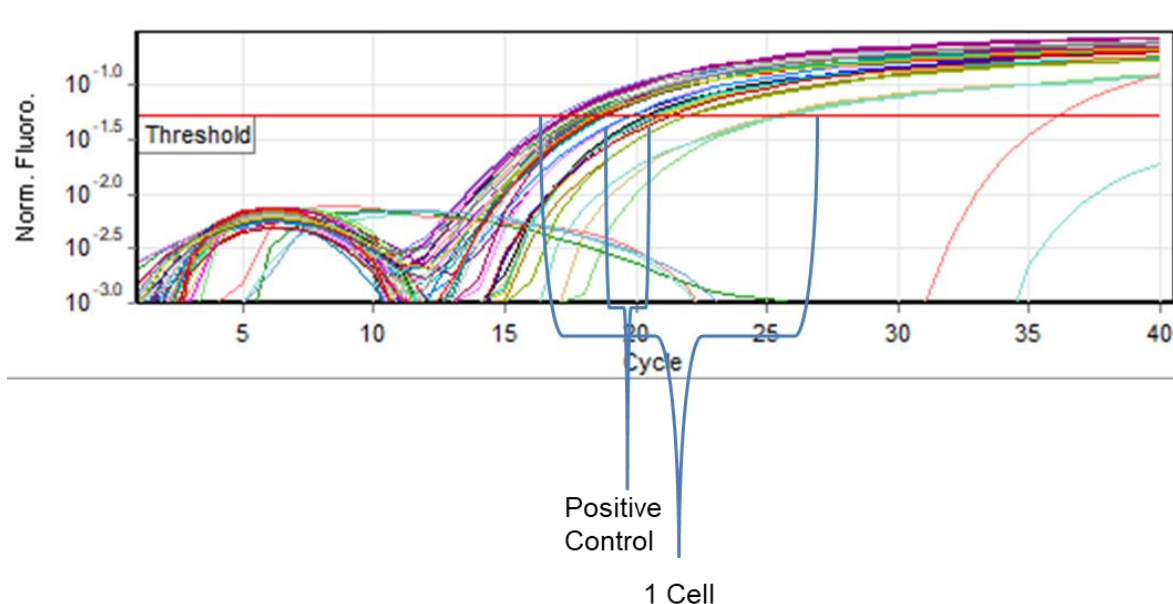
Figure 45. Gene expression of *Hprt* in single cells amplified with the CellsDirect™ kit.

QRT-PCR amplification plot of CBA/Ca cDNA from 15 single cells amplified with the Cells Direct kit including a pre-amplification step as detailed in Moignard et al. A positive control sample consisting of RNA extracted with the miRNeasy kit and reverse transcribed using the High Capacity Reverse Transcription Kit is labelled.

3.3.5 Single cell protocol validation

The selected amplification step involving single cells allows the use of more stringent cell surface markers which isolates smaller, although more distinct, populations of stem cells. A recent paper by Wilson et al. isolated HSCs based on the use of the cell surface markers c-Kit, Sca1, CD48, CD150, CD34 and CD135 which allowed for HSPCs to be separated into 1 HSC and 4 multipotent progenitor populations (Wilson et al. 2008). These populations consist of a HSC population with an LSK CD34- CD48- CD135- CD150+ phenotype. This HSC develops into MPP1 with the acquisition of CD34. The MPP1 then develops into MPP2 upon acquisition of CD48. Loss of CD150 leads to the development into MPP3 and finally acquisition of CD135 leads to the development of MPP4. These markers were used to isolate the 5 HSPC populations as in Wilson et al.

Primer design in single cell experiments must ensure that mRNA, and not DNA, is amplified, which here is crucial as there is no DNase step. For genes of interest, primers were therefore designed to specifically span exon boundaries. Primers were first tested by a SYBR green melt curve analysis to ensure the primers were specific. SYBR green analysis of the *Hoxb5* design is illustrated in Figure 46. Designs were also tested by SYBR green analysis on both cDNA and DNA samples. Primers were selected which showed upregulation in cDNA samples and not DNA samples and, if upregulation was observed in DNA samples, the design was accepted if at least 5 Ct difference was seen between cDNA and DNA samples.

Designs for the genes *Hprt* and *Hoxb5* had a Ct difference of less than 5 Ct between cDNA and DNA amplification (Figure 47). The design for sestrin 2 (*Sesn2)* had a Ct difference of 9 Ct between cDNA and DNA amplification. The *Hprt* and *Hoxb5* designs were further tested on 500 sorted cells which were pre-amplified using the primer designs including minus RT controls to test for DNA amplification (Figure 48). A Ct

difference of greater than 5 Ct was detected, and the designs validated. Furthermore, a minus RT control was also included on each QRT-PCR plate to detect any DNA contamination.
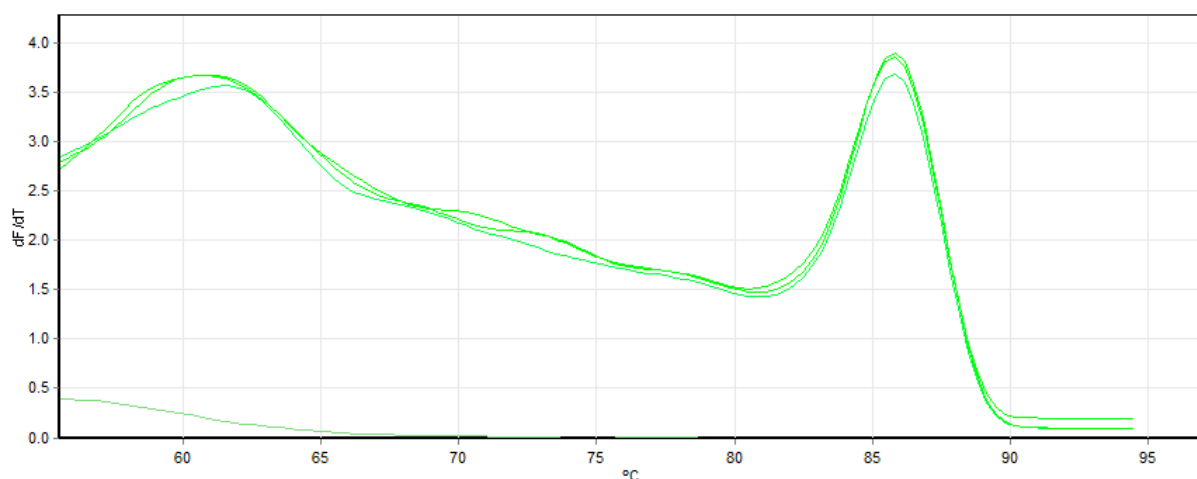
Figure 46. Melt curve analysis of QPCR primers for the gene *Hoxb5*.

The melt curve shows one sharp peak with a melting temperature of 85 °C which means that the primers are specific to the area of interest and do not amplify unspecific products.
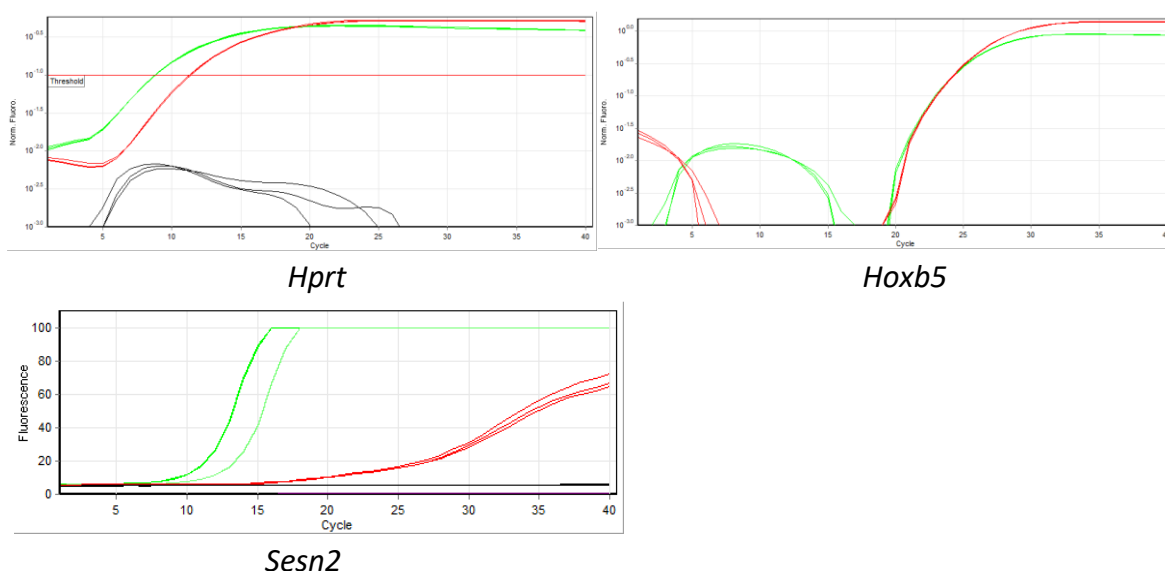
Hprt

Hoxb5

Sesn2

Figure 47. QPCR expression of the genes *Hprt*, *Hoxb5* and *Sesn2* in cDNA and DNA samples.

cDNA samples are labelled in green and DNA samples in red.

Hprt   *Hoxb5*
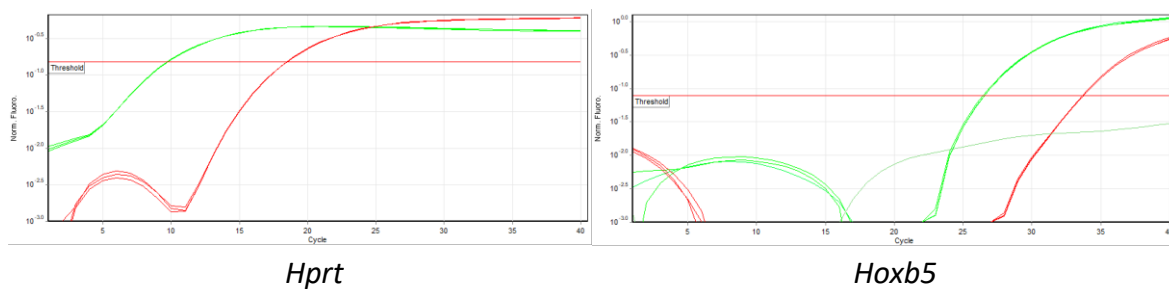
Figure 48. QPCR expression of the genes *Hprt* and *Hoxb5* in cDNA and minus RT controls.

MQRT-PCR expression was pre-amplified from 500 sorted HSPCs using the CellsDirect Kit. cDNA samples are labelled in green and minus RT control samples in red.

In order to determine if a cell was present in the well and calculate the efficiency of the cell sorting protocol, a single cell was deemed present if the expression of *Hprt* could be detected in the well. A total of 30 single cells were used for each QPCR with one well retained as a minus RT control. The number of wells where *Hprt* was detected was measured for both CBA/Ca and C57/BL6 mice (Figure 49). No difference in the number of wells with *Hprt* expression was detected between strains or between sub-populations. Overall a mean of 25 out of 29 wells was determined to have *Hprt* expression.
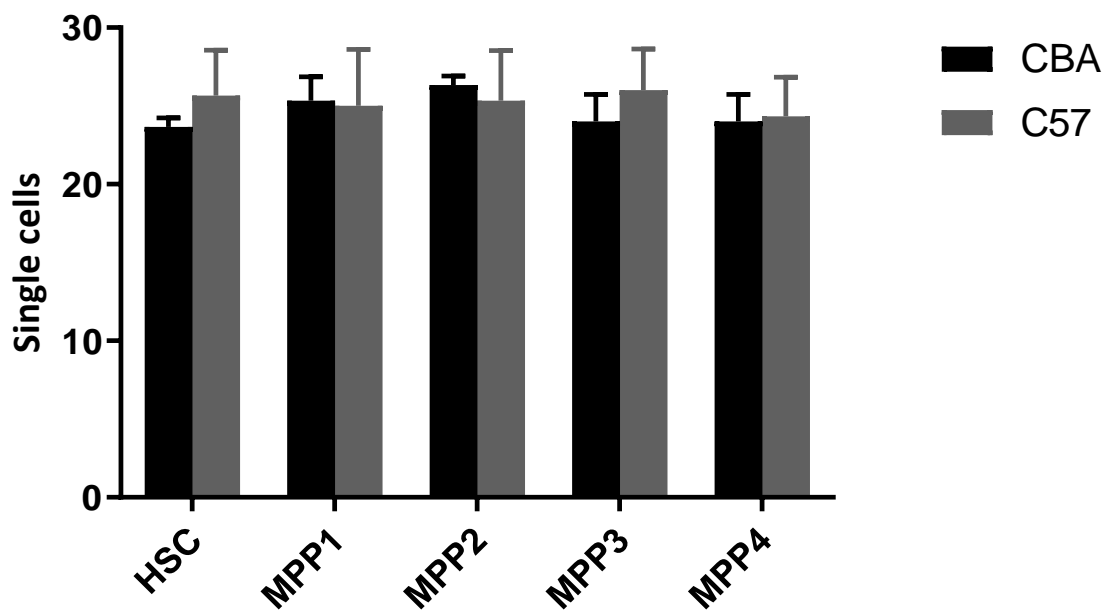
Figure 49. Single cell amplification efficiency.

The number of cell sorting wells in which a single cell was present, as indicated by the expression of *Hprt*, in both CBA/Ca and C57BL/6 strains.

In order to validate that the sorted cells belong to the specified sub-populations, we tested the expression of a known marker for long-term hematopoietic stem cells, *Hoxb5* (Chen, Miyanishi, et al. 2016). Using the same markers as this study, *Hoxb5* has previously been shown to be highly expressed in hematopoietic stem cells while expressed at a much lower level in multipotent progenitors populations (Cabezas-Wallscheid et al. 2014). In CBA/Ca HSPCs using these cell surface markers, expression of *Hoxb5* could be clearly seen in HSCs, even though the number of cells expressing *Hoxb5* was limited (Figure 50). The detection of *Hprt* expression was used to confirm the presence of a single cell. The limited numbers of cells expressing *Hoxb5* are not due to technical issues, as *Hprt* expression was confirmed, but rather could be due to the difference between the previously reported bulk transcriptional expression (Chen, Miyanishi, et al. 2016) and single cell expression. For the remining activating MPP cell populations, there was little to no *Hoxb5* expression, with a significant reduction seen in MPP2. Although the number of cells expressing *Hoxb5* are few, the

expression of *Hoxb5* shows that these cell surface markers are specifically able to isolate HSCs.
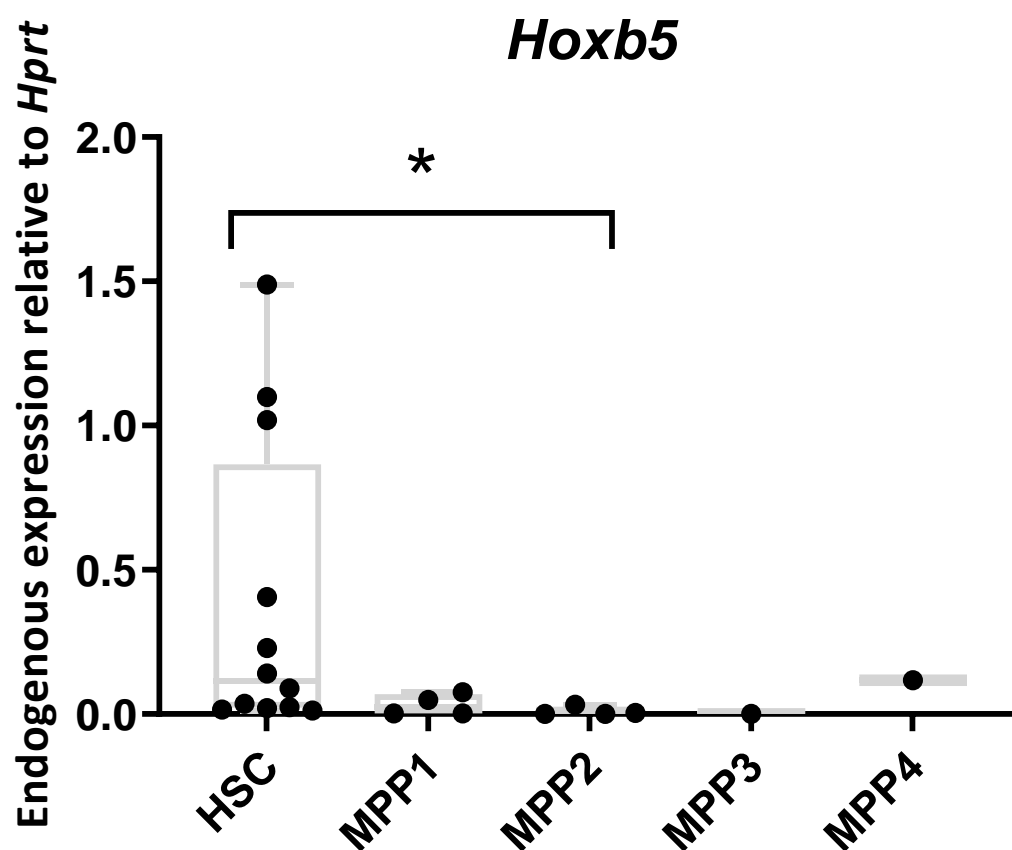
Figure 50. *Hoxb5* expression in CBA HSPCs populations HSCs, MPP1, MPP2, MPP3 and MPP4.

Data represent the endogenous expression of 29 single cells for each population in CBA/Ca mice normalised to *Hprt.* Error bars represent the minimum and maximum values. Significance was calculated by a Mann Whitney test and indicated by an astrix where p≤0.05.

To further validate this single cell protocol and ensure that the expression changes normally detected during QPCR without amplification are still accurately detected in single cells with amplification, QPCR analysis was performed comparing gene

expression changes in total bone marrow in comparison to single cell amplification QPCR. CBA/Ca and C57/BL6 mice were irradiated with 2 Gy, total bone marrow isolated at 24 hr post exposure and transcriptional changes in validated DNA damage response marker genes analysed.

Expression of the gene *Sesn2* is upregulated in both human and mouse following exposure to radiation. *Sesn2* is regulated by p53 in response to DNA damage (Kabacik et al. 2011) and in human *ex vivo* irradiated blood samples, shows an up-regulation after X-ray exposure (Brzoska and Kruszewski 2015). *Sesn2* was also upregulated in mouse blood samples following X-ray exposure (Kabacik et al. 2011).

After exposure to a 2 Gy dose of ionising radiation, *Sesn2* expression was significantly up-regulated 1.7 and 1.8-fold, respectively, in both CBA/Ca and C57BL/6 total bone marrow at 24 hr (Figure 51). Following single cell analysis, *Sesn2* was up-regulated both CBA/Ca and C57BL/6 strains (Figure 52). In CBA/Ca HSPCs, *Sesn2* was significantly up-regulated in HSC, MPP2, MPP3 and MPP4 populations. In C57/BL6 HSPCs, *Sesn2* was significantly up-regulated in two populations, MPP1 and MPP2 and, although not significant, was up-regulated in all other populations.
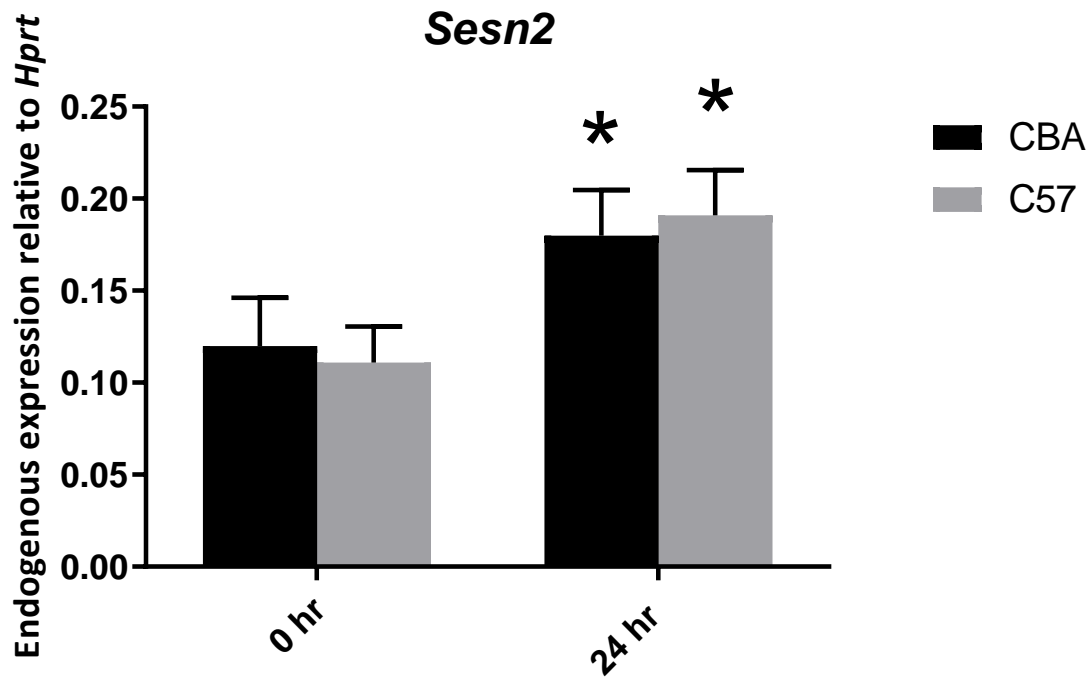
Figure 51. *Sesn2* expression in CBA and C57 bone marrow at 24 hr following a 2 Gy dose.

Error bars represent the standard deviation of 3 mice per dose. A t-test was performed to test for significance between time-points and indicated by an astrix where p=0.05.
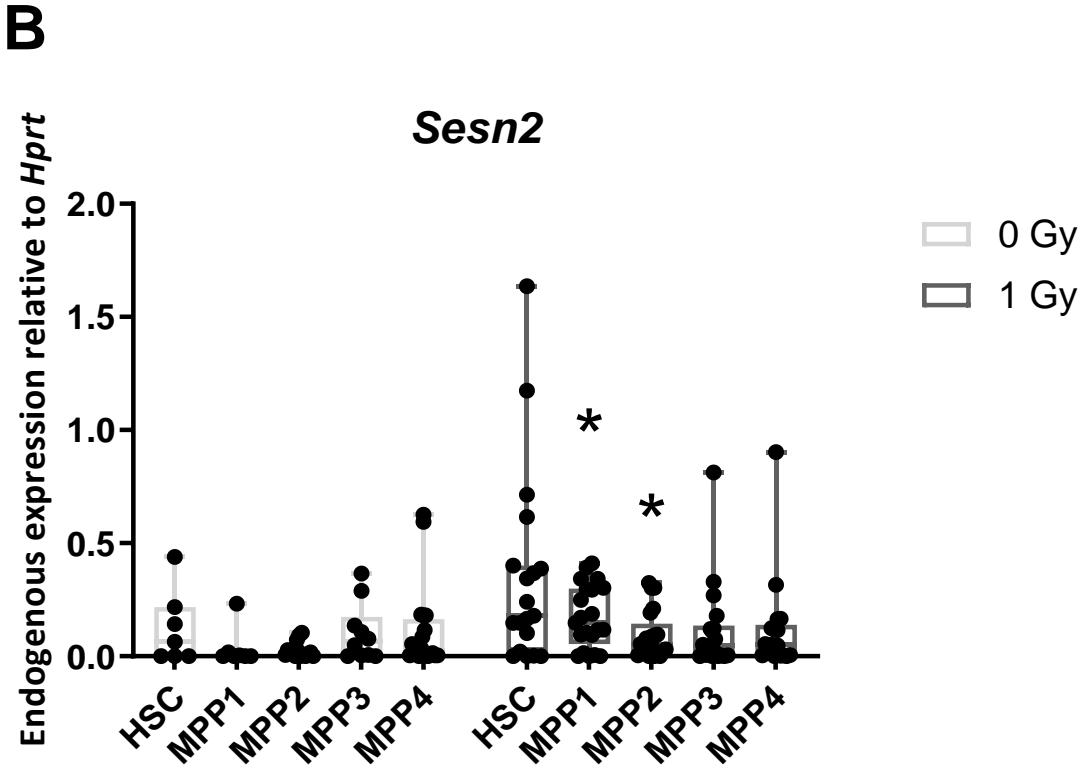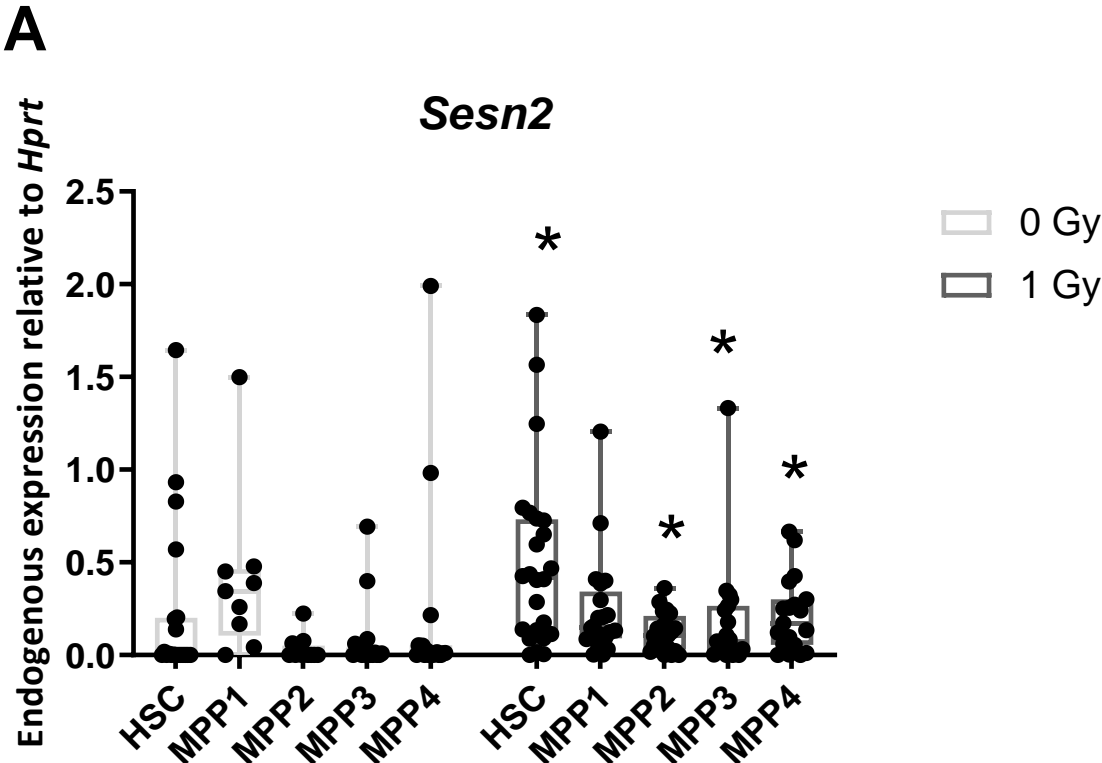
Figure 52. *Sesn2* expression in CBA (A) and C57 (B) HSPC populations.

MQRT-PCR analysis in HSCs, MPP1, MPP2, MPP3 and MPP4 populations at 24 hr following a 0 Gy and a 1 Gy dose. Each population contains 29 single cells which were individually analysed by QMRT-PCR. A Mann Whitney test was performed to test for significance between control and irradiated samples and indicated by an astrix where p=0.05.

To identify long term biomarkers of irradiation exposure in the bone marrow, cells from HSCs and the progenitor populations MPP1, MPP2, MPP3 and MPP4 were analysed using the nCounter system which can assess transcriptional levels of up to 800 genes. CBA/Ca and C57BL/6 mice were irradiated with 1 Gy and bone marrow isolated at 1 month, 3 month and 6 months following exposure (Figure 53). Approximately 300 cells from each population were pooled and analysed using the Mouse PanCancer Pathways Panel on the nCounter system. The PanCancer Pathways Panel was used as it measures the expression of 770 genes from cancer associated pathways such as MAPK, STAT, PI3K, RAS cell cycle, apoptosis, Hedgehog, Wnt, DNA damage control, transcriptional regulation, chromatin modification and transforming growth factor beta (TGF-β).
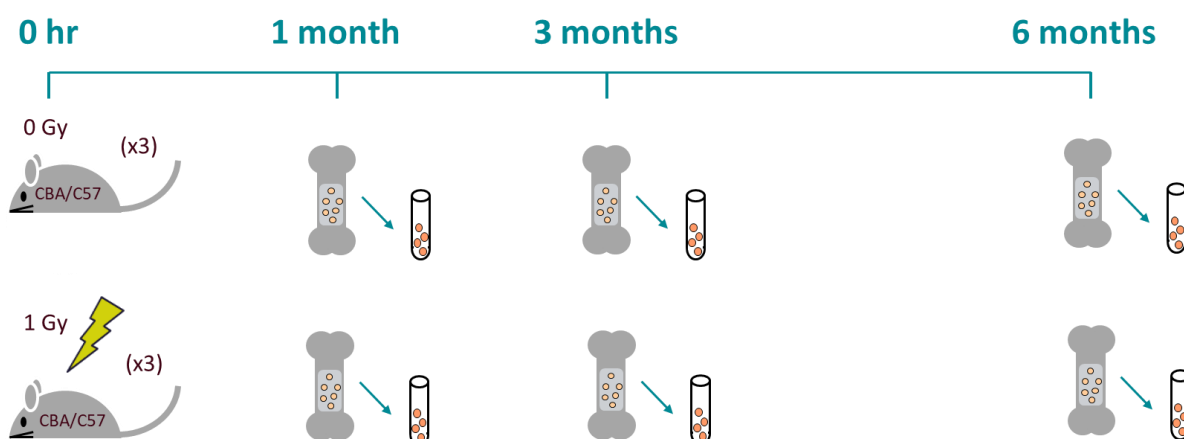
Figure 53. Single cell experimental plan.

CBA/Ca and C57/BL6 mice, 3 mice per strain, were irradiated with 0 Gy and 1 Gy and bone marrow isolated at 1 month, 3 month and 6 month time-points.

The molecular barcode counts from the nCounter system could also add further information on transcriptional expression of specific genes within each HSPC sub-population and validate the sorting protocol. Previous studies identified many genes, such as suppressor of cytokine signaling 2 (*Socs2)* (Zhong et al. 2005), cyclin dependent kinase inhibitor 1c (*Cdkn1c)*, integrin subunit alpha 6 (*Itga6),* mixed-lineage leukaemia 4 (*Mllt4)*, PBX homeobox 1 (*Pbx1)*, Cyclin D3 (*Ccnd3)*, Cyclin A2 (*Ccna2)*, Cyclin B1 (*Ccnb1)* and interleukin 1 receptor type 1 (*Il1r1)* (Forsberg et al. 2005), which had differential expression between murine HSCs and MPPs, and these genes were also included on the PanCancer Pathways panel, allowing a comparison between previous studies and between CBA/Ca and C57BL/6 strains to be made (Figure 54). When analysing the molecular counts of these genes a strong expression of the genes *Itga6*, *Socs2*, *Mllt4*, *Pbx3*, *Cdkn1c* and *Ccnd3* in HSCs was detected with a lower expression in MPP populations (Figure 54 A and B). Also, a strong expression of the genes *Ccna2*, *Ccnb1* and *Il1r1* was reported in MPP populations with a lower expression in HSC (Figure 54 C and D). This analysis is consistent with the previously published studies.
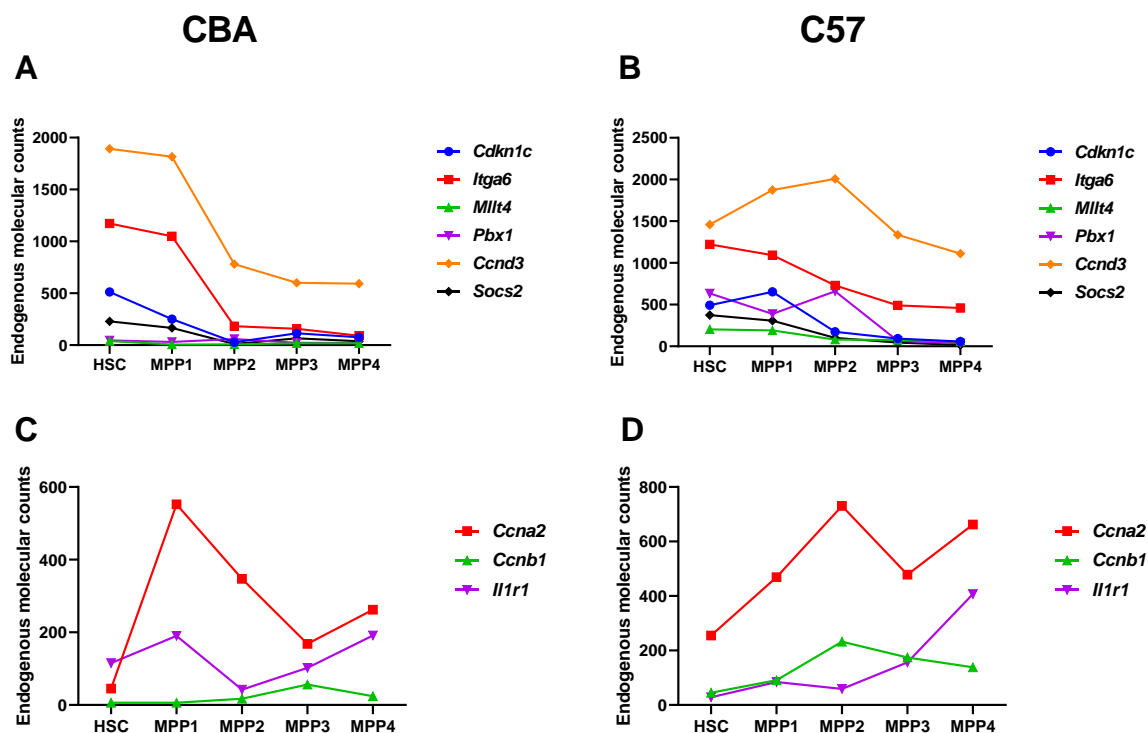
Figure 54. Transcriptional analysis of CBA (A, C) and C57 (B, D) HSC, MPP1, MPP2, MPP3 and MPP4 populations.

The data represent molecular barcode counts, measured by the nCounter system, for the genes *Cdkn1c, Itga6, Mllt4, Pbx1, Ccnd3* and *Soc2* which have a high level of expression in HSC and a lower level of expression in MPPs (A, B) and genes *Ccna2, Ccnb1* and *Il1r1* which have a low level of expression in HSC and a higher level of expression in MPPs (C, D). Data represent the endogenous molecular counts of 300 cells for each population normalised to *Hprt.*

Genes of interest were identified using the BRB array tools algorithm which identified genes whose transcription was significantly altered after irradiation (Figure 55). The gene *Stat1* was the second most significantly expressed gene at 1 month between all CBA/Ca 0 Gy and 1 Gy cell populations with a 3-fold up-regulation at $p=0.001$. Another gene *Jak3*, which is another member of the Jak/Stat pathway was found to be significantly expressed in MPP cells. These genes were analysed further at each time-point.

Other interesting genes were found to have a significantly differentially expressed level following radiation exposure, such as Fanconi anaemia, complementary group A (*Fanca)* and *Dnmt3a*. Mutations in the gene *FANCA* are associated with Fanconi anaemia and, also, acute myeloid leukaemia (Alter 2014). Transcriptional analysis of *FANCA* expression in human AML patients suggests that mutations and reduced expression of *FANCA* may be associated with AML (Tischkowitz et al. 2004). Here, a strong upregulation was identified after radiation exposure which may be due to its role in replication repair or cell cycle checkpoint (Benitez et al. 2018). The nCounter analysis also revealed a decreased in *Dnmt3a* expression which has previously been reported in DNMT3A protein levels in C57BL/6 thymus 10 days after fractionated and acute radiation exposure (Pogribny et al. 2005). Also, 4 weeks after exposure to ^{56}Fe ions, a decrease in the transcriptional expression in the *Dnmt* genes *Dnmt1*, *Dnmt3a* and *Dnmt3a* was reported in C57/BL6 HSPCs (Miousse et al. 2014). As a methyltransferase involved in DNA methylation, decrease in expression of *Dnmt3a* could play an important role in the development of and further investigation would be of interest. The nCounter molecular counts of *Fanca* and *Dnmt3a*, however, were very low and so analysis by single cell MQRT-PCR might not be possible.

Perhaps of concern, the gene *Kit*, also one of our cell surface markers in the single cell isolation protocol, showed a significant decrease in transcriptional expression at 1 month after a 1 Gy dose. Previous studies have reported an initial decrease in *Kit* expression in murine HSCs after irradiation which eventually returned to control levels after 10 weeks (Simonnet et al. 2009). However, a recent study by Chen et al. has shown that a down-regulation of *Kit* did not affect the ability of the HSPCs to be transplanted (Chen, Faltusova, et al. 2016). Therefore, although a down-regulation is evident at 1 month, there is no evidence to suggest that it affects the function of the HSPCs.
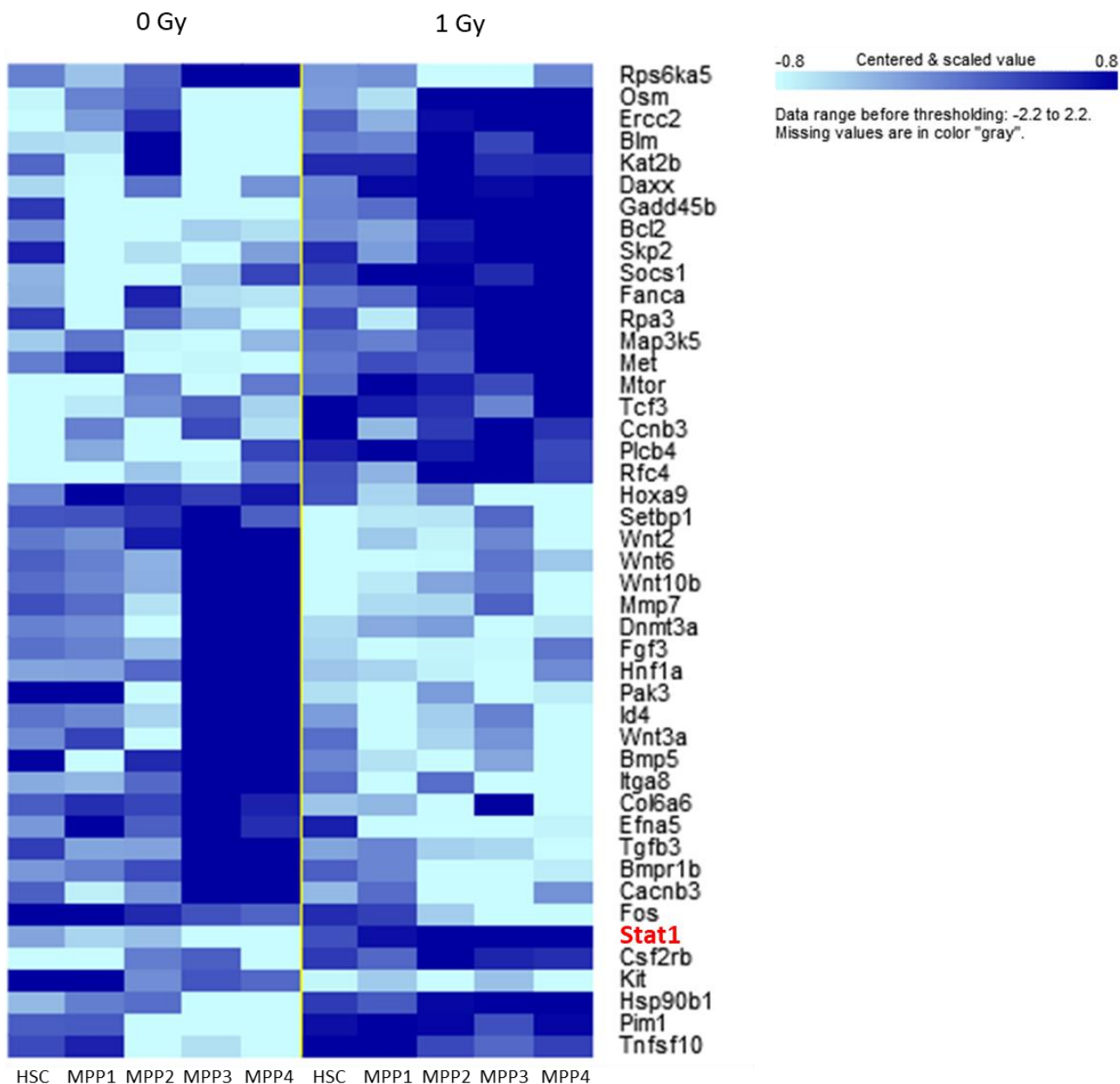
Figure 55. BRBArray Tools analysis of nCounter PanCancer Pathways Panel data.

A class comparison analysis was performed on samples using BRB-ArrayTools developed by Dr. Richard Simon and the BRB-ArrayTools Development Team. Analysis of CBA/Ca HSC, MPP1, MPP2, MPP3 and MPP4 populations, each containing data from approximately 300 cells, at 1 month identify 45 genes which have a significant molecular count of p≤0.05 between 0 Gy and 1 Gy. The level of expression is indicated by a colour chart with a light blue indicating low expression and a dark blue indicating high expression.

In CBA/Ca mice, *Stat1* was also significantly expressed when analysing MPP populations alone (p=0.002) (Figure 56 A). This up-regulation was still found at 3 months, although not significant. Interestingly, when analysing C57/BL6 1 month cell populations, *Stat1* was not found to be differentially expressed between 0 Gy and 1 Gy samples at any time-point (Figure 56 B). The gene *Jak3* was also significantly up-regulated 2.9-fold when analysing CBA/Ca MPP populations at 1 month following 0 Gy and 1 Gy (p=0.02) (Figure 57 A) and again this was not found in C57/BL6 samples (Figure 57 B). Since these genes are both present in the Jak/Stat pathway, they were identified as genes of interest and primers were designed to try to validate this expression on a single cell level by QPCR.

Q-PCR primers were designed for *Stat1* and *Jak3* and single cells were first pre-amplified and then analysed by Q-PCR. Single cells from CBA/Ca and C57/BL6 strains were isolated after 0 Gy and 1 Gy at 1 month, 3 months and 6 months for 5 HSPC populations (HSCs, MPP1, MPP2, MPP3 and MPP4). QPCR analysis validated the nCounter results with a significant up-regulation in *Stat1* expression in CBA/Ca samples after 1 Gy at 1 month in HSCs, MPP2 and MPP4 populations (Figure 58 A) while no significant increase in expression was found in C57/BL6 samples at 1 month (Figure 58 B). At 3 months this up-regulation was still found to be significant in HSCs, MPP1 and MPP2 populations in CBA/Ca samples (Figure 58 C) while in C57/BL6 samples a significant down-regulation was found in MPP3 (Figure 58 D). At 6 months the response is still found to be up-regulated in MPP1 and MPP3 and significantly up-regulated in the MPP4 population for CBA/Ca (Figure 58 E), while for C57/BL6 up-regulated in MPP2 and is significantly up-regulated in MPP3 cells (Figure 58 F). *Jak3* meanwhile does not have as strong a response as *Stat1.* CBA/Ca cells show a significant up-regulation in expression at 1 month in HSCs and a down regulation in expression in MPP4 cells (Figure 59 A). At later timepoints of 3 and 6 months, no CBA/Ca cell population showed a significant change in expression (Figure 59 C, E). In C57BL/6 cells there was no significant change in expression at 1 month (Figure 59 B) and, at 3 months, MPP4 had a significant down-regulation (Figure 59 D) while at 6 months MPP2 had a significant up-regulation in expression (Figure 59 F).

Overall, this QPCR data validates the response seen using the nCounter system, that CBA/Ca HSPCs have a strong up-regulation in *Stat1* expression that persists 6 months after X-ray exposure. Importantly, this response is not seen in C57BL/6 mice

and is clearly a strain and HSC specific response which could be associated with an overexpressed pathway in the CBA/Ca mouse model.
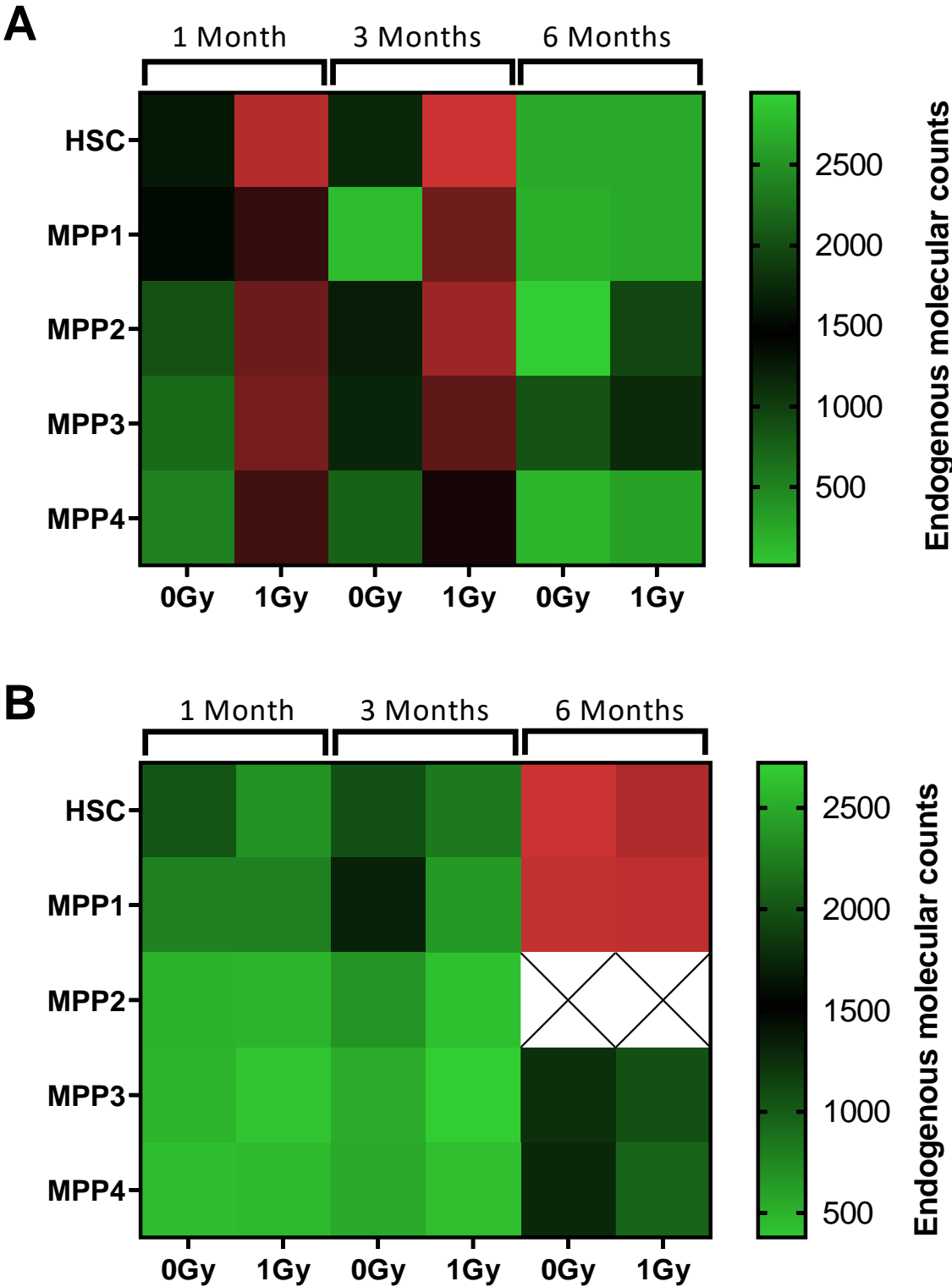
Figure 56. Long term nCounter *Stat1* expression in CBA and C57 HSPCs.

Stat1 expression in CBA/Ca (A) and C57/BL6 (B) HSCs, MPP1, MPP2, MPP3 and MPP4 populations at 1 month, 3 months and 6 months after 0 Gy and 1 Gy X-rays.

Data represent the endogenous molecular counts of 300 cells for each population normalised to *Hprt.* Data represent results of a single experiment for each time point and strain. Boxes with an X represent timepoints where no sample was available.
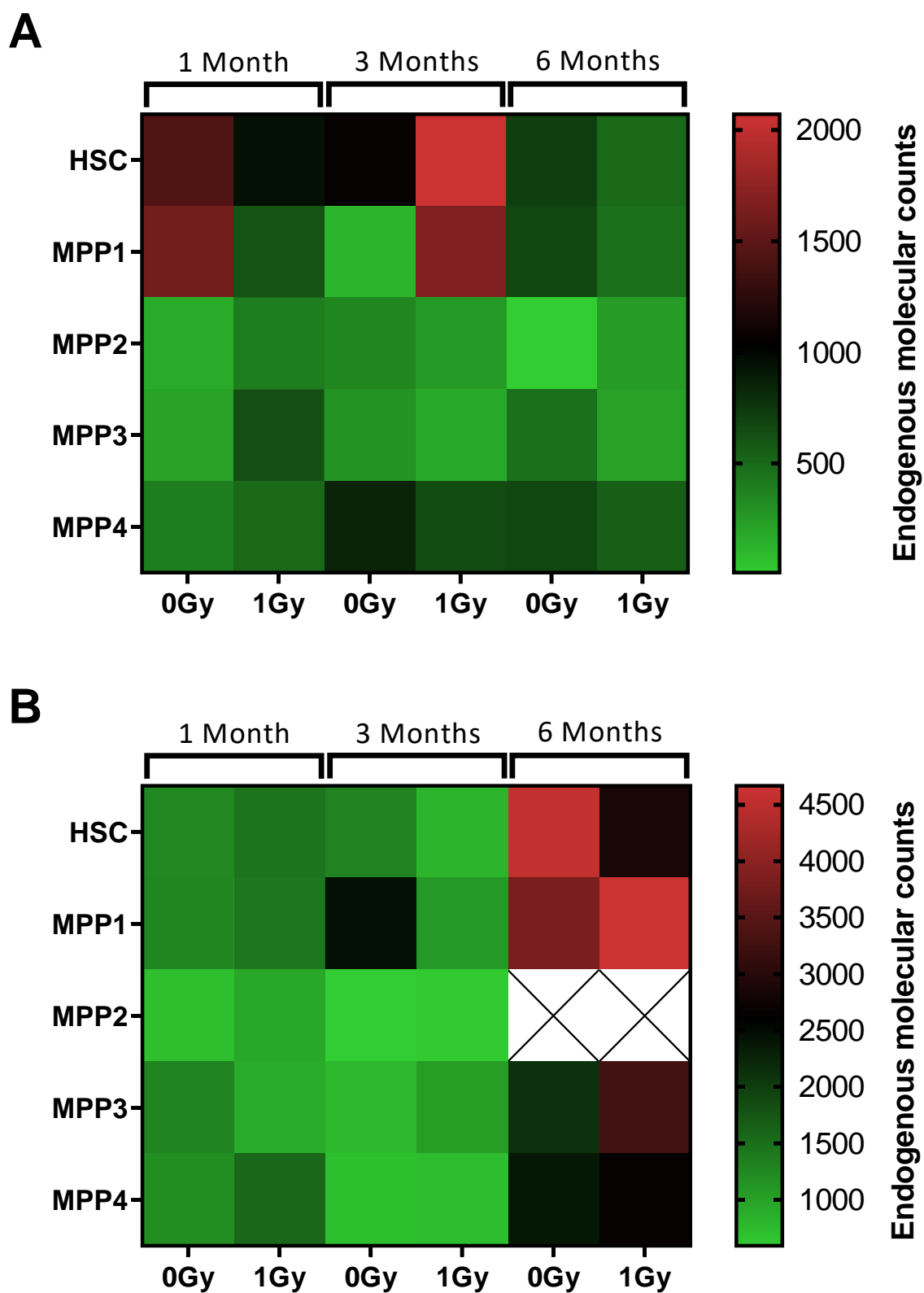
Figure 57. Long term nCounter *Jak3* expression in CBA and C57 HSPCs.

Jak3 expression in CBA/Ca (A) and C57/BL6 (B) HSCs, MPP1, MPP2, MPP3 and MPP4 populations at 1 month, 3 months and 6 months after 0 Gy and 1 Gy X-rays. Data represent the endogenous molecular counts of 300 cells for each population normalised to *Hprt.* Data represent results of a single experiment for each time point and strain. Boxes with an X represent timepoints where no sample was available.
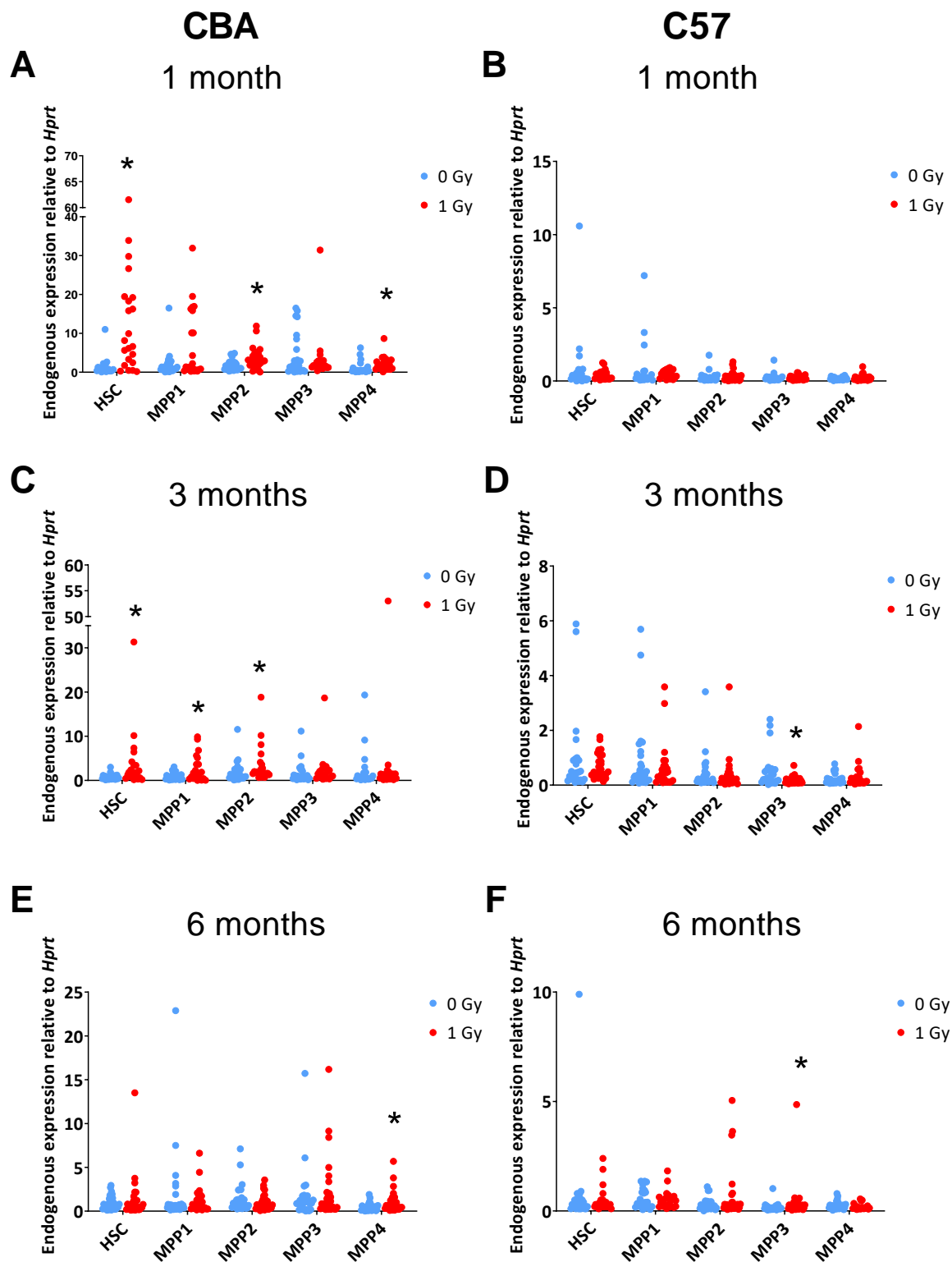
Figure 58. Long term MQRT-PCR *Stat1* expression in CBA and C57 HSPCs.

Stat1 expression in CBA/Ca (A, C E) and C57BL/6 (B, D, F) HSCs, MPP1, MPP2, MPP3 and MPP4 populations at 1 month (A, B), 3 months (C, D) and 6 months (E, F) after 0 Gy and 1 Gy X-rays. Data represent the endogenous expression of 29 single

cells for each population normalised to *Hprt.* Error bars represent the minimum and maximum values. Data represent results of a single experiment for each time point and strain. Significance was calculated by a Mann Whitney test and indicated by an astrix where p≤0.05.
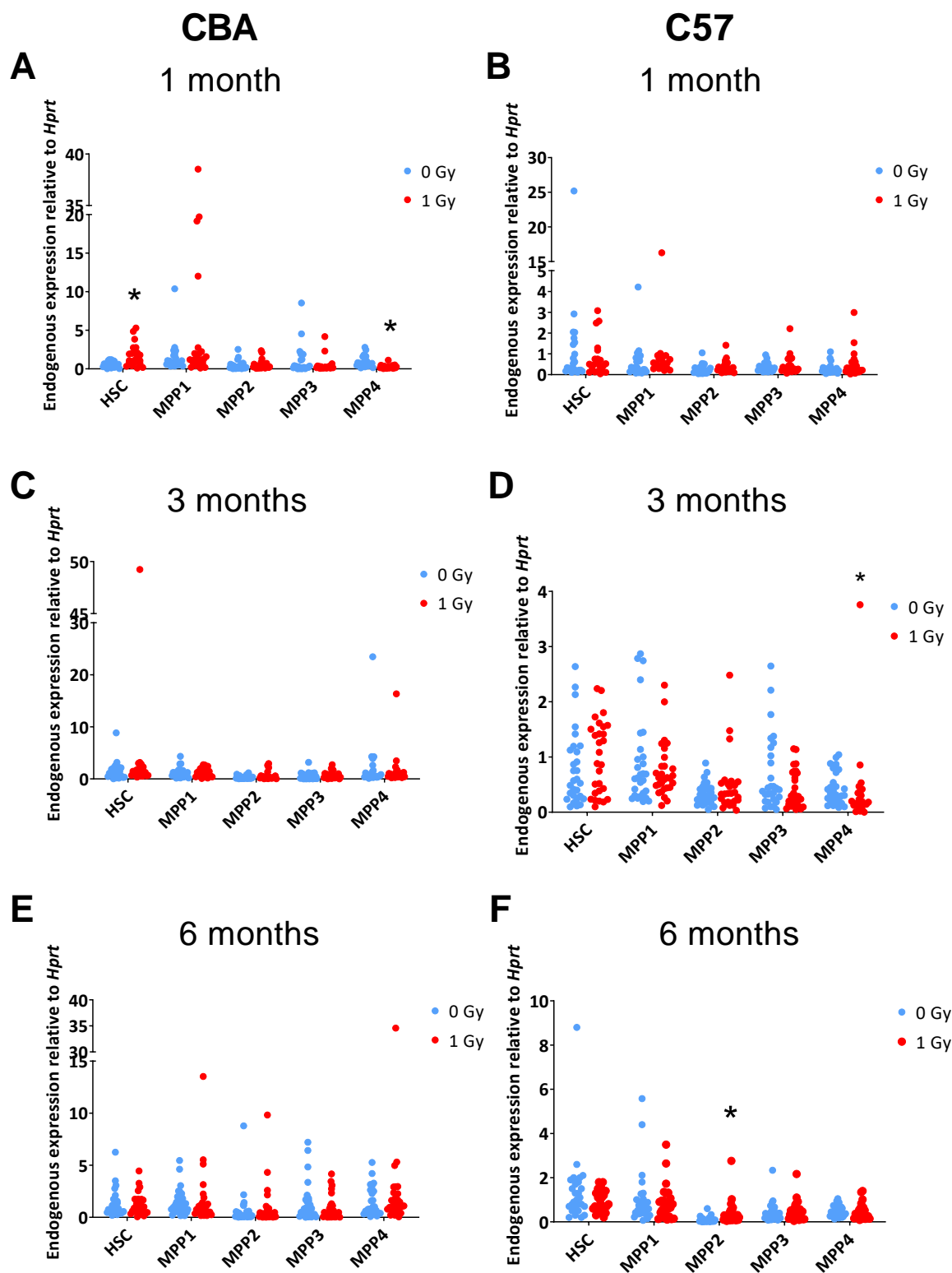
Figure 59. Long term MQRT-PCR *Jak3* expression in CBA and C57 HSPCs.

Jak3 expression in CBA/Ca (A, C E) and C57BL/6 (B, D, F) HSCs, MPP1, MPP2, MPP3 and MPP4 populations at 1 month (A, B), 3 months (C, D) and 6 months (E, F)

after 0 Gy and 1 Gy X-rays. Data represent the endogenous expression of 29 single cells for each population normalised to *Hprt.* Error bars represent the minimum and maximum values. Data represent results of a single experiment for each time point and strain. Significance was calculated by a Mann Whitney test and indicated by an astrix where p≤0.05.

Discussion and future directions

The overall challenging goal was to study the in vivo transcriptional response of individual cells from bone marrow compartment progenitors and stem cells to irradiation. This work developed a single cell QPCR protocol and in long-term experiments identifies the JAK/STAT pathway, with Stat1 in particular, as being upregulated in response to irradiation in the CBA mouse strain.

In order to develop a working protocol for analysis of transcriptional changes in HSPC populations, standard RNA extraction kits were first tested. A protocol with standard RNA extraction and reverse transcription would allow genes of interest to be analysed without the introduction of another step, such as pre-amplification, which may introduce further variability into the experiment. The cell surface markers CD201+CD27+ were used allowing the isolation of a larger HSPC population in the CBA mouse and overcoming the problem of using SLAM markers in CBA mice due to the lack of Sca-1. RNA extraction using standard RNA extraction kits and kits specific for low amounts of cells, however, did not provide sufficient RNA for reverse transcription. The Single Cell RNA Purification Kit extracted the greatest amount of RNA from the smallest number of cells, however this amount was just 45 ng from one population. Although the CD27 and CD201 markers can isolate more HSPCs than the traditional SLAM markers, pooling of many mouse samples would be required due to the low numbers of cells isolated.

The use of pre-amplification kits was next investigated and the use of the markers c-Kit, Sca-1, CD48, CD150, CD34 and CD135 allowing the isolation of 5 HSPC sub populations; HSCs, MPP1, MPP2, MPP3 and MPP4. The development of a single cell protocol, or protocol involving few cells, allows the isolation of rarer HSPC populations. SLAM markers were therefore chosen for single cell sorting due to their validated use

for isolation of HSCs and MMP. A pre-amplification protocol by Moignard et al. using the Cells Direct Kit and primers specific for the genes of interest was determined to be the best protocol for the isolation of single cells (Figure 60). The REPLI-g WGA & WTA kit and REPLI-g WTA kit worked best with cell numbers of 100 and 10, respectively, while the use of smaller cell numbers resulted in Ct values greater than 35 Ct.
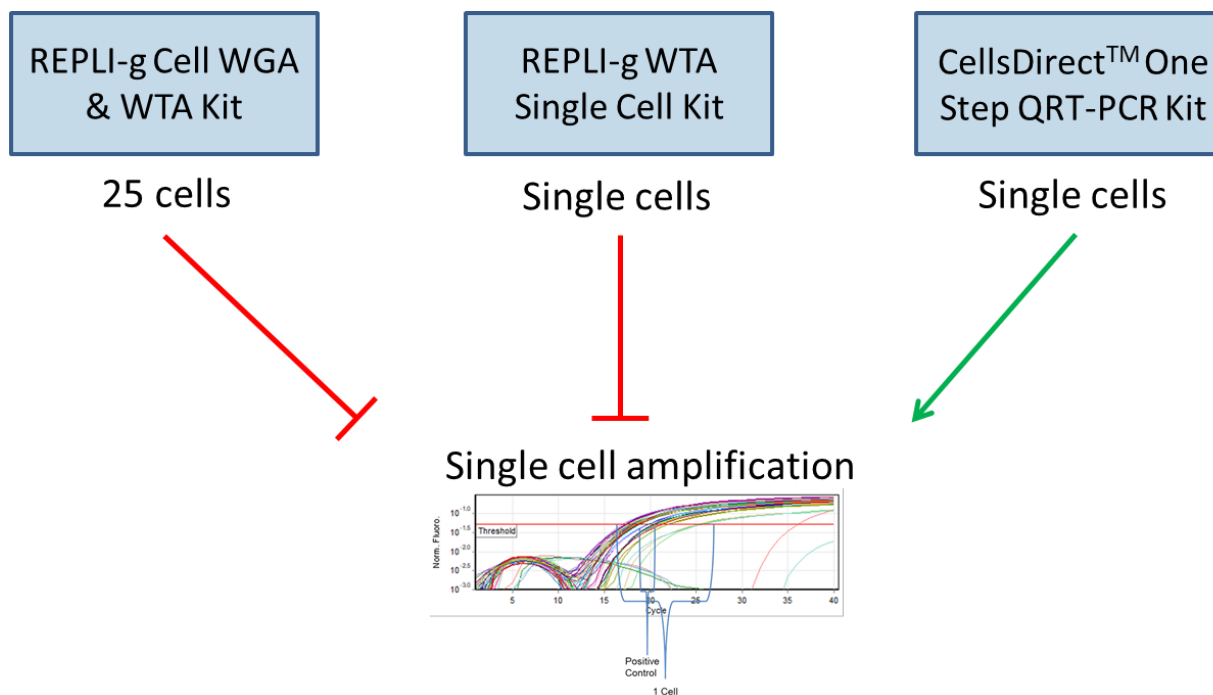
Figure 60. Validation of RNA amplification kits.

The kits were tested with 25 cells (REPLI-g Cell WGA & WTA Kit) and single cells (REPLI-g WTA Single Cell Kit and CellsDirect™ One Step QRT-PCR Kit) and only the CellsDirect™ One Step QRT-PCR Kit resulted in single cell amplification using a modified protocol by Moignard et al.

To confirm that the HSC populations isolated are indeed long-term, expression of the gene *Hoxb5*, a gene previously reported to be expressed in long term HSCs (Chen et al. 2016), was investigated and found to have a higher expression in the HSC population. This expression was significant in comparison to the MPP2 population. Significance could not be determined for MPP3 and MPP4 populations as there was only one cell in each population that expressed *Hoxb5*. However, with only one cell

expressing a low level of *Hoxb5* in each MPP3 and MPP4, the difference is clear. To ensure that the transcriptional expression in single cell analysis is representative of bulk RNA transcriptional expression, the expression of a known DNA damage response gene, *Sesn2*, was analysed 24 hr after radiation exposure in total bone marrow using standard QPCR in comparison to single cell analysis of HSPCs. A significant up-regulation was detected in both CBA/Ca and C57/BL6 mice. At single cell level a significant up-regulation was also detected in MPP populations of CBA/Ca and C57/BL6 mice demonstrating that the single cell protocol does represent the transcriptional changes seen in bulk RNA samples and is therefore validated.

Another step was also taken to validate the protocol by comparing molecular counts obtained using the nCounter system to the transcriptional profile of previously published studies. The genes *Itga6*, *Socs2*, *Mllt4*, *Pbx3*, *Cdkn1c* and *Ccnd3* are highly expressed in HSCs, as previously reported (Zhong et al. 2005; Forsberg et al. 2005), with expression gradually reducing in each MPP population each with interesting possible roles. *Itga6* is a potential stem cell marker and HSC cell surface expression of Itga6 has been previously reported (Wagers, Allsopp, and Weissman 2002). *Socs2* is a cytokine suppressor with its reduction in MPP populations possibly allowing cytokine signalling for the differentiation of HSCs and *Mllt4* is an actin filament-binding protein. The gene *Pbx3* interacts with *Hox* genes and so its expression could be co-operating with *Hoxb5* expression as detected by the MQRT-PCR expression. With regards cell proliferation, previous work has shown that HSCs cell proliferation is possibly regulated by cell cycle inhibitors *Cdkn1c* and *Ccnd3* (Forsberg et al. 2005), as also reported in this study. The genes *Ccna2*, *Ccnb1* and *Il1r1* have previously been reported to have a low level of expression in HSC and a higher level of expression in MPPs (Forsberg et al. 2005), which again can be seen in this study. *Ccna2* and *Ccnb1* are involved in cell cycle proliferation and have a high level of expression in MPP populations, probably due to the increased proliferation of the cells as they differentiate. *Il1r1* is a cytokine receptor and its higher level of expression in MPPs may be due to the differentiation signals needed by the cells. When comparing the expression between strains, although slight variations in expression levels between strains was evident the response was very similar considering that only one sample was used per population. Overall, the expression analysis by the nCounter

PanCancer Pathways panel further validates the isolation of HSPC populations in this study and confirms previously reported transcriptional analysis of these populations.

To investigate gene expression changes in this study, nCounter data were first analysed to examine expression levels. Since the nCounter protocol for low input RNA already has an amplification step, the final counts give an indication of the level of expression expected after QPCR analysis and identify genes of interest with a low level of expression. nCounter data of *Stat1* and *Jak3* reported counts of approximately 500 or higher which, in comparison to other genes, seemed to give a high level of expression. *Hprt* expression allowed the identification of the presence of a single cell after sorting. However, difficulties in the detection of mRNA using single cell techniques has been well documented (Stahlberg and Kubista 2018). This preliminary step of analysing nCounter counts increased the possibility of detectable QPCR expression in the samples and avoiding dropout issues. Other techniques could also be used which could overcome issues with detection. Digitial droplet PCR (DDPCR) is a fluidic system, based on water-oil emulsion droplet technology, available through either RainDance's RainDrop Digital PCR System or Bio-Rad's Droplet DigitalTM PCR System. A sample is fractionated into approximately 20,000 droplets and then the template in each individual droplet is amplified by PCR. Each individual droplet is then passed through a detection system. The system reduces error rates by removing the amplification efficiency reliance of qPCR. Another emerging technique is the MinION from Nanopore Technologies, which utilises an electric current to detect the passing of transcripts through 512 channels, can provide direct RNA sequencing, eliminating PCR bias. New technologies could be used in future analysis to quantify with increasing accuracy the *Stat1* and *Jak3* changes in expression.

Validation of the single cell protocol allowed for long-term single cell experiments to be performed at 1 month, 3 months and 6 months after irradiation in vivo. The nCounter system allowed for the identification of genes of interest by analysing approximately 300 cells from each HSPC sub-population with the Low RNA Input Kit and the PanCancer Pathways Panel. Two genes of interest were identified using the nCounter system, *Stat1* and *Jak3,* showing significant changes in expression after irradiation in some of the CBA/Ca HSPC populations while not showing the same response in C57BL/6 HSPC populations. *Stat1* has more HSPC populations with significant changes in transcription after irradiation in comparison to *Jak3*, but both are

of interest as they are both involved in the JAK/STAT pathway. The increased transcriptional expression of *Stat1* in CBA/Ca mice and not in C57BL/6 mice was of particular interest as CBA/Ca is prone to developing AML after radiation exposure while the C57BL/6 strain does not. Remarkably, this *Stat1* expression, although more pronounced at 1 and 3 months, is still expressed at 6 months after IR in CBA/Ca mice and so may be important for AML development, as has been previously reported (Kovacic et al. 2006).

The Jak/Stat pathway is a signalling pathway involved in processes such as immune response, cell proliferation and apoptosis. The pathway is usually activated by cytokines, such as interferons (IFNs) and interleukins (ILs), that bind to cell surface receptors and cause JAKs to phosphorylate tyrosine residues on the receptor which creates docking sites for STAT proteins (Figure 61). The STAT proteins are then tyrosine-phosphorylated by JAKs which results in dissociation of STAT from the receptor and formation of a STAT dimer which enters the nucleus to transcribe target genes. As a transcription factor, STAT1 binds to and co-operates with a number of co-activators such as CREB binding protein (CBP), p300, minichromosome maintenance complex component 5 (Mcm5), n-Myc and STAT interactor (Nmi), breast cancer type 1 susceptibility protein 1 (BRCA1), specificity protein 1 (SP1), Jun and nuclear factor kappa B subunit 1 (NFκB) (Ramana et al. 2000). It targets B-cell lymphoma 2 (Bcl-2) family members Bcl-2 and Bcl-Xl (Cao et al. 2015; Stephanou et al. 2000) and caspases 1, 2, 3 and 8 (Kumar et al. 1997; Fulda and Debatin 2002) to regulate apoptosis, cell cycle regulators p21, p27, c-myc and cyclins (Chin et al. 1996; Dimberg et al. 2003) to regulate cell cycle arrest and also the tumour suppressor transcription factor p53 (Townsend et al. 2004). Some other members of the JAK/STAT pathway were also included on the PanCancer Pathways panel such as *Jak1*, *Jak2*, *Stat3* and *Stat4*. Some show differential levels of expression both after irradiation and between strains. They have not yet been analysed due to difficulties in primer design and may not be feasible because of low molecular barcode counts indicating very low levels of expression. However, these will be further assessed in future studies.
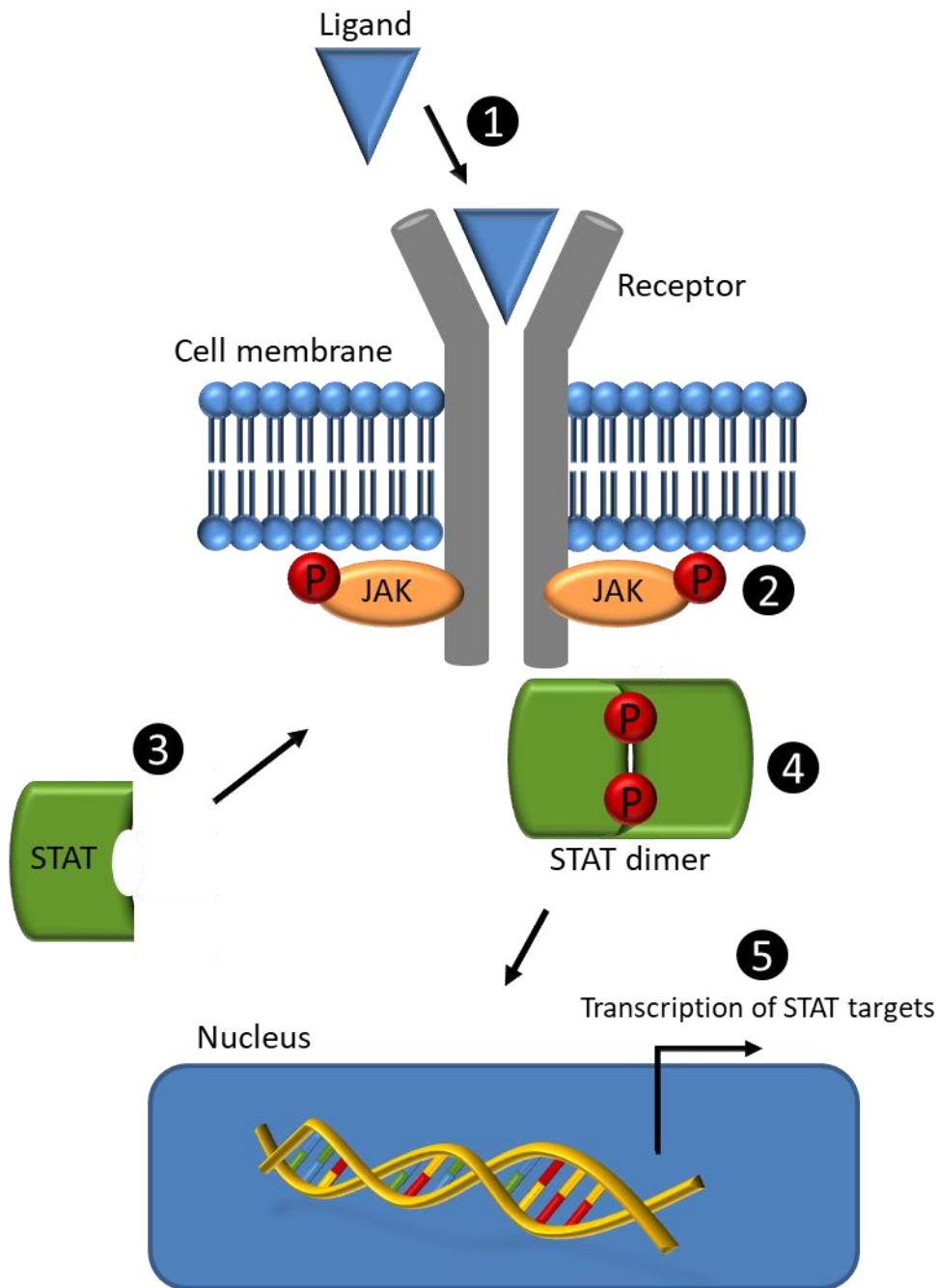
Figure 61. The JAK/STAT pathway.

Activation of the JAK/STAT pathway begins with (1) binding of a ligand to a cell surface receptor such as cytokine receptors, G protein coupled receptors, growth factor receptors or tyrosine kinase receptors. This results in (2) JAKs to phosphorylate tyrosine residues on the receptor. (3) STATs are then recruited to the receptor and phosphorylated and can form dimer complexes (4). The STAT dimer can enter the nucleus and activate STAT target genes.

The Jak/Stat pathway is of interest as protein levels of STAT1 has previously been found to be constitutively activated in human AML cell lines (Spiekermann et al. 2001), blood samples from AML patients (Gouilleux-Gruart et al. 1996) and AML leukemic cells (Aronica et al. 1996; Weber-Nordt et al. 1996). In another study, STAT1 protein was found to be constitutively activated in primary AML blasts in 12 out of 16 AML patients (Spiekermann et al. 2002). Other JAK/STAT family members have also been found to be differentially expressed in AML cells such as STAT3 and STAT5 (Spiekermann et al. 2001; Gouilleux-Gruart et al. 1996; Benekli et al. 2002). Activation of STAT1 is not limited to AML and has also been detected in other cancers such as breast cancer (Perou et al. 1999) and lung cancer (Chen et al. 2007). The JAK/STAT pathway is clearly activated in AML, however, specific sub groups show higher activation that others. Mouse models of AML with AML-associated fusion proteins PML/RARa and DEK/NUP214 show a clear enhanced level of STAT5 expression in comparison to non-leukemic cells (Oancea et al. 2014). Enhanced phosphorylation STAT5 was confirmed in primary t(6;9) AML patient samples in comparison to healthy donors, with no difference, however, of STAT1 phosphorylation levels between groups (Oancea et al. 2014). The relevance of STAT activation in the development of normal cells to leukemic cells has also been demonstrated by the increased colony growth in NIH 3T3 cells transfected with the constitutively activated tyrosine kinase v-src and *Stat3* plasmids, and the decrease in colony formation with dominant negative *Stat3* mutants (Bromberg et al. 1998).

In HSCs, the JAK/STAT pathway plays a vital role in HSC functions such as differentiation and self-renewal. STATs are essential for the differentiation of myeloid derived cells from in vitro studies, with STAT5 mutant HSCs producing proliferative multilineage cells (Stine and Matunis 2013; Kato et al. 2005), however, recent in vivo work using STAT knockout mice has questioned the role STAT1, STAT3 and STAT6 play in myeloid differentiation (Coffer, Koenderman, and de Groot 2000). STATs also play a role in leukemic HSCs self-renewal. Transplantation assays have revealed that STAT5 deficient stem cells failed to sufficiently repopulate the blood, spleen, thymus or bone marrow of irradiated recipient mice (Snow et al. 2002). Also, STAT 1, 3 and 5 expression was enhanced in the AML oncogene model MN1/HOXA9, while Stat5b-null MN1/HOXA9 cells had a reduced proliferative rate compared to wild-type cells

(Heuser et al. 2009). Similarly, in human AML patients, enhanced expression of STAT genes was also found in patients with a MN1 and HOXA9 co-overexpression (Heuser et al. 2009). STAT1 knockout mouse models have illustrated the importance of Stat1 in leukemic formation. STAT1-/- cell lines showed a decrease in leukemic development in recipient RAG2-/- mice with a decrease in spleen and liver leukemic infiltration (Kovacic et al. 2006).

Expression of STAT1 is mainly considered to have a tumour suppressor function, however, in recent years, studies reported conflicting results on its function. A constitutive activation of the JAK/STAT pathway has been demonstrated in leukemic cells. The function of Stat1 in AML cells appears to be different to that in normal functioning cells. The JAK/STAT1 pathway has been targeted in the treatment of AML with the drug Atiprimod decreasing phosphorylation of Stat3 and Stat5, inhibiting cell proliferation and inducing apoptosis of AML cell lines (Faderl et al. 2007). STAT1 has also been shown to promote the proliferation of AML cell lines with suppression of STAT1 phosphorylation by the CDK8 inhibitor cortistatin A (CA) resulting in growth arrest (Nitulescu et al. 2017). STAT1 overexpression has also been identified in a radiation-resistant human tumour xenograft in comparison to a radiosensitive SCC-61 tumour (Khodarev et al. 2004). Moreover, transfection of STAT1 vectors into the radiosensitive SCC-61 tumour cells increased cell survival after 3 Gy in comparison to cell survival of SCC-61 cells alone and so conferring a radioresistant phenotype (Khodarev et al. 2004). The function of Stat1 therefore seems to be multifaceted, depending on cell type and requires further investigation.

STAT1 transcriptional expression can be upregulated by interferon-alpha, interferon-gamma (Lehtonen, Matikainen, and Julkunen 1997), epidermal growth factor receptor (*EGFR*) and human epidermal growth factor (*HER2*) (Han et al. 2013) and potentially controlled by *STAT1* distal regulatory elements (Yuasa and Hijikata 2016). *STAT1* and *STAT1*-dependent genes have been shown to be activated by IR in mouse head and neck cancer and breast cancer tumour xenografts (Khodarev et al. 2007) and has been identified as a radiation responsive gene in the peripheral blood of radiotherapy patients (Amundson et al. 2004). Activation of *Stat1* by interferons remains possibly the most widely studied pathway of activation. *Stat1* was initially thought to be a tumour-suppressor as *Stat1* deficient mice were found to be tumour-prone (Kaplan et al. 1998). This may have been more to do with the IFN activation of *Stat1* rather than

Stat1 itself as tumours were found to grow faster in IFN deficient mice than controls, while *Stat1*$^{-/-}$ mice showed a decrease in leukemic formation in comparison to *Stat1*$^{+/-}$ mice (Kovacic et al. 2006). For decades interferon-α therapy was used for the treatment of AML due to its proapoptotic and antiproliferative effects however, the clinical outcomes were variable and success was achieved only in a minority of patients (Smits, Anguille, and Berneman 2013). This clinical failure of interferon therapy highlighted the need to investigate other activated pathways in AML and other possible treatment options.

Another factor to activate *STAT1* is EGFR. A recent study demonstrated how *STAT1* transcriptional expression is upregulated by nuclear EGFR in cooperation with STAT3 (Han et al. 2013). EGFR amplification has been seen predominantly in glioblastoma multiforme (GBM) but also in breast, ovarian, prostate and lung carcinomas (Kuan, Wikstrand, and Bigner 2001). Enhanced expression of receptors has been implicated in the progression of cancers but the mechanism of EGFR amplification, however, is unclear. Larger studies in GBMs have identified a mutation in EGFR, known as EGFRvIII, which is commonly present in GBMs with EGFR amplification. EGFRvIII is a gain of function mutation which is constitutively activated (Greenall and Johns 2016). The presence of EGFRvIII has not been widely investigated in AML cases, however recent reports have identified EGFR expression in 15-33% of AML patients (Sun et al. 2012; Mahmud et al. 2016) and in a murine AML tumour (Ben-Ishay 2014). EGF inhibitors such as erlotinib and gefitinib have had mixed results with erlotinib treatment leading to complete remission of AML in two patients (Chan and Pilichowska 2007; Pitini, Arrigo, and Altavilla 2008) while gefitinib treatment had no effect on patient outcome (Deangelo et al. 2014). Further research into the biological mechanisms of EGFR expression in AML is therefore required.

In recent years, however, a wider range of methods of STAT1 protein activation has been identified. STAT1 can be activated by erythropoietin (Kirito et al. 2002), hypoxic conditions (Lee et al. 2006) and mutated kinases. The Philadelphia chromosome translocation, which results in a constitutively activated tyrosine kinase BCR-ABL, has also been shown to target STAT family members for protein activation (Ilaria and Van Etten 1996) and results in the development of leukaemia in murine cases (Daley, Van Etten, and Baltimore 1991). The BCR-ABL fusion oncogene has mostly been reported in chronic myelogenous leukaemia but since 2016 has been included in the WHO

classification as a rare subtype of AML (Neuendorff et al. 2016). Another chromosomal translocation at t(9;12)(p24;p13) produces a TEL-JAK2 fusion protein results in a constitutively active tyrosine kinase (Lacronique et al. 1997). TEL-JAK2 has been reported to induce strong activation of STAT1, STAT3 and STAT5 in HEK 293 cells (Spiekermann et al. 2002). Long-term activation of STAT1 may be the result of autocrine stimulation from, for example, IL-6 from leukemic blasts which results in Stat3 phosphorylation (Schuringa et al. 2000). STAT3 has been shown to cooperate with nuclear EGFR to increase STAT1 expression (Han et al. 2013), and so interleukins may also be responsible for activation of STAT1. However in vitro studies has shown Il-6 to modulate clonogenic blast cell growth in AML cells showing both proliferative and antiproliferative effects depending on the presence of different hematopoietic growth factor combinations (Koistinen et al. 1997).

Activation of Stat1, therefore, can occur by many different mechanisms and expression of Stat1 can have different consequences. It has been proposed that a threshold level of Stat1 expression can occur whereby low levels of Stat1 expression leads to activation of cytotoxic genes and high levels of expression leads to activation of pro-survival genes (Khodarev et al. 2007). This theory is also supported by investigations into the clinical failure of interferon- therapy to treat AML. Interferon treatment of AML showed a higher success rate when high doses (3000 IU/ml) were used during a prolonged period (Benjamin et al. 2007). This prolonged high serum level can be achieved by modification of IFN by conjugation to a polyethylene glycol moiety (pegylated-IFNα), which was not used in previous clinical trials. Use of pegylated IFNα has since resulted in complete remission of AML in two recorded cases, highlighting the rationale for its use in future clinical trials (Smits, Anguille, and Berneman 2013). The role of STAT1 in the development of solid cancers has recently been reviewed and mechanisms by which it achieves this (Meissl et al. 2017). Tumour promoting functions of STAT1 were highlighted, such as, generation of an immunosuppressive tumour microenvironment by mobilisation of myeloid derived suppressor cells and infiltration of tumour-associated macrophages, and promoting tumour growth and invasiveness by interaction with the oncoprotein mucin 1 (MUC1) which is associated with reduced patient survival as well as radio-resistance (Meissl et al. 2017). This overexpression of *STAT1* and its role in cancer progression however has been investigated in cell lines and cancer tissues. Little work has been done on

its role in normal or preleukemic cells. Although much of the research of AML development involves analysis using AML cell lines and primary AML, this work may identify a pathway of interest in the development of preleukemic AML which is still activated in AML at diagnosis. The difference in *Stat1* expression in CBA/Ca and C57BL/6 strains could perhaps be explained by the consistent upregulation in CBA/Ca at 1, 3 and 6 month time-points after irradiation representing a consistent activation of cytotoxic genes. Although a high expression of *Stat1* is not detected in the C57BL/6 strain, perhaps the response is of shorter duration, before 1 month, with pro-survival genes already activated.

A role for *Jak3* in the development of AML is unclear. Previous work has shown that activation of STAT1 without affecting the phosphorylation of JAK3 (Jiang et al. 2011). Also, JAK3 was not found to be constitutively activated in AML cell lines (Spiekermann et al. 2001). In the acute megakaryoblastic leukaemia (AMKL) cell line CMK, JAK3 downregulation was shown to inhibit CMK cell growth and silencing of JAK3 increased apoptosis (Walters et al. 2006). Activating *JAK3* mutations, resulting in the constitutive activation of the JAK3 protein have been identified in AMLK patients but JAK mutations in AML patients have only been in JAK1 and JAK2 (Lee et al. 2013). JAK3 therefore does not appear to play a vital role in AML development.

The mechanism of *Stat1* activation in CBA/Ca HSPCs has yet to be determined. In human AML cell lines STAT1 has been found to be activated by TEL-JAK2, TEL-ABL and BCR-ABL and not activated by FLT3-ITD (Spiekermann et al. 2002). Other common AML features such as PML-RARα showed no change in STAT1 expression and RAS mutations also appeared to have no effect on STAT1 protein levels (Spiekermann et al. 2002). The presence of these translocations, mutations such as EGFRvIII and autocrine interleukin activation should all be investigated as mechanisms of *Stat1* induction in the CBA/Ca mouse after IR. Also DNA methylation, with members of the JAK/STAT pathway identified as having different DNA methylation levels in leukemic samples in comparison to case controls, is also of interest (Jiang et al. 2014).

4 General discussion and future perspectives

AML is an acute cancer of the blood with often limited treatment options. T-AML is a subset of AML which is increasing in number due to increased use of chemotherapy and radiotherapy for a previous malignancy, increased survival of cancer patients with improving cancer treatment and an aging population (Leone et al. 1999). A diagnosis of t-AML has quite a poor prognosis. Options are limited with many studies finding chemo resistance to affect treatment, leaving stem cell transplantation as a final option, however, at this stage many patients are too ill for this treatment. Upon diagnosis, cytogenetic analysis provides the main and most established method of prognosis and risk assessment. However, around 50% of AML patients have a normal karyotype (Byrd et al. 2002; Grimwade et al. 1998) and even with cytogenetic analysis, patients can be classified as having an intermediate risk where outcome can vary. To improve on this classification system other changes such as mutations have been investigated. Genetic analysis has identified many mutated genes such as *FLT3*, *NPM1* and *CEPBA* now also included into the WHO system (Khasawneh and Abdel-Wahab 2014). An understanding of the molecular mechanisms of AML development and the originating leukemic cell is vital in order to develop the appropriate treatment. Here, in order to better characterise rAML, we investigated the genetic and epigenetic alterations that are present in AML samples in human and murine samples taking an in depth look at mutations, gene expression, miRNA expression and DNA methylation.

This study has further demonstrated the common disruption of *PU.1* expression in both murine and human AML. In this study, DNA methylation of the *PU.1* promoter in human AML patients was evident with a reduction of *PU.1* expression also found in one patient. This is also seen in murine cases where a minor pathway without a *PU.1* mutation but with down-regulation of *PU.1* expression has been established. This work clearly validates the use of the CBA mouse as a model of leukemogenesis and the importance of further work into *PU.1* regulation in AML development. The CBA leukemogenic mouse model is characterised by a chromosome 2 deletion, which affects one allele of PU.1 and a point mutation in the DNA binding domain of the second allele. In contrast, this mutation is rare in human AML, however downregulation of PU.1 is evident in human AML cases. Dysregulation of PU.1 is clearly a common path in both human and murine AML. In this study murine cases of rAML without a PU.1 mutation showed transcriptional downregulation. Further

research is needed to identify how this downregulation is caused and if it is a similar mechanism to human AML.

Expression of *Sfpi1* is regulated by many transcription factors such as RUNX1, GATA-1, Oct-1 and Spi-B and it also interacts with many cofactors such as c-Jun, CBP and c-Myb (Gupta et al. 2009). Its dysregulation in the CBA mouse could be due to a number of factors regulating its expression. PU.1 is a major downstream target of RUNX1. RUNX1 binds to sites within the URE of PU.1 and can positively and negatively regulate PU.1 expression (Huang et al. 2008). RUNX1 is also commonly mutated in human AML cases by point mutations and translocations (Osato et al. 1999; Osato 2004; Harada et al. 2004). PU.1 expression is also regulated by a number of factors such as Notch1 (Schroeder et al. 2003) with Notch signalling silenced in human primary AML cells (Lobry et al. 2013) and down-regulation of NOTCH-1 shown to directly decrease PU.1 signalling in human AML cell lines (Chen et al. 2008). GATA-1 and PU.1 are transcription factors for the development of cells into the erythroid or myeloid lineage and mutually inhibit each other. GATA-1 can inhibit PU.1 expression through its promoter (Chou et al. 2009) and, in co-operation with DNMT1, by binding to the URE leading to DNA methylation and H3K9-trimethylation (Burda et al. 2016). The downregulation of PU.1 in AML could therefore be caused by GATA-1, RUNX1 or NOTCH-1.

The DNA methylation analysis in this study highlights the strong epigenetic changes occurring in AML. Histone modifications should also be investigated in rAML cases, in particular in cases with a high level of DNA methylation, to fully investigate if there is a structural change to the chromatin in these cases. Mutations in genes involved in epigenetic modifications such as DNA methylation and chromatin modification has been reported in AML and so the DNA methylation changes in the murine rAML and human AML cases in this study could be indicating a modification of the chromatin structure by dysregulated upstream chromatin modifiers.

Single cell transcriptional analysis of the increased *Stat1* expression in the CBA mouse after irradiation exposure. This work indicates that *Stat1* expression is specific to a CBA response to irradiation. Whether it contributes to AML development in the CBA model remains to be seen. *Stat1* expression has been found to be highly expressed in human AML patient samples and its dysregulation is a key event in a number of

hematological malignancies (Furqan et al. 2013). Indeed, STAT5 has been shown to drive AML development alongside overexpression of HOXA9 in T-cell acute lymphoblastic leukemia (de Bock et al. 2018). The potential role of JAK/STAT dysregulation in murine cases is not well understood. The JAK/STAT and PU.1 pathway may be both instrumental in AML development. A connection between both pathways is not well established, although PU.1 has previously been shown to co-operate with Stat1 in the activation of the FCGR1 gene (Aittomaki et al. 2000). An increase in *PU.1* expression has also been seen with an overexpression of JAK2 in JAK2 V617F mutated cells, possibly through STAT3, STAT5A or STAT5B (Irino et al. 2011). The possibility of a link between *Stat1* and *Pu.1* expression in the CBA RAML cases, however, requires further investigation.

In this work, the identification of activated members of the JAK/STAT pathway following radiation exposure is an exciting finding in HSPC characterisation. Further studies are needed to fully investigate its role. Activity of STAT1 inhibitors, including SOCS, protein inhibitors of activated STATs (PIAS) and nuclear ubiquitin E3 ligases also require investigation. Methylation of the promoter of SOCS1 was found in the leukemic cells of 60% of AML patients in a study involving 53 patients (Chen et al. 2003). Expression of other JAK/STAT members such as *Stat3*, *Stat5* and *Jak2* need to be investigated along with identification of its upstream activator to further characterise the activated pathway. Interaction of STAT1 with STAT3 is of particular interest as STAT3 has been consistently identified as constitutively activated alongside STAT1 in many AML cases (Spiekermann et al. 2001; Spiekermann et al. 2002; Weber-Nordt et al. 1996; Gouilleux-Gruart et al. 1996). They are considered to have opposing roles with STAT-1 initiating proapoptotic functions by inhibiting prosurvival genes Bcl-2 and Bcl-x and STAT3 promoting proliferation by activating them (Stephanou et al. 2000). Their relationship is complex, at times showing opposing functions and at other times reportedly working together in tumour growth, angiogenesis and tumour-associated inflammation (Avalle et al. 2012). In this study the transcriptional levels of *Stat1* expression were analysed but future work must also analyse the protein levels of STAT1 and other Jak/Stat pathway members. Use of mouse models, such as the mCherry mouse model, could be used to study the *Stat1* expression in mice with a chromosome 2 deletion. The CBA *Sfpi1*[GFP/mCh] mouse model has previously been used to study chromosome 2 deletion after irradiation, indicated

by the loss of mCherry expression (Verbiest et al. 2018). To confirm its role in preleukemic cells in AML development *Stat1* expression could be analysed in CBA *Sfpi1*^{GFP/mCh} mice at different time points after irradiation. *Stat1* expression in chromosome 2 deleted cells, indicated by loss of mCherry expression, could be studied using the single cell MQRT-PCR protocol developed in this work and compared to *Stat1* expression in mCherry positive cells for the HSC and MPP populations. Its role could be further analysed by analysing *Stat1* expression in immunodeficient mice after irradiation. It would be of great interest to study and identify the ligand triggering the activation of the JAK/STAT pathway, possibly a chronic inflammation enhanced expression of a cytokine. Further research into this relationship will no doubt shed light on the pathway of leukemic development in hematopoietic stem cells.

Characterisation of the genetic mutations and epigenetic changes in murine AML cases in this work has identified new pathways of mutation. This work focused on specific mutations in commonly mutated genes in human AML. Other mutations could still be contributing to AML development. Whole genome DNA sequencing is needed to fully investigate the mutational landscape of CBA AML cases and the samples could also be screened for transcriptional changes using the mouse PanCancer Pathways panel to identify pathways of interest. Changes in transcriptional expression was detected but protein analysis must also be analysed in future work. This work has confirmed chromosome 2 deletion and *PU.1* mutation as a major pathway of AML development as well as minor pathways of interest, which could also be investigated further. In human AML patients with *FLT3*-ITD, a high expression of *WT1* was also detected (Handschuh et al. 2018). Transcriptional analysis of *WT1* could be analysed in murine cases to investigate if this minor pathway is similar between human and murine cases.

In this study, work on human AML samples was limited by the type and amount of sample received. Human samples of AML, and t-AML in particular, are difficult to obtain due to a bone marrow aspirate procedure being required to obtain a sample. A full patient treatment history is needed to identify t-AML and, more specifically, rAML. This proved to be a difficult task as it requires time and effort to go through confidential files and build a full retrospective treatment history. To continue this work, bone marrow aspirate samples would ideally be obtained for a source of DNA and RNA.

With a larger amount of human AML samples, the DNA methylation and transcriptional level of *PU.1* could be fully investigated. This work would further validate the use of the CBA mouse as a model of leukaemogenesis, if proved true. Samples from different types of AML patients allowing for a comparison between treatment types would also be needed. Mutational and epigenetic analysis could identify signatures specific to the treatment and could add further insights into AML development, chemoresistance and patient prognosis.

Continuation of this work is vital to fully understand the molecular mechanisms occurring in the development of radiation-induced AML, and AML in general. Full investigation of the genetic and epigenetic landscape of AML samples is needed to understand the pathways affected in leukaemogenesis. This information would also improve patient prognosis and risk stratification. By understanding the sequence of these events in AML progression, preleukemic clones could also be identified, allowing the development of AML to be interrupted earlier and treatment administered. Patients could also be monitored for AML progression if preleukemic markers could be determined. Identifying the cell population where AML develops will help the development of specific therapeutic drugs and the targeting of leukaemic cells. This tailored treatment would improve treatment success rates and also patient outcome as it would specifically target leukaemic cells hence avoiding damage to normal HSCs. Further investigation into the development of AML is therefore vital for improved patient treatment and survival.

In summary, we have further established a link between human and murine AML, identifying a rare *Kras* mutated pathway in mouse AML and validated the use of the CBA mouse as a model of leukaemogenesis by illustrating a possible common occurrence of *PU.1/Sfpi1* promoter methylation and transcriptional repression, an alternative pathway to *PU.1* mutations in human AML cases. This work also identified JAK/STAT as an affected pathway after radiation exposure, particularly in the CBA strain which may be a pathway required for radiation induced AML development. Overall this work highlights the interplay between genetic and epigenetic factors, including point mutations, DNA methylation and transcription, driving both human and mouse radiation-induced leukaemogenesis.

Bibliography

Adolfsson, J., O. J. Borge, D. Bryder, K. Theilgaard-Monch, I. Astrand-Grundstrom, E. Sitnicka, Y. Sasaki, and S. E. Jacobsen. 2001. 'Upregulation of Flt3 expression within the bone marrow Lin(-)Sca1(+)c-kit(+) stem cell compartment is accompanied by loss of self-renewal capacity', *Immunity*, 15: 659-69.

Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, and S. R. Sunyaev. 2010. 'A method and server for predicting damaging missense mutations', *Nat Methods*, 7: 248-9.

Agrawal, S., M. Unterberg, S. Koschmieder, U. zur Stadt, U. Brunnberg, W. Verbeek, T. Buchner, W. E. Berdel, H. Serve, and C. Muller-Tidow. 2007. 'DNA methylation of tumor suppressor genes in clinical remission predicts the relapse risk in acute myeloid leukemia', *Cancer Res*, 67: 1370-7.

Aikawa, Y., T. Katsumoto, P. Zhang, H. Shima, M. Shino, K. Terui, E. Ito, H. Ohno, E. R. Stanley, H. Singh, D. G. Tenen, and I. Kitabayashi. 2010. 'PU.1-mediated upregulation of CSF1R is crucial for leukemia stem cell potential induced by MOZ-TIF2', *Nat Med*, 16: 580-5, 1p following 85.

Aittomaki, S., M. Pesu, B. Groner, O. A. Janne, J. J. Palvimo, and O. Silvennoinen. 2000. 'Cooperation among Stat1, glucocorticoid receptor, and PU.1 in transcriptional activation of the high-affinity Fc gamma receptor I in monocytes', *J Immunol*, 164: 5689-97.

Akashi, K., D. Traver, T. Miyamoto, and I. L. Weissman. 2000. 'A clonogenic common myeloid progenitor that gives rise to all myeloid lineages', *Nature*, 404: 193-7.

Alexandrov, L. B., P. H. Jones, D. C. Wedge, J. E. Sale, P. J. Campbell, S. Nik-Zainal, and M. R. Stratton. 2015. 'Clock-like mutational processes in human somatic cells', *Nat Genet*, 47: 1402-7.

Alter, B. P. 2014. 'Fanconi anemia and the development of leukemia', *Best Pract Res Clin Haematol*, 27: 214-21.

Amaravadi, L., and M. J. Klemsz. 1999. 'DNA methylation and chromatin structure regulate PU.1 expression', *DNA Cell Biol*, 18: 875-84.

Amundson, S. A., M. B. Grace, C. B. McLeland, M. W. Epperly, A. Yeager, Q. Zhan, J. S. Greenberger, and A. J. Fornace, Jr. 2004. 'Human in vivo radiation-induced biomarkers: gene expression changes in radiotherapy patients', *Cancer Res*, 64: 6368-71.

Anderson, J. E., T. A. Gooley, G. Schoch, C. Anasetti, W. I. Bensinger, R. A. Clift, J. A. Hansen, J. E. Sanders, R. Storb, and F. R. Appelbaum. 1997. 'Stem cell transplantation for secondary acute myeloid leukemia: evaluation of transplantation as initial therapy or following induction chemotherapy', *Blood*, 89: 2578-85.

Antonchuk, J., G. Sauvageau, and R. K. Humphries. 2001. 'HOXB4 overexpression mediates very rapid stem cell regeneration and competitive hematopoietic repopulation', *Exp Hematol*, 29: 1125-34.

Arbab Jafari, P., H. Ayatollahi, R. Sadeghi, M. Sheikhi, and A. Asghari. 2018. 'Prognostic significance of SRSF2 mutations in myelodysplastic syndromes and chronic myelomonocytic leukemia: a meta-analysis', *Hematology*, 23: 778-84.

Armand, P., H. T. Kim, D. J. DeAngelo, V. T. Ho, C. S. Cutler, R. M. Stone, J. Ritz, E. P. Alyea, J. H. Antin, and R. J. Soiffer. 2007. 'Impact of cytogenetics on outcome of de novo and therapy-related AML and MDS after allogeneic transplantation', *Biol Blood Marrow Transplant*, 13: 655-64.

Aronica, M. G., M. F. Brizzi, P. Dentelli, A. Rosso, Y. Yarden, and L. Pegoraro. 1996. 'p91 STAT1 activation in interleukin-3-stimulated primary acute myeloid leukemia cells', *Oncogene*, 13: 1017-26.

Aslanyan, M. G., L. I. Kroeze, S. M. Langemeijer, T. N. Koorenhof-Scheele, M. Massop, P. van Hoogen, E. Stevens-Linders, L. T. van de Locht, E. Tonnissen, A. van der Heijden, P. da Silva-Coelho, D. Cilloni, G. Saglio, J. P. Marie, R. Tang, B. Labar, S. Amadori, P. Muus, R. Willemze, E. W. Marijt, T. de Witte, B. A. van der Reijden, S. Suciu, and J. H. Jansen. 2014. 'Clinical and biological impact of TET2 mutations and expression in younger adult AML patients treated within the EORTC/GIMEMA AML-12 clinical trial', *Ann Hematol*, 93: 1401-12.

Avalle, L., S. Pensa, G. Regis, F. Novelli, and V. Poli. 2012. 'STAT1 and STAT3 in tumorigenesis: A matter of balance', *JAKSTAT*, 1: 65-72.

Azumi, J. I., and L. Sachs. 1977. 'Chromosome mapping of the genes that control differentiation and malignancy in myeloid leukemic cells', *Proc Natl Acad Sci U S A*, 74: 253-7.

Badie, C., A. Blachowicz, Z. Barjaktarovic, R. Finnon, A. Michaux, H. Sarioglu, N. Brown, G. Manning, M. A. Benotmane, S. Tapio, J. Polanska, and S. D. Bouffler. 2016. 'Transcriptomic and proteomic analysis of mouse radiation-induced acute myeloid leukaemia (AML)', *Oncotarget*, 7: 40461-80.

Ballen, K. K., and J. H. Antin. 1993. 'Treatment of therapy-related acute myelogenous leukemia and myelodysplastic syndromes', *Hematol Oncol Clin North Am*, 7: 477-93.

Bannister, A. J., and T. Kouzarides. 2011. 'Regulation of chromatin by histone modifications', *Cell Res*, 21: 381-95.

Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. 2007. 'High-resolution profiling of histone methylations in the human genome', *Cell*, 129: 823-37.

Basova, P., V. Pospisil, F. Savvulidi, P. Burda, K. Vargova, L. Stanek, M. Dluhosova, E. Kuzmova, A. Jonasova, U. Steidl, P. Laslo, and T. Stopka. 2014. 'Aggressive acute myeloid leukemia in PU.1/p53 double-mutant mice', *Oncogene*, 33: 4735-45.

Beerman, I., C. Bock, B. S. Garrison, Z. D. Smith, H. Gu, A. Meissner, and D. J. Rossi. 2013. 'Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging', *Cell Stem Cell*, 12: 413-25.

Behjati, S., G. Gundem, D. C. Wedge, N. D. Roberts, P. S. Tarpey, S. L. Cooke, P. Van Loo, L. B. Alexandrov, M. Ramakrishna, H. Davies, S. Nik-Zainal, C. Hardy, C. Latimer, K. M. Raine, L. Stebbings, A. Menzies, D. Jones, R. Shepherd, A. P. Butler, J. W. Teague, M. Jorgensen, B. Khatri, N. Pillay, A. Shlien, P. A. Futreal, C. Badie, Icgc Prostate Group, U. McDermott, G. S. Bova, A. L. Richardson, A. M. Flanagan, M. R. Stratton, and P. J. Campbell. 2016. 'Mutational signatures of ionizing radiation in second malignancies', *Nat Commun*, 7: 12605.

Ben-Ishay, Z. 2014. 'Expression of epidermal growth factor receptor by an experimental murine tumor of acute myeloid leukemia origin', *Acta Haematol*, 131: 183-6.

Bendl, J., M. Musil, J. Stourac, J. Zendulka, J. Damborsky, and J. Brezovsky. 2016. 'PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions', *PLoS Comput Biol*, 12: e1004962.

Bendl, J., J. Stourac, O. Salanda, A. Pavelka, E. D. Wieben, J. Zendulka, J. Brezovsky, and J. Damborsky. 2014. 'PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations', *PLoS Comput Biol*, 10: e1003440.

Benekli, M., Z. Xia, K. A. Donohue, L. A. Ford, L. A. Pixley, M. R. Baer, H. Baumann, and M. Wetzler. 2002. 'Constitutive activity of signal transducer and activator of transcription 3 protein in acute myeloid leukemia blasts is associated with short disease-free survival', *Blood*, 99: 252-7.

Benitez, A., W. Liu, A. Palovcak, G. Wang, J. Moon, K. An, A. Kim, K. Zheng, Y. Zhang, F. Bai, A. V. Mazin, X. H. Pei, F. Yuan, and Y. Zhang. 2018. 'FANCA Promotes DNA Double-Strand Break Repair by Catalyzing Single-Strand Annealing and Strand Exchange', *Mol Cell*, 71: 621-28 e4.

Benjamin, R., A. Khwaja, N. Singh, J. McIntosh, A. Meager, M. Wadhwa, C. Streck, C. Ng, A. M. Davidoff, and A. C. Nathwani. 2007. 'Continuous delivery of human type I interferons (alpha/beta) has significant activity against acute myeloid leukemia cells in vitro and in a xenograft model', *Blood*, 109: 1244-7.

Benveniste, P., C. Frelin, S. Janmohamed, M. Barbara, R. Herrington, D. Hyam, and N. N. Iscove. 2010. 'Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential', *Cell Stem Cell*, 6: 48-58.

Besplug, J., P. Burke, A. Ponton, J. Filkowski, V. Titov, I. Kovalchuk, and O. Kovalchuk. 2005. 'Sex and tissue-specific differences in low-dose radiation-induced oncogenic signaling', *Int J Radiat Biol*, 81: 157-68.

Bizzozero, O. J., Jr., K. G. Johnson, and A. Ciocco. 1966. 'Radiation-related leukemia in Hiroshima and Nagasaki, 1946-1964. I. Distribution, incidence and appearance time', *N Engl J Med*, 274: 1095-101.

Bochtler, T., F. Stolzel, C. E. Heilig, C. Kunz, B. Mohr, A. Jauch, J. W. Janssen, M. Kramer, A. Benner, M. Bornhauser, A. D. Ho, G. Ehninger, M. Schaich, and A. Kramer. 2013. 'Clonal heterogeneity as detected by metaphase karyotyping is an indicator of poor prognosis in acute myeloid leukemia', *J Clin Oncol*, 31: 3898-905.

Bogdanovic, O., and G. J. Veenstra. 2009. 'DNA methylation and methyl-CpG binding proteins: developmental requirements and function', *Chromosoma*, 118: 549-65.

Boice, J. D., Jr., M. Blettner, R. A. Kleinerman, M. Stovall, W. C. Moloney, G. Engholm, D. F. Austin, A. Bosch, D. L. Cookfair, E. T. Krementz, H. B. Latourette, L. J. Peters, M. D. Schulz, M. Lundell, F. Pettersson, H. H. Storm, C. M. Bell, M. P. Coleman, P. Fraser, M. Palmer, P. Prior, N. W. Choi, T. G. Hislop, M. Koch, D. Robb, D. Robson, R. F. Spengler, D. von Fournier, R. Frischkorn, H. Lochmuller, V. Pompe-Kirn, A. Rimpela, K. Kjorstad, M. H. Pejovic, K. Sigurdsson, P. Pisani, H. Kucera, and G. B. Hutchison. 1987.

'Radiation dose and leukemia risk in patients treated for cancer of the cervix', *J Natl Cancer Inst*, 79: 1295-311.

Bolouri, H., J. E. Farrar, T. Triche, Jr., R. E. Ries, E. L. Lim, T. A. Alonzo, Y. Ma, R. Moore, A. J. Mungall, M. A. Marra, J. Zhang, X. Ma, Y. Liu, Y. Liu, J. M. G. Auvil, T. M. Davidsen, P. Gesuwan, L. C. Hermida, B. Salhia, S. Capone, G. Ramsingh, C. M. Zwaan, S. Noort, S. R. Piccolo, E. A. Kolb, A. S. Gamis, M. A. Smith, D. S. Gerhard, and S. Meshinchi. 2018. 'The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions', *Nat Med*, 24: 103-12.

Bonadies, N., T. Pabst, and B. U. Mueller. 2010. 'Heterozygous deletion of the PU.1 locus in human AML', *Blood*, 115: 331-4.

British Committee for Standards in Haematology, D. W. Milligan, D. Grimwade, J. O. Cullis, L. Bond, D. Swirsky, C. Craddock, J. Kell, J. Homewood, K. Campbell, S. McGinley, K. Wheatley, and G. Jackson. 2006. 'Guidelines on the management of acute myeloid leukaemia in adults', *Br J Haematol*, 135: 450-74.

Bromberg, J. F., C. M. Horvath, D. Besser, W. W. Lathem, and J. E. Darnell, Jr. 1998. 'Stat3 activation is required for cellular transformation by v-src', *Mol Cell Biol*, 18: 2553-8.

Brown, N., R. Finnon, G. Manning, S. Bouffler, and C. Badie. 2015. 'Influence of radiation quality on mouse chromosome 2 deletions in radiation-induced acute myeloid leukaemia', *Mutat Res Genet Toxicol Environ Mutagen*, 793: 48-54.

Brzoska, K., and M. Kruszewski. 2015. 'Toward the development of transcriptional biodosimetry for the identification of irradiated individuals and assessment of absorbed radiation dose', *Radiat Environ Biophys*, 54: 353-63.

Bullinger, L., K. Dohner, and H. Dohner. 2017. 'Genomics of Acute Myeloid Leukemia Diagnosis and Pathways', *J Clin Oncol*, 35: 934-46.

Burda, P., J. Vargova, N. Curik, C. Salek, G. L. Papadopoulos, J. Strouboulis, and T. Stopka. 2016. 'GATA-1 Inhibits PU.1 Gene via DNA and Histone H3K9 Methylation of Its Distal Enhancer in Erythroleukemia', *PLoS One*, 11: e0152234.

Bustin, S. A., and T. Nolan. 2004. 'Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction', *J Biomol Tech*, 15: 155-66.

Butler, J. S., and S. Y. Dent. 2013. 'The role of chromatin modifiers in normal and malignant hematopoiesis', *Blood*, 121: 3076-84.

Byrd, J. C., K. Mrozek, R. K. Dodge, A. J. Carroll, C. G. Edwards, D. C. Arthur, M. J. Pettenati, S. R. Patil, K. W. Rao, M. S. Watson, P. R. Koduru, J. O. Moore, R. M. Stone, R. J. Mayer, E. J. Feldman, F. R. Davey, C. A. Schiffer, R. A. Larson, C. D. Bloomfield, Cancer, and B. Leukemia Group. 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: 4325-36.

Cabezas-Wallscheid, N., D. Klimmeck, J. Hansson, D. B. Lipka, A. Reyes, Q. Wang, D. Weichenhan, A. Lier, L. von Paleske, S. Renders, P. Wunsche, P. Zeisberger, D. Brocks, L. Gu, C. Herrmann, S. Haas, M. A. G. Essers, B. Brors, R. Eils, W. Huber, M. D. Milsom, C. Plass, J. Krijgsveld, and A.

Trumpp. 2014. 'Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis', *Cell Stem Cell*, 15: 507-22.

Cai, Y., H. Guo, H. Z. Li, W. D. Wang, and Y. S. Zhang. 2017. '[MicroRNA differential expression profile in tuberous sclerosis complex cell line TSC2(-/-) MEFs and normal cell line TSC2(+/+) MEFs]', *Beijing Da Xue Xue Bao Yi Xue Ban*, 49: 580-84.

Cancer Genome Atlas Research, Network, T. J. Ley, C. Miller, L. Ding, B. J. Raphael, A. J. Mungall, A. Robertson, K. Hoadley, T. J. Triche, Jr., P. W. Laird, J. D. Baty, L. L. Fulton, R. Fulton, S. E. Heath, J. Kalicki-Veizer, C. Kandoth, J. M. Klco, D. C. Koboldt, K. L. Kanchi, S. Kulkarni, T. L. Lamprecht, D. E. Larson, L. Lin, C. Lu, M. D. McLellan, J. F. McMichael, J. Payton, H. Schmidt, D. H. Spencer, M. H. Tomasson, J. W. Wallis, L. D. Wartman, M. A. Watson, J. Welch, M. C. Wendl, A. Ally, M. Balasundaram, I. Birol, Y. Butterfield, R. Chiu, A. Chu, E. Chuah, H. J. Chun, R. Corbett, N. Dhalla, R. Guin, A. He, C. Hirst, M. Hirst, R. A. Holt, S. Jones, A. Karsan, D. Lee, H. I. Li, M. A. Marra, M. Mayo, R. A. Moore, K. Mungall, J. Parker, E. Pleasance, P. Plettner, J. Schein, D. Stoll, L. Swanson, A. Tam, N. Thiessen, R. Varhol, N. Wye, Y. Zhao, S. Gabriel, G. Getz, C. Sougnez, L. Zou, M. D. Leiserson, F. Vandin, H. T. Wu, F. Applebaum, S. B. Baylin, R. Akbani, B. M. Broom, K. Chen, T. C. Motter, K. Nguyen, J. N. Weinstein, N. Zhang, M. L. Ferguson, C. Adams, A. Black, J. Bowen, J. Gastier-Foster, T. Grossman, T. Lichtenberg, L. Wise, T. Davidsen, J. A. Demchok, K. R. Shaw, M. Sheth, H. J. Sofia, L. Yang, J. R. Downing, and G. Eley. 2013. 'Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia', *N Engl J Med*, 368: 2059-74.

Cao, Z. H., Q. Y. Zheng, G. Q. Li, X. B. Hu, S. L. Feng, G. L. Xu, and K. Q. Zhang. 2015. 'STAT1-mediated down-regulation of Bcl-2 expression is involved in IFN-gamma/TNF-alpha-induced apoptosis in NIT-1 cells', *PLoS One*, 10: e0120921.

Carow, C. E., M. Levenstein, S. H. Kaufmann, J. Chen, S. Amin, P. Rockwell, L. Witte, M. J. Borowitz, C. I. Civin, and D. Small. 1996. 'Expression of the hematopoietic growth factor receptor FLT3 (STK-1/Flk2) in human leukemias', *Blood*, 87: 1089-96.

Castilla, L. H., L. Garrett, N. Adya, D. Orlic, A. Dutra, S. Anderson, J. Owens, M. Eckhaus, D. Bodine, and P. P. Liu. 1999. 'The fusion gene Cbfb-MYH11 blocks myeloid differentiation and predisposes mice to acute myelomonocytic leukaemia', *Nat Genet*, 23: 144-6.

Chan, G., and M. Pilichowska. 2007. 'Complete remission in a patient with acute myelogenous leukemia treated with erlotinib for non small-cell lung cancer', *Blood*, 110: 1079-80.

Chao, M. P., A. J. Gentles, S. Chatterjee, F. Lan, A. Reinisch, M. R. Corces, S. Xavy, J. Shen, D. Haag, S. Chanda, R. Sinha, R. M. Morganti, T. Nishimura, M. Ameen, H. Wu, M. Wernig, J. C. Wu, and R. Majeti. 2017. 'Human AML-iPSCs Reacquire Leukemic Properties after Differentiation and Model Clonal Variation of Disease', *Cell Stem Cell*, 20: 329-44 e7.

Chen, C. L., K. Faltusova, M. Molik, F. Savvulidi, K. T. Chang, and E. Necas. 2016. 'Low c-Kit Expression Level Induced by Stem Cell Factor Does Not Compromise Transplantation of Hematopoietic Stem Cells', *Biol Blood Marrow Transplant*, 22: 1167-72.

Chen, C. Y., W. Tsay, J. L. Tang, H. L. Shen, S. W. Lin, S. Y. Huang, M. Yao, Y. C. Chen, M. C. Shen, C. H. Wang, and H. F. Tien. 2003. 'SOCS1 methylation in patients with newly diagnosed acute myeloid leukemia', *Genes Chromosomes Cancer*, 37: 300-5.

Chen, H. M., P. Zhang, M. T. Voso, S. Hohaus, D. A. Gonzalez, C. K. Glass, D. E. Zhang, and D. G. Tenen. 1995. 'Neutrophils and monocytes express high levels of PU.1 (Spi-1) but not Spi-B', *Blood*, 85: 2918-28.

Chen, H., D. Ray-Gallet, P. Zhang, C. J. Hetherington, D. A. Gonzalez, D. E. Zhang, F. Moreau-Gachelin, and D. G. Tenen. 1995. 'PU.1 (Spi-1) autoregulates its expression in myeloid cells', *Oncogene*, 11: 1549-60.

Chen, H. Y., S. L. Yu, C. H. Chen, G. C. Chang, C. Y. Chen, A. Yuan, C. L. Cheng, C. H. Wang, H. J. Terng, S. F. Kao, W. K. Chan, H. N. Li, C. C. Liu, S. Singh, W. J. Chen, J. J. Chen, and P. C. Yang. 2007. 'A five-gene signature and clinical outcome in non-small-cell lung cancer', *N Engl J Med*, 356: 11-20.

Chen, J. Y., M. Miyanishi, S. K. Wang, S. Yamazaki, R. Sinha, K. S. Kao, J. Seita, D. Sahoo, H. Nakauchi, and I. L. Weissman. 2016. 'Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche', *Nature*, 530: 223-7.

Chen, P. M., C. C. Yen, W. S. Wang, Y. J. Lin, C. J. Chu, T. J. Chiou, J. H. Liu, and M. H. Yang. 2008. 'Down-regulation of Notch-1 expression decreases PU.1-mediated myeloid differentiation signaling in acute myeloid leukemia', *Int J Oncol*, 32: 1335-41.

Cheng, T., H. Shen, D. Giokas, J. Gere, D. G. Tenen, and D. T. Scadden. 1996. 'Temporal mapping of gene expression levels during the differentiation of individual primary hematopoietic cells', *Proc Natl Acad Sci U S A*, 93: 13158-63.

Chin, Y. E., M. Kitagawa, W. C. Su, Z. H. You, Y. Iwamoto, and X. Y. Fu. 1996. 'Cell growth arrest and induction of cyclin-dependent kinase inhibitor p21 WAF1/CIP1 mediated by STAT1', *Science*, 272: 719-22.

Choi, Y., and A. P. Chan. 2015. 'PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels', *Bioinformatics*, 31: 2745-7.

Chou, S. T., E. Khandros, L. C. Bailey, K. E. Nichols, C. R. Vakoc, Y. Yao, Z. Huang, J. D. Crispino, R. C. Hardison, G. A. Blobel, and M. J. Weiss. 2009. 'Graded repression of PU.1/Sfpi1 gene transcription by GATA factors regulates hematopoietic cell fate', *Blood*, 114: 983-94.

Chou, W. C., S. C. Chou, C. Y. Liu, C. Y. Chen, H. A. Hou, Y. Y. Kuo, M. C. Lee, B. S. Ko, J. L. Tang, M. Yao, W. Tsay, S. J. Wu, S. Y. Huang, S. C. Hsu, Y. C. Chen, Y. C. Chang, Y. Y. Kuo, K. T. Kuo, F. Y. Lee, M. C. Liu, C. W. Liu, M. H. Tseng, C. F. Huang, and H. F. Tien. 2011. 'TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics', *Blood*, 118: 3803-10.

Christiansen, D. H., M. K. Andersen, F. Desta, and J. Pedersen-Bjergaard. 2005. 'Mutations of genes in the receptor tyrosine kinase (RTK)/RAS-BRAF signal transduction pathway in therapy-related myelodysplasia and acute myeloid leukemia', *Leukemia*, 19: 2232-40.

Clark, D. J., E. I. Meijne, S. D. Bouffler, R. Huiskamp, C. J. Skidmore, R. Cox, and A. R. Silver. 1996. 'Microsatellite analysis of recurrent chromosome 2 deletions in acute myeloid leukaemia induced by radiation in F1 hybrid mice', *Genes Chromosomes Cancer*, 16: 238-46.

Claus, R., and M. Lubbert. 2003. 'Epigenetic targets in hematopoietic malignancies', *Oncogene*, 22: 6489-96.

Coffer, P. J., L. Koenderman, and R. P. de Groot. 2000. 'The role of STATs in myeloid differentiation and leukemia', *Oncogene*, 19: 2511-22.

Cook, W. D., B. J. McCaw, C. Herring, D. L. John, S. J. Foote, S. L. Nutt, and J. M. Adams. 2004. 'PU.1 is a suppressor of myeloid leukemia, inactivated in mice by gene deletion and mutation of its DNA binding domain', *Blood*, 104: 3437-44.

Corces-Zimmerman, M. R., W. J. Hong, I. L. Weissman, B. C. Medeiros, and R. Majeti. 2014. 'Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission', *Proc Natl Acad Sci U S A*, 111: 2548-53.

Cristobal, I., L. Garcia-Orti, C. Cirauqui, X. Cortes-Lavaud, M. A. Garcia-Sanchez, M. J. Calasanz, and M. D. Odero. 2012. 'Overexpression of SET is a recurrent event associated with poor outcome and contributes to protein phosphatase 2A inhibition in acute myeloid leukemia', *Haematologica*, 97: 543-50.

Daley, G. Q., R. A. Van Etten, and D. Baltimore. 1991. 'Blast crisis in a murine model of chronic myelogenous leukemia', *Proc Natl Acad Sci U S A*, 88: 11335-8.

Daraki, A., S. Zachaki, T. Koromila, P. Diamantopoulou, G. E. Pantelias, C. Sambani, V. Aleporou, P. Kollia, and K. N. Manola. 2014. 'The G(5)(1)(6)T CYP2B6 germline polymorphism affects the risk of acute myeloid leukemia and is associated with specific chromosomal abnormalities', *PLoS One*, 9: e88879.

Darby, S. C., G. Reeves, T. Key, R. Doll, and M. Stovall. 1994. 'Mortality in a cohort of women given X-ray therapy for metropathia haemorrhagica', *Int J Cancer*, 56: 793-801.

Daubner, G. M., A. Clery, S. Jayne, J. Stevenin, and F. H. Allain. 2012. 'A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well', *EMBO J*, 31: 162-74.

de Bock, C. E., S. Demeyer, S. Degryse, D. Verbeke, B. Sweron, O. Gielen, R. Vandepoel, C. Vicente, M. Vanden Bempt, A. Dagklis, E. Geerdens, S. Bornschein, R. Gijsbers, J. Soulier, J. P. Meijerink, M. Heinaniemi, S. Teppo, M. Bouvy-Liivrand, O. Lohi, E. Radaelli, and J. Cools. 2018. 'HOXA9 Cooperates with Activated JAK/STAT Signaling to Drive Leukemia Development', *Cancer Discov*, 8: 616-31.

De Kouchkovsky, I., and M. Abdul-Hay. 2016. ''Acute myeloid leukemia: a comprehensive review and 2016 update'', *Blood Cancer J*, 6: e441.

Deangelo, D. J., D. Neuberg, P. C. Amrein, J. Berchuck, M. Wadleigh, L. A. Sirulnik, I. Galinsky, T. Golub, K. Stegmaier, and R. M. Stone. 2014. 'A phase II study of the EGFR inhibitor gefitinib in patients with acute myeloid leukemia', *Leuk Res*, 38: 430-4.

Delhommeau, F., S. Dupont, V. Della Valle, C. James, S. Trannoy, A. Masse, O. Kosmider, J. P. Le Couedic, F. Robert, A. Alberdi, Y. Lecluse, I. Plo, F. J. Dreyfus, C. Marzac, N. Casadevall, C. Lacombe, S. P. Romana, P. Dessen, J. Soulier, F. Viguie, M. Fontenay, W. Vainchenker, and O. A. Bernard. 2009. 'Mutation in TET2 in myeloid cancers', *N Engl J Med*, 360: 2289-301.

Dimberg, A., I. Karlberg, K. Nilsson, and F. Oberg. 2003. 'Ser727/Tyr701-phosphorylated Stat1 is required for the regulation of c-Myc, cyclins, and

p27Kip1 associated with ATRA-induced G0/G1 arrest of U-937 cells', *Blood*, 102: 254-61.

Ding, L., T. J. Ley, D. E. Larson, C. A. Miller, D. C. Koboldt, J. S. Welch, J. K. Ritchey, M. A. Young, T. Lamprecht, M. D. McLellan, J. F. McMichael, J. W. Wallis, C. Lu, D. Shen, C. C. Harris, D. J. Dooling, R. S. Fulton, L. L. Fulton, K. Chen, H. Schmidt, J. Kalicki-Veizer, V. J. Magrini, L. Cook, S. D. McGrath, T. L. Vickery, M. C. Wendl, S. Heath, M. A. Watson, D. C. Link, M. H. Tomasson, W. D. Shannon, J. E. Payton, S. Kulkarni, P. Westervelt, M. J. Walter, T. A. Graubert, E. R. Mardis, R. K. Wilson, and J. F. DiPersio. 2012. 'Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing', *Nature*, 481: 506-10.

Dohner, H., E. Estey, D. Grimwade, S. Amadori, F. R. Appelbaum, T. Buchner, H. Dombret, B. L. Ebert, P. Fenaux, R. A. Larson, R. L. Levine, F. Lo-Coco, T. Naoe, D. Niederwieser, G. J. Ossenkoppele, M. Sanz, J. Sierra, M. S. Tallman, H. F. Tien, A. H. Wei, B. Lowenberg, and C. D. Bloomfield. 2017. 'Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel', *Blood*, 129: 424-47.

Dohner, H., E. H. Estey, S. Amadori, F. R. Appelbaum, T. Buchner, A. K. Burnett, H. Dombret, P. Fenaux, D. Grimwade, R. A. Larson, F. Lo-Coco, T. Naoe, D. Niederwieser, G. J. Ossenkoppele, M. A. Sanz, J. Sierra, M. S. Tallman, B. Lowenberg, C. D. Bloomfield, and LeukemiaNet European. 2010. 'Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet', *Blood*, 115: 453-74.

Dohner, H., D. J. Weisdorf, and C. D. Bloomfield. 2015. 'Acute Myeloid Leukemia', *N Engl J Med*, 373: 1136-52.

Dohner, K., and H. Dohner. 2008. 'Molecular characterization of acute myeloid leukemia', *Haematologica*, 93: 976-82.

Dohner, K., R. F. Schlenk, M. Habdank, C. Scholl, F. G. Rucker, A. Corbacioglu, L. Bullinger, S. Frohling, and H. Dohner. 2005. 'Mutant nucleophosmin (NPM1) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations', *Blood*, 106: 3740-6.

Druz, A., Y. C. Chen, R. Guha, M. Betenbaugh, S. E. Martin, and J. Shiloach. 2013. 'Large-scale screening identifies a novel microRNA, miR-15a-3p, which induces apoptosis in human cancer cell lines', *RNA Biol*, 10: 287-300.

Du, Q., P. L. Luu, C. Stirzaker, and S. J. Clark. 2015. 'Methyl-CpG-binding domain proteins: readers of the epigenome', *Epigenomics*, 7: 1051-73.

Edinger, R. S., C. Coronnello, A. J. Bodnar, M. Labarca, V. Bhalla, W. A. LaFramboise, P. V. Benos, J. Ho, J. P. Johnson, and M. B. Butterworth. 2014. 'Aldosterone regulates microRNAs in the cortical collecting duct to alter sodium transport', *J Am Soc Nephrol*, 25: 2445-57.

Eis, P. S., W. Tam, L. Sun, A. Chadburn, Z. Li, M. F. Gomez, E. Lund, and J. E. Dahlberg. 2005. 'Accumulation of miR-155 and BIC RNA in human B cell lymphomas', *Proc Natl Acad Sci U S A*, 102: 3627-32.

Eppert, K., K. Takenaka, E. R. Lechman, L. Waldron, B. Nilsson, P. van Galen, K. H. Metzeler, A. Poeppl, V. Ling, J. Beyene, A. J. Canty, J. S. Danska, S. K. Bohlander, C. Buske, M. D. Minden, T. R. Golub, I. Jurisica, B. L. Ebert, and J. E. Dick. 2011. 'Stem cell gene expression programs influence clinical outcome in human leukemia', *Nat Med*, 17: 1086-93.

Evans, D. I., and J. K. Steward. 1972. 'Down's syndrome and leukaemia', *Lancet*, 2: 1322.

Faderl, S., A. Ferrajoli, D. Harris, Q. Van, H. M. Kantarjian, and Z. Estrov. 2007. 'Atiprimod blocks phosphorylation of JAK-STAT and inhibits proliferation of acute myeloid leukemia (AML) cells', *Leuk Res*, 31: 91-5.

Fang, L., J. Cai, B. Chen, S. Wu, R. Li, X. Xu, Y. Yang, H. Guan, X. Zhu, L. Zhang, J. Yuan, J. Wu, and M. Li. 2015. 'Aberrantly expressed miR-582-3p maintains lung cancer stem cell-like traits by activating Wnt/beta-catenin signalling', *Nat Commun*, 6: 8640.

Faraoni, I., S. Laterza, D. Ardiri, C. Ciardi, F. Fazi, and F. Lo-Coco. 2012. 'MiR-424 and miR-155 deregulated expression in cytogenetically normal acute myeloid leukaemia: correlation with NPM1 and FLT3 mutation status', *J Hematol Oncol*, 5: 26.

Feng, Y., X. Li, K. Cassady, Z. Zou, and X. Zhang. 2019. 'TET2 Function in Hematopoietic Malignancies, Immune Regulation, and DNA Repair', *Front Oncol*, 9: 210.

Fernandez-Mercado, M., B. H. Yip, A. Pellagatti, C. Davies, M. J. Larrayoz, T. Kondo, C. Perez, S. Killick, E. J. McDonald, M. D. Odero, X. Agirre, F. Prosper, M. J. Calasanz, J. S. Wainscoat, and J. Boultwood. 2012. 'Mutation patterns of 16 genes in primary and secondary acute myeloid leukemia (AML) with normal cytogenetics', *PLoS One*, 7: e42334.

Fernandez-Nestosa, M. J., E. Monturus, Z. Sanchez, F. S. Torres, A. F. Fernandez, M. F. Fraga, P. Hernandez, J. B. Schvartzman, and D. B. Krimer. 2013. 'DNA methylation-mediated silencing of PU.1 in leukemia cells resistant to cell differentiation', *Springerplus*, 2: 392.

Ferreira, H. J., H. Heyn, M. Vizoso, C. Moutinho, E. Vidal, A. Gomez, A. Martinez-Cardus, L. Simo-Riudalbas, S. Moran, E. Jost, and M. Esteller. 2016. 'DNMT3A mutations mediate the epigenetic reactivation of the leukemogenic factor MEIS1 in acute myeloid leukemia', *Oncogene*, 35: 3079-82.

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

Figueroa, M. E., L. Skrabanek, Y. Li, A. Jiemjit, T. E. Fandy, E. Paietta, H. Fernandez, M. S. Tallman, J. M. Greally, H. Carraway, J. D. Licht, S. D. Gore, and A. Melnick. 2009. 'MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation', *Blood*, 114: 3448-58.

Finnon, R., N. Brown, J. Moody, C. Badie, C. H. Olme, R. Huiskamp, E. Meijne, M. Sutmuller, M. Rosemann, and S. D. Bouffler. 2012. 'Flt3-ITD mutations in a mouse model of radiation-induced acute myeloid leukaemia', *Leukemia*, 26: 1445-6.

Forsberg, E. C., S. S. Prohaska, S. Katzman, G. C. Heffner, J. M. Stuart, and I. L. Weissman. 2005. 'Differential expression of novel potential regulators in hematopoietic stem cells', *PLoS Genet*, 1: e28.

Foucar, K.; Reichard, K.; Czuchlewski, D. 2010. 'Acute Myeloid Leukaemia.' in, *Acute Myeloid Leukaemia: Bone Marrow Pathology* (ASCP: Chicago, IL).

Fried, I., C. Bodner, M. M. Pichler, K. Lind, C. Beham-Schmid, F. Quehenberger, W. R. Sperr, W. Linkesch, H. Sill, and A. Wolfler. 2012. 'Frequency, onset and clinical impact of somatic DNMT3A mutations in therapy-related and secondary acute myeloid leukemia', *Haematologica*, 97: 246-50.

Fulda, S., and K. M. Debatin. 2002. 'IFNgamma sensitizes for apoptosis by upregulating caspase-8 expression through the Stat1 pathway', *Oncogene*, 21: 2295-308.

Furqan, M., N. Mukhi, B. Lee, and D. Liu. 2013. 'Dysregulation of JAK-STAT pathway in hematological malignancies and JAK inhibitors for clinical application', *Biomark Res*, 1: 5.

Gaidzik, V. I., P. Paschka, D. Spath, M. Habdank, C. H. Kohne, U. Germing, M. von Lilienfeld-Toal, G. Held, H. A. Horst, D. Haase, M. Bentz, K. Gotze, H. Dohner, R. F. Schlenk, L. Bullinger, and K. Dohner. 2012. 'TET2 mutations in acute myeloid leukemia (AML): results from a comprehensive genetic and clinical analysis of the AML study group', *J Clin Oncol*, 30: 1350-7.

Gaidzik, V. I., V. Teleanu, E. Papaemmanuil, D. Weber, P. Paschka, J. Hahn, T. Wallrabenstein, B. Kolbinger, C. H. Kohne, H. A. Horst, P. Brossart, G. Held, A. Kundgen, M. Ringhoffer, K. Gotze, M. Rummel, M. Gerstung, P. Campbell, J. M. Kraus, H. A. Kestler, F. Thol, M. Heuser, B. Schlegelberger, A. Ganser, L. Bullinger, R. F. Schlenk, K. Dohner, and H. Dohner. 2016. 'RUNX1 mutations in acute myeloid leukemia are associated with distinct clinico-pathologic and genetic features', *Leukemia*, 30: 2282.

Genovese, G., A. K. Kahler, R. E. Handsaker, J. Lindberg, S. A. Rose, S. F. Bakhoum, K. Chambert, E. Mick, B. M. Neale, M. Fromer, S. M. Purcell, O. Svantesson, M. Landen, M. Hoglund, S. Lehmann, S. B. Gabriel, J. L. Moran, E. S. Lander, P. F. Sullivan, P. Sklar, H. Gronberg, C. M. Hultman, and S. A. McCarroll. 2014. 'Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence', *N Engl J Med*, 371: 2477-87.

Gerloff, D., R. Grundler, A. A. Wurm, D. Brauer-Hartmann, C. Katzerke, J. U. Hartmann, V. Madan, C. Muller-Tidow, J. Duyster, D. G. Tenen, D. Niederwieser, and G. Behre. 2015. 'NF-kappaB/STAT5/miR-155 network targets PU.1 in FLT3-ITD-driven acute myeloid leukemia', *Leukemia*, 29: 535-47.

Ghavifekr Fakhr, M., M. Farshdousti Hagh, D. Shanehbandi, and B. Baradaran. 2013. 'DNA methylation pattern as important epigenetic criterion in cancer', *Genet Res Int*, 2013: 317569.

Giustacchini, A., S. Thongjuea, N. Barkas, P. S. Woll, B. J. Povinelli, C. A. G. Booth, P. Sopp, R. Norfo, A. Rodriguez-Meira, N. Ashley, L. Jamieson, P. Vyas, K. Anderson, A. Segerstolpe, H. Qian, U. Olsson-Stromberg, S. Mustjoki, R. Sandberg, S. E. W. Jacobsen, and A. J. Mead. 2017. 'Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia', *Nat Med*, 23: 692-702.

Gouilleux-Gruart, V., F. Gouilleux, C. Desaint, J. F. Claisse, J. C. Capiod, J. Delobel, R. Weber-Nordt, I. Dusanter-Fourt, F. Dreyfus, B. Groner, and L. Prin. 1996. 'STAT-related transcription factors are constitutively activated in peripheral blood cells from acute leukemia patients', *Blood*, 87: 1692-7.

Greaves, M. F. 1997. 'Aetiology of acute leukaemia', *Lancet*, 349: 344-9.

Greenall, S. A., and T. G. Johns. 2016. 'EGFRvIII: the promiscuous mutation', *Cell Death Discov*, 2: 16049.

Grimwade, D., H. Walker, F. Oliver, K. Wheatley, C. Harrison, G. Harrison, J. Rees, I. Hann, R. Stevens, A. Burnett, and A. Goldstone. 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: 2322-33.

Guerrero, I., A. Villasante, V. Corces, and A. Pellicer. 1984. 'Activation of a c-K-ras oncogene by somatic mutation in mouse lymphomas induced by gamma radiation', *Science*, 225: 1159-62.

Gupta, P., G. U. Gurudutta, D. Saluja, and R. P. Tripathi. 2009. 'PU.1 and partners: regulation of haematopoietic stem cell fate in normal and malignant haematopoiesis', *J Cell Mol Med*, 13: 4349-63.

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

Han, W., R. L. Carpenter, X. Cao, and H. W. Lo. 2013. 'STAT1 gene expression is enhanced by nuclear EGFR and HER2 via cooperation with STAT3', *Mol Carcinog*, 52: 959-69.

Handschuh, L., M. Kazmierczak, M. C. Milewski, M. Goralski, M. Luczak, M. Wojtaszewska, B. Uszczynska-Ratajczak, K. Lewandowski, M. Komarnicki, and M. Figlerowicz. 2018. 'Gene expression profiling of acute myeloid leukemia samples from adult patients with AML-M1 and -M2 through boutique microarrays, real-time PCR and droplet digital PCR', *Int J Oncol*, 52: 656-78.

Harada, H., Y. Harada, H. Niimi, T. Kyo, A. Kimura, and T. Inaba. 2004. 'High incidence of somatic mutations in the AML1/RUNX1 gene in myelodysplastic syndrome and low blast percentage myeloid leukemia with myelodysplasia', *Blood*, 103: 2316-24.

Heaney, M. L., and D. W. Golde. 1999. 'Myelodysplasia', *N Engl J Med*, 340: 1649-60.

Henry-Amar, M., and P. Y. Dietrich. 1993. 'Acute leukemia after the treatment of Hodgkin's disease', *Hematol Oncol Clin North Am*, 7: 369-87.

Heuser, M., L. M. Sly, B. Argiropoulos, F. Kuchenbauer, C. Lai, A. Weng, M. Leung, G. Lin, C. Brookes, S. Fung, P. J. Valk, R. Delwel, B. Lowenberg, G. Krystal, and R. K. Humphries. 2009. 'Modeling the functional heterogeneity of leukemia stem cells: role of STAT5 in leukemia stem cell self-renewal', *Blood*, 114: 3983-93.

Heyn, H., N. Li, H. J. Ferreira, S. Moran, D. G. Pisano, A. Gomez, J. Diez, J. V. Sanchez-Mut, F. Setien, F. J. Carmona, A. A. Puca, S. Sayols, M. A. Pujana, J. Serra-Musach, I. Iglesias-Platas, F. Formiga, A. F. Fernandez, M. F. Fraga, S. C. Heath, A. Valencia, I. G. Gut, J. Wang, and M. Esteller. 2012. 'Distinct DNA methylomes of newborns and centenarians', *Proc Natl Acad Sci U S A*, 109: 10522-7.

Hohaus, S., M. S. Petrovick, M. T. Voso, Z. Sun, D. E. Zhang, and D. G. Tenen. 1995. 'PU.1 (Spi-1) and C/EBP alpha regulate expression of the granulocyte-macrophage colony-stimulating factor receptor alpha gene', *Mol Cell Biol*, 15: 5830-45.

Hosokawa, H., J. Ungerback, X. Wang, M. Matsumoto, K. I. Nakayama, S. M. Cohen, T. Tanaka, and E. V. Rothenberg. 2018. 'Transcription Factor PU.1

Represses and Activates Gene Expression in Early T Cells by Redirecting Partner Transcription Factor Binding', *Immunity*, 49: 782.

Hou, H. A., Y. Y. Kuo, C. Y. Liu, W. C. Chou, M. C. Lee, C. Y. Chen, L. I. Lin, M. H. Tseng, C. F. Huang, Y. C. Chiang, F. Y. Lee, M. C. Liu, C. W. Liu, J. L. Tang, M. Yao, S. Y. Huang, B. S. Ko, S. C. Hsu, S. J. Wu, W. Tsay, Y. C. Chen, and H. F. Tien. 2012. 'DNMT3A mutations in acute myeloid leukemia: stability during disease evolution and clinical implications', *Blood*, 119: 559-68.

Hsu, W. L., D. L. Preston, M. Soda, H. Sugiyama, S. Funamoto, K. Kodama, A. Kimura, N. Kamada, H. Dohy, M. Tomonaga, M. Iwanaga, Y. Miyazaki, H. M. Cullings, A. Suyama, K. Ozasa, R. E. Shore, and K. Mabuchi. 2013. 'The incidence of leukemia, lymphoma and multiple myeloma among atomic bomb survivors: 1950-2001', *Radiat Res*, 179: 361-82.

Huang, G., P. Zhang, H. Hirai, S. Elf, X. Yan, Z. Chen, S. Koschmieder, Y. Okuno, T. Dayaram, J. D. Growney, R. A. Shivdasani, D. G. Gilliland, N. A. Speck, S. D. Nimer, and D. G. Tenen. 2008. 'PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis', *Nat Genet*, 40: 51-60.

Ilaria, R. L., Jr., and R. A. Van Etten. 1996. 'P210 and P190(BCR/ABL) induce the tyrosine phosphorylation and DNA binding activity of multiple specific STAT family members', *J Biol Chem*, 271: 31704-10.

Illmer, T., C. Thiede, A. Fredersdorf, S. Stadler, A. Neubauer, G. Ehninger, and M. Schaich. 2005. 'Activation of the RAS pathway is predictive for a chemosensitive phenotype of acute myelogenous leukemia blasts', *Clin Cancer Res*, 11: 3217-24.

Inomata, M., S. Takahashi, H. Harigae, J. Kameoka, M. Kaku, and T. Sasaki. 2006. 'Inverse correlation between Flt3 and PU.1 expression in acute myeloblastic leukemias', *Leuk Res*, 30: 659-64.

Irino, T., M. Uemura, H. Yamane, S. Umemura, T. Utsumi, N. Kakazu, T. Shirakawa, M. Ito, T. Suzuki, and K. Kinoshita. 2011. 'JAK2 V617F-dependent upregulation of PU.1 expression in the peripheral blood of myeloproliferative neoplasm patients', *PLoS One*, 6: e22148.

Jacobs, K. B., M. Yeager, W. Zhou, S. Wacholder, Z. Wang, B. Rodriguez-Santiago, A. Hutchinson, X. Deng, C. Liu, M. J. Horner, M. Cullen, C. G. Epstein, L. Burdett, M. C. Dean, N. Chatterjee, J. Sampson, C. C. Chung, J. Kovaks, S. M. Gapstur, V. L. Stevens, L. T. Teras, M. M. Gaudet, D. Albanes, S. J. Weinstein, J. Virtamo, P. R. Taylor, N. D. Freedman, C. C. Abnet, A. M. Goldstein, N. Hu, K. Yu, J. M. Yuan, L. Liao, T. Ding, Y. L. Qiao, Y. T. Gao, W. P. Koh, Y. B. Xiang, Z. Z. Tang, J. H. Fan, M. C. Aldrich, C. Amos, W. J. Blot, C. H. Bock, E. M. Gillanders, C. C. Harris, C. A. Haiman, B. E. Henderson, L. N. Kolonel, L. Le Marchand, L. H. McNeill, B. A. Rybicki, A. G. Schwartz, L. B. Signorello, M. R. Spitz, J. K. Wiencke, M. Wrensch, X. Wu, K. A. Zanetti, R. G. Ziegler, J. D. Figueroa, M. Garcia-Closas, N. Malats, G. Marenne, L. Prokunina-Olsson, D. Baris, M. Schwenn, A. Johnson, M. T. Landi, L. Goldin, D. Consonni, P. A. Bertazzi, M. Rotunno, P. Rajaraman, U. Andersson, L. E. Beane Freeman, C. D. Berg, J. E. Buring, M. A. Butler, T. Carreon, M. Feychting, A. Ahlbom, J. M. Gaziano, G. G. Giles, G. Hallmans, S. E. Hankinson, P. Hartge, R. Henriksson, P. D. Inskip, C. Johansen, A. Landgren, R. McKean-Cowdin, D. S. Michaud, B. S. Melin, U. Peters, A. M. Ruder, H. D. Sesso, G. Severi, X. O. Shu, K. Visvanathan, E. White, A. Wolk, A. Zeleniuch-Jacquotte, W.

Zheng, D. T. Silverman, M. Kogevinas, J. R. Gonzalez, O. Villa, D. Li, E. J. Duell, H. A. Risch, S. H. Olson, C. Kooperberg, B. M. Wolpin, L. Jiao, M. Hassan, W. Wheeler, A. A. Arslan, H. B. Bueno-de-Mesquita, C. S. Fuchs, S. Gallinger, M. D. Gross, E. A. Holly, A. P. Klein, A. LaCroix, M. T. Mandelson, G. Petersen, M. C. Boutron-Ruault, P. M. Bracci, F. Canzian, K. Chang, M. Cotterchio, E. L. Giovannucci, M. Goggins, J. A. Hoffman Bolton, M. Jenab, K. T. Khaw, V. Krogh, R. C. Kurtz, R. R. McWilliams, J. B. Mendelsohn, K. G. Rabe, E. Riboli, A. Tjonneland, G. S. Tobias, D. Trichopoulos, J. W. Elena, H. Yu, L. Amundadottir, R. Z. Stolzenberg-Solomon, P. Kraft, F. Schumacher, D. Stram, S. A. Savage, L. Mirabello, I. L. Andrulis, J. S. Wunder, A. Patino Garcia, L. Sierrasesumaga, D. A. Barkauskas, R. G. Gorlick, M. Purdue, W. H. Chow, L. E. Moore, K. L. Schwartz, F. G. Davis, A. W. Hsing, S. I. Berndt, A. Black, N. Wentzensen, L. A. Brinton, J. Lissowska, B. Peplonska, K. A. McGlynn, M. B. Cook, B. I. Graubard, C. P. Kratz, M. H. Greene, R. L. Erickson, D. J. Hunter, G. Thomas, R. N. Hoover, F. X. Real, J. F. Fraumeni, Jr., N. E. Caporaso, M. Tucker, N. Rothman, L. A. Perez-Jurado, and S. J. Chanock. 2012. 'Detectable clonal mosaicism and its relationship to aging and cancer', *Nat Genet*, 44: 651-8.

Jaiswal, S., P. Fontanillas, J. Flannick, A. Manning, P. V. Grauman, B. G. Mar, R. C. Lindsley, C. H. Mermel, N. Burtt, A. Chavez, J. M. Higgins, V. Moltchanov, F. C. Kuo, M. J. Kluk, B. Henderson, L. Kinnunen, H. A. Koistinen, C. Ladenvall, G. Getz, A. Correa, B. F. Banahan, S. Gabriel, S. Kathiresan, H. M. Stringham, M. I. McCarthy, M. Boehnke, J. Tuomilehto, C. Haiman, L. Groop, G. Atzmon, J. G. Wilson, D. Neuberg, D. Altshuler, and B. L. Ebert. 2014. 'Age-related clonal hematopoiesis associated with adverse outcomes', *N Engl J Med*, 371: 2488-98.

Jalili, M., M. Yaghmaie, M. Ahmadvand, K. Alimoghaddam, S. A. Mousavi, M. Vaezi, and A. Ghavamzadeh. 2018. 'Prognostic Value of RUNX1 Mutations in AML: A Meta-Analysis', *Asian Pac J Cancer Prev*, 19: 325-29.

Jiang, D., Q. Hong, Y. Shen, Y. Xu, H. Zhu, Y. Li, C. Xu, G. Ouyang, and S. Duan. 2014. 'The diagnostic value of DNA methylation in leukemia: a systematic review and meta-analysis', *PLoS One*, 9: e96822.

Jiang, Lin-Jia, Nan-Nan Zhang, Fei Ding, Xian-Yang Li, Lei Chen, Hong-Xin Zhang, Wu Zhang, Sai-Juan Chen, Zhu-Gang Wang, Jun-Min Li, Zhu Chen, and Jiang Zhu. 2011. 'RA-inducible gene-I induction augments STAT1 activation to inhibit leukemia cell proliferation', *Proceedings of the National Academy of Sciences of the United States of America*, 108: 1897-902.

Jin, X., Y. Guan, H. Sheng, and Y. Liu. 2017. 'Crosstalk in competing endogenous RNA network reveals the complex molecular mechanism underlying lung cancer', *Oncotarget*, 8: 91270-80.

Kabacik, S., A. Ortega-Molina, A. Efeyan, P. Finnon, S. Bouffler, M. Serrano, and C. Badie. 2011. 'A minimally invasive assay for individual assessment of the ATM/CHEK2/p53 pathway activity', *Cell Cycle*, 10: 1152-61.

Kao, H. W., D. C. Liang, J. H. Wu, M. C. Kuo, P. N. Wang, C. P. Yang, Y. S. Shih, T. H. Lin, Y. H. Huang, and L. Y. Shih. 2014. 'Gene mutation patterns in patients with minimally differentiated acute myeloid leukemia', *Neoplasia*, 16: 481-8.

Kaplan, D. H., V. Shankaran, A. S. Dighe, E. Stockert, M. Aguet, L. J. Old, and R. D. Schreiber. 1998. 'Demonstration of an interferon gamma-dependent tumor surveillance system in immunocompetent mice', *Proc Natl Acad Sci U S A*, 95: 7556-61.

Kato, Y., A. Iwama, Y. Tadokoro, K. Shimoda, M. Minoguchi, S. Akira, M. Tanaka, A. Miyajima, T. Kitamura, and H. Nakauchi. 2005. 'Selective activation of STAT5 unveils its role in stem cell self-renewal in normal and leukemic hematopoiesis', *J Exp Med*, 202: 169-79.

Kayser, S., K. Dohner, J. Krauter, C. H. Kohne, H. A. Horst, G. Held, M. von Lilienfeld-Toal, S. Wilhelm, A. Kundgen, K. Gotze, M. Rummel, D. Nachbaur, B. Schlegelberger, G. Gohring, D. Spath, C. Morlok, M. Zucknick, A. Ganser, H. Dohner, R. F. Schlenk, and Amlsg German-Austrian. 2011. 'The impact of therapy-related acute myeloid leukemia (AML) on outcome in 2853 adult patients with newly diagnosed AML', *Blood*, 117: 2137-45.

Kelly, L. M., and D. G. Gilliland. 2002. 'Genetics of myeloid leukemias', *Annu Rev Genomics Hum Genet*, 3: 179-98.

Kelly, M. J., J. So, A. J. Rogers, G. Gregory, J. Li, M. Zethoven, M. D. Gearhart, V. J. Bardwell, R. W. Johnstone, S. J. Vervoort, and L. M. Kats. 2019. 'Bcor loss perturbs myeloid differentiation and promotes leukaemogenesis', *Nat Commun*, 10: 1347.

Kern, W., T. Haferlach, S. Schnittger, W. D. Ludwig, W. Hiddemann, and C. Schoch. 2002. 'Karyotype instability between diagnosis and relapse in 117 patients with acute myeloid leukemia: implications for resistance against therapy', *Leukemia*, 16: 2084-91.

Khasawneh, M. K., and O. Abdel-Wahab. 2014. 'Recent discoveries in molecular characterization of acute myeloid leukemia', *Curr Hematol Malig Rep*, 9: 93-9.

Khodarev, N. N., M. Beckett, E. Labay, T. Darga, B. Roizman, and R. R. Weichselbaum. 2004. 'STAT1 is overexpressed in tumors selected for radioresistance and confers protection from radiation in transduced sensitive cells', *Proc Natl Acad Sci U S A*, 101: 1714-9.

Khodarev, N. N., A. J. Minn, E. V. Efimova, T. E. Darga, E. Labay, M. Beckett, H. J. Mauceri, B. Roizman, and R. R. Weichselbaum. 2007. 'Signal transducer and activator of transcription 1 regulates both cytotoxic and prosurvival functions in tumor cells', *Cancer Res*, 67: 9214-20.

Kiel, M. J., O. H. Yilmaz, T. Iwashita, O. H. Yilmaz, C. Terhorst, and S. J. Morrison. 2005. 'SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells', *Cell*, 121: 1109-21.

Kihara, R., Y. Nagata, H. Kiyoi, T. Kato, E. Yamamoto, K. Suzuki, F. Chen, N. Asou, S. Ohtake, S. Miyawaki, Y. Miyazaki, T. Sakura, Y. Ozawa, N. Usui, H. Kanamori, T. Kiguchi, K. Imai, N. Uike, F. Kimura, K. Kitamura, C. Nakaseko, M. Onizuka, A. Takeshita, F. Ishida, H. Suzushima, Y. Kato, H. Miwa, Y. Shiraishi, K. Chiba, H. Tanaka, S. Miyano, S. Ogawa, and T. Naoe. 2014. 'Comprehensive analysis of genetic alterations and their prognostic impacts in adult acute myeloid leukemia patients', *Leukemia*, 28: 1586-95.

Kirito, K., K. Nakajima, T. Watanabe, M. Uchida, M. Tanaka, K. Ozawa, and N. Komatsu. 2002. 'Identification of the human erythropoietin receptor region required for Stat1 and Stat3 activation', *Blood*, 99: 102-10.

Kiyoi, H., M. Towatari, S. Yokota, M. Hamaguchi, R. Ohno, H. Saito, and T. Naoe. 1998. 'Internal tandem duplication of the FLT3 gene is a novel modality of elongation mutation which causes constitutive activation of the product', *Leukemia*, 12: 1333-7.

Kluiver, J., S. Poppema, D. de Jong, T. Blokzijl, G. Harms, S. Jacobs, B. J. Kroesen, and A. van den Berg. 2005. 'BIC and miR-155 are highly expressed in Hodgkin, primary mediastinal and diffuse large B cell lymphomas', *J Pathol*, 207: 243-9.

Ko, M., Y. Huang, A. M. Jankowska, U. J. Pape, M. Tahiliani, H. S. Bandukwala, J. An, E. D. Lamperti, K. P. Koh, R. Ganetzky, X. S. Liu, L. Aravind, S. Agarwal, J. P. Maciejewski, and A. Rao. 2010. 'Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2', *Nature*, 468: 839-43.

Kogan, S. C., J. M. Ward, M. R. Anver, J. J. Berman, C. Brayton, R. D. Cardiff, J. S. Carter, S. de Coronado, J. R. Downing, T. N. Fredrickson, D. C. Haines, A. W. Harris, N. L. Harris, H. Hiai, E. S. Jaffe, I. C. MacLennan, P. P. Pandolfi, P. K. Pattengale, A. S. Perkins, R. M. Simpson, M. S. Tuttle, J. F. Wong, H. C. Morse, 3rd, and Consortium Hematopathology subcommittee of the Mouse Models of Human Cancers. 2002. 'Bethesda proposals for classification of nonlymphoid hematopoietic neoplasms in mice', *Blood*, 100: 238-45.

Koistinen, P., M. Saily, N. Poromaa, and E. R. Savolainen. 1997. 'Complex effects of interleukin 6 on clonogenic blast cell growth in acute myeloblastic leukemia', *Acta Haematol*, 98: 14-21.

Kondo, M., I. L. Weissman, and K. Akashi. 1997. 'Identification of clonogenic common lymphoid progenitors in mouse bone marrow', *Cell*, 91: 661-72.

Kovacic, B., D. Stoiber, R. Moriggl, E. Weisz, R. G. Ott, R. Kreibich, D. E. Levy, H. Beug, M. Freissmuth, and V. Sexl. 2006. 'STAT1 acts as a tumor promoter for leukemia development', *Cancer Cell*, 10: 77-87.

Kovalchuk, O., A. Ponton, J. Filkowski, and I. Kovalchuk. 2004. 'Dissimilar genome response to acute and chronic low-dose radiation in male and female mice', *Mutat Res*, 550: 59-72.

Krainer, A. R., G. C. Conway, and D. Kozak. 1990. 'Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells', *Genes Dev*, 4: 1158-71.

Krestinina, L. Y., F. G. Davis, S. Schonfeld, D. L. Preston, M. Degteva, S. Epifanova, and A. V. Akleyev. 2013. 'Leukaemia incidence in the Techa River Cohort: 1953-2007', *Br J Cancer*, 109: 2886-93.

Kroger, N., R. Brand, A. van Biezen, J. Y. Cahn, S. Slavin, D. Blaise, J. Sierra, A. Zander, D. Niederwieser, T. de Witte, Blood Myelodysplastic Syndromes Subcommittee of The Chronic Leukaemia Working Party of the European Group for, and Transplantation Marrow. 2006. 'Autologous stem cell transplantation for therapy-related acute myeloid leukemia and myelodysplastic syndrome', *Bone Marrow Transplant*, 37: 183-9.

Kuan, C. T., C. J. Wikstrand, and D. D. Bigner. 2001. 'EGF mutant receptor vIII as a molecular target in cancer therapy', *Endocr Relat Cancer*, 8: 83-96.

Kumar, A., M. Commane, T. W. Flickinger, C. M. Horvath, and G. R. Stark. 1997. 'Defective TNF-alpha-induced apoptosis in STAT1-null cells due to low constitutive levels of caspases', *Science*, 278: 1630-2.

Lacronique, Virginie, Anthony Boureux, Véronique Della Valle, Helene Poirel, Christine Tran Quang, Martine Mauchauffé, Christian Berthou, Michel Lessard, Roland Berger, Jacques Ghysdael, and Olivier A. Bernard. 1997. 'A TEL-JAK2 Fusion Protein with Constitutive Kinase Activity in Human Leukemia', *Science*, 278: 1309-12.

Lamandin, C., C. Sagot, C. Roumier, P. Lepelley, S. De Botton, A. Cosson, P. Fenaux, and C. Preudhomme. 2002. 'Are PU.1 mutations frequent genetic events in acute myeloid leukemia (AML)?', *Blood*, 100: 4680-1.

Larson, R. A., Y. Wang, M. Banerjee, J. Wiemels, C. Hartford, M. M. Le Beau, and M. T. Smith. 1999. 'Prevalence of the inactivating 609C-->T polymorphism in the NAD(P)H:quinone oxidoreductase (NQO1) gene in patients with primary and therapy-related myeloid leukemia', *Blood*, 94: 803-7.

Le Deley, M. C., F. Suzan, B. Cutuli, S. Delaloge, A. Shamsaldin, C. Linassier, S. Clisant, F. de Vathaire, P. Fenaux, and C. Hill. 2007. 'Anthracyclines, mitoxantrone, radiotherapy, and granulocyte colony-stimulating factor: risk factors for leukemia and myelodysplastic syndrome after breast cancer', *J Clin Oncol*, 25: 292-300.

Lee, B. H., Z. Tothova, R. L. Levine, K. Anderson, N. Buza-Vidas, D. E. Cullen, E. P. McDowell, J. Adelsperger, S. Frohling, B. J. Huntly, M. Beran, S. E. Jacobsen, and D. G. Gilliland. 2007. 'FLT3 mutations confer enhanced proliferation and survival properties to multipotent progenitors in a murine model of chronic myelomonocytic leukemia', *Cancer Cell*, 12: 367-80.

Lee, H. J., N. Daver, H. M. Kantarjian, S. Verstovsek, and F. Ravandi. 2013. 'The role of JAK pathway dysregulation in the pathogenesis and treatment of acute myeloid leukemia', *Clin Cancer Res*, 19: 327-35.

Lee, M. Y., Y. H. Joung, E. J. Lim, J. H. Park, S. K. Ye, T. Park, Z. Zhang, D. K. Park, K. J. Lee, and Y. M. Yang. 2006. 'Phosphorylation and activation of STAT proteins by hypoxia in breast cancer cells', *Breast*, 15: 187-95.

Lehtonen, A., S. Matikainen, and I. Julkunen. 1997. 'Interferons up-regulate STAT1, STAT2, and IRF family transcription factor gene expression in human peripheral blood mononuclear cells and macrophages', *J Immunol*, 159: 794-803.

Leone, G., L. Mele, A. Pulsoni, F. Equitani, and L. Pagano. 1999. 'The incidence of secondary leukemias', *Haematologica*, 84: 937-45.

Ley, T. J., L. Ding, M. J. Walter, M. D. McLellan, T. Lamprecht, D. E. Larson, C. Kandoth, J. E. Payton, J. Baty, J. Welch, C. C. Harris, C. F. Lichti, R. R. Townsend, R. S. Fulton, D. J. Dooling, D. C. Koboldt, H. Schmidt, Q. Zhang, J. R. Osborne, L. Lin, M. O'Laughlin, J. F. McMichael, K. D. Delehaunty, S. D. McGrath, L. A. Fulton, V. J. Magrini, T. L. Vickery, J. Hundal, L. L. Cook, J. J. Conyers, G. W. Swift, J. P. Reed, P. A. Alldredge, T. Wylie, J. Walker, J. Kalicki, M. A. Watson, S. Heath, W. D. Shannon, N. Varghese, R. Nagarajan, P. Westervelt, M. H. Tomasson, D. C. Link, T. A. Graubert, J. F. DiPersio, E. R. Mardis, and R. K. Wilson. 2010. 'DNMT3A mutations in acute myeloid leukemia', *N Engl J Med*, 363: 2424-33.

Li, L., and L. Ma. 2018. 'Upregulation of miR-582-5p regulates cell proliferation and apoptosis by targeting AKT3 in human endometrial carcinoma', *Saudi J Biol Sci*, 25: 965-70.

Li, Q., K. M. Haigis, A. McDaniel, E. Harding-Theobald, S. C. Kogan, K. Akagi, J. C. Wong, B. S. Braun, L. Wolff, T. Jacks, and K. Shannon. 2011.

'Hematopoiesis and leukemogenesis in mice expressing oncogenic NrasG12D from the endogenous locus', *Blood*, 117: 2022-32.

Li, Y., Y. Okuno, P. Zhang, H. S. Radomska, H. Chen, H. Iwasaki, K. Akashi, M. J. Klemsz, S. R. McKercher, R. A. Maki, and D. G. Tenen. 2001. 'Regulation of the PU.1 gene by distal elements', *Blood*, 98: 2958-65.

Lindsley, R. C., B. G. Mar, E. Mazzola, P. V. Grauman, S. Shareef, S. L. Allen, A. Pigneux, M. Wetzler, R. K. Stuart, H. P. Erba, L. E. Damon, B. L. Powell, N. Lindeman, D. P. Steensma, M. Wadleigh, D. J. DeAngelo, D. Neuberg, R. M. Stone, and B. L. Ebert. 2015. 'Acute myeloid leukemia ontogeny is defined by distinct somatic mutations', *Blood*, 125: 1367-76.

Linet, M. S., S. N. Yin, E. S. Gilbert, G. M. Dores, R. B. Hayes, R. Vermeulen, H. Y. Tian, Q. Lan, L. Portengen, B. T. Ji, G. L. Li, N. Rothman, Control Chinese Center for Disease, and U. S. National Cancer Institute Benzene Study Group Prevention. 2015. 'A retrospective cohort study of cause-specific mortality and incidence of hematopoietic malignancies in Chinese benzene-exposed workers', *Int J Cancer*, 137: 2184-97.

Lobry, C., P. Ntziachristos, D. Ndiaye-Lobry, P. Oh, L. Cimmino, N. Zhu, E. Araldi, W. Hu, J. Freund, O. Abdel-Wahab, S. Ibrahim, D. Skokos, S. A. Armstrong, R. L. Levine, C. Y. Park, and I. Aifantis. 2013. 'Notch pathway activation targets AML-initiating cell homeostasis and differentiation', *J Exp Med*, 210: 301-19.

Loeffler, M., O. Brosteanu, D. Hasenclever, M. Sextro, D. Assouline, A. A. Bartolucci, P. A. Cassileth, D. Crowther, V. Diehl, R. I. Fisher, R. T. Hoppe, P. Jacobs, J. L. Pater, S. Pavlovsky, E. Thompson, and P. Wiernik. 1998. 'Meta-analysis of chemotherapy versus combined modality treatment trials in Hodgkin's disease. International Database on Hodgkin's Disease Overview Study Group', *J Clin Oncol*, 16: 818-29.

Lu, Y., Y. H. Loh, H. Li, M. Cesana, S. B. Ficarro, J. R. Parikh, N. Salomonis, C. X. Toh, S. T. Andreadis, C. J. Luckey, J. J. Collins, G. Q. Daley, and J. A. Marto. 2014. 'Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells', *Cell Stem Cell*, 15: 92-101.

Lyman, S. D., and S. E. Jacobsen. 1998. 'c-kit ligand and Flt3 ligand: stem/progenitor cell factors with overlapping yet distinct activities', *Blood*, 91: 1101-34.

Mahmud, H., S. M. Kornblau, A. Ter Elst, F. J. Scherpen, Y. H. Qiu, K. R. Coombes, and E. S. de Bont. 2016. 'Epidermal growth factor receptor is expressed and active in a subset of acute myeloid leukemia', *J Hematol Oncol*, 9: 64.

Major, I. R. 1979. 'Induction of myeloid leukaemia by whole-body single exposure of CBA male mice to x-rays', *Br J Cancer*, 40: 903-13.

Major, I. R., and R. H. Mole. 1978. 'Myeloid leukaemia in x-ray irradiated CBA mice', *Nature*, 272: 455-6.

Mathews, J. D., A. V. Forsythe, Z. Brady, M. W. Butler, S. K. Goergen, G. B. Byrnes, G. G. Giles, A. B. Wallace, P. R. Anderson, T. A. Guiver, P. McGale, T. M. Cain, J. G. Dowty, A. C. Bickerstaffe, and S. C. Darby. 2013. 'Cancer risk in 680,000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians', *BMJ*, 346: f2360.

Maunakea, A. K., I. Chepelev, K. Cui, and K. Zhao. 2013. 'Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition', *Cell Res*, 23: 1256-69.

Mauritzson, N., M. Albin, L. Rylander, R. Billstrom, T. Ahlgren, Z. Mikoczy, J. Bjork, U. Stromberg, P. G. Nilsson, F. Mitelman, L. Hagmar, and B. Johansson. 2002. 'Pooled analysis of clinical and cytogenetic features in treatment-related and de novo adult acute myeloid leukemia and myelodysplastic syndromes based on a consecutive series of 761 patients analyzed 1976-1993 and on 5098 unselected cases reported in the literature 1974-2001', *Leukemia*, 16: 2366-78.

Mayle, A., M. Luo, M. Jeong, and M. A. Goodell. 2013. 'Flow cytometry analysis of murine hematopoietic stem cells', *Cytometry A*, 83: 27-37.

McCulloch, E. A., and J. E. Till. 1960. 'The radiation sensitivity of normal mouse bone marrow cells, determined by quantitative marrow transplantation into irradiated mice', *Radiat Res*, 13: 115-25.

McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. 2016. 'The Ensembl Variant Effect Predictor', *Genome Biol*, 17: 122.

Medinger, M., and J. R. Passweg. 2017. 'Acute myeloid leukaemia genomics', *Br J Haematol*, 179: 530-42.

Meissl, K., S. Macho-Maschler, M. Muller, and B. Strobl. 2017. 'The good and the bad faces of STAT1 in solid tumours', *Cytokine*, 89: 12-20.

Metzeler, K. H., K. Maharry, J. Kohlschmidt, S. Volinia, K. Mrozek, H. Becker, D. Nicolet, S. P. Whitman, J. H. Mendler, S. Schwind, A. K. Eisfeld, Y. Z. Wu, B. L. Powell, T. H. Carter, M. Wetzler, J. E. Kolitz, M. R. Baer, A. J. Carroll, R. M. Stone, M. A. Caligiuri, G. Marcucci, and C. D. Bloomfield. 2013. 'A stem cell-like gene expression signature associates with inferior outcomes and a distinct microRNA expression profile in adults with primary cytogenetically normal acute myeloid leukemia', *Leukemia*, 27: 2023-31.

Metzeler, K. H., A. Walker, S. Geyer, R. Garzon, R. B. Klisovic, C. D. Bloomfield, W. Blum, and G. Marcucci. 2012. 'DNMT3A mutations and response to the hypomethylating agent decitabine in acute myeloid leukemia', *Leukemia*, 26: 1106-7.

Miousse, I. R., L. Shao, J. Chang, W. Feng, Y. Wang, A. R. Allen, J. Turner, B. Stewart, J. Raber, D. Zhou, and I. Koturbash. 2014. 'Exposure to low-dose (56)Fe-ion radiation induces long-term epigenetic alterations in mouse bone marrow hematopoietic progenitor and stem cells', *Radiat Res*, 182: 92-101.

Moignard, V., I. C. Macaulay, G. Swiers, F. Buettner, J. Schutte, F. J. Calero-Nieto, S. Kinston, A. Joshi, R. Hannah, F. J. Theis, S. E. Jacobsen, M. F. de Bruijn, and B. Gottgens. 2013. 'Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis', *Nat Cell Biol*, 15: 363-72.

Morey, L., C. Brenner, F. Fazi, R. Villa, A. Gutierrez, M. Buschbeck, C. Nervi, S. Minucci, F. Fuks, and L. Di Croce. 2008. 'MBD3, a component of the NuRD complex, facilitates chromatin alteration and deposition of epigenetic marks', *Mol Cell Biol*, 28: 5912-23.

Morrison, S. J., and I. L. Weissman. 1994. 'The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype', *Immunity*, 1: 661-73.

Mrozek, K., N. A. Heerema, and C. D. Bloomfield. 2004. 'Cytogenetics in acute leukemia', *Blood Rev*, 18: 115-36.

Mrozek, K., K. Heinonen, A. de la Chapelle, and C. D. Bloomfield. 1997. 'Clinical significance of cytogenetics in acute myeloid leukemia', *Semin Oncol*, 24: 17-31.

Mueller, B. U., T. Pabst, M. Osato, N. Asou, L. M. Johansen, M. D. Minden, G. Behre, W. Hiddemann, Y. Ito, and D. G. Tenen. 2002. 'Heterozygous PU.1 mutations are associated with acute myeloid leukemia', *Blood*, 100: 998-1007.

Mupo, A., L. Celani, O. Dovey, J. L. Cooper, C. Grove, R. Rad, P. Sportoletti, B. Falini, A. Bradley, and G. S. Vassiliou. 2013. 'A powerful molecular synergy between mutant Nucleophosmin and Flt3-ITD drives acute myeloid leukemia in mice', *Leukemia*, 27: 1917-20.

Nakada, D., H. Oguro, B. P. Levi, N. Ryan, A. Kitano, Y. Saitoh, M. Takeichi, G. R. Wendt, and S. J. Morrison. 2014. 'Oestrogen increases haematopoietic stem-cell self-renewal in females and during pregnancy', *Nature*, 505: 555-8.

Naoe, T., K. Takeyama, T. Yokozawa, H. Kiyoi, M. Seto, N. Uike, T. Ino, A. Utsunomiya, A. Maruta, I. Jin-nai, N. Kamada, Y. Kubota, H. Nakamura, C. Shimazaki, S. Horiike, Y. Kodera, H. Saito, R. Ueda, J. Wiemels, and R. Ohno. 2000. 'Analysis of genetic polymorphism in NQO1, GST-M1, GST-T1, and CYP3A4 in 469 Japanese patients with therapy-related leukemia/myelodysplastic syndrome and de novo acute myeloid leukemia', *Clin Cancer Res*, 6: 4091-5.

Neubauer, A., R. K. Dodge, S. L. George, F. R. Davey, R. T. Silver, C. A. Schiffer, R. J. Mayer, E. D. Ball, D. Wurster-Hill, C. D. Bloomfield, and et al. 1994. 'Prognostic importance of mutations in the ras proto-oncogenes in de novo acute myeloid leukemia', *Blood*, 83: 1603-11.

Neuendorff, N. R., T. Burmeister, B. Dorken, and J. Westermann. 2016. 'BCR-ABL-positive acute myeloid leukemia: a new entity? Analysis of clinical and molecular features', *Ann Hematol*, 95: 1211-21.

Nitulescu, II, S. C. Meyer, Q. J. Wen, J. D. Crispino, M. E. Lemieux, R. L. Levine, H. E. Pelish, and M. D. Shair. 2017. 'Mediator Kinase Phosphorylation of STAT1 S727 Promotes Growth of Neoplasms With JAK-STAT Activation', *EBioMedicine*, 26: 112-25.

O'Connell, R. M., D. S. Rao, A. A. Chaudhuri, M. P. Boldin, K. D. Taganov, J. Nicoll, R. L. Paquette, and D. Baltimore. 2008. 'Sustained expression of microRNA-155 in hematopoietic stem cells causes a myeloproliferative disorder', *J Exp Med*, 205: 585-94.

Oancea, C., B. Ruster, B. Brill, J. Roos, M. Heinssmann, G. Bug, A. A. Mian, N. A. Guillen, S. M. Kornblau, R. Henschler, and M. Ruthardt. 2014. 'STAT activation status differentiates leukemogenic from non-leukemogenic stem cells in AML and is suppressed by arsenic in t(6;9)-positive AML', *Genes Cancer*, 5: 378-92.

Oben, K. Z., B. W. Gachuki, S. S. Alhakeem, M. K. McKenna, Y. Liang, D. K. St Clair, V. M. Rangnekar, and S. Bondada. 2017. 'Radiation Induced

Apoptosis of Murine Bone Marrow Cells Is Independent of Early Growth Response 1 (EGR1)', *PLoS One*, 12: e0169767.

Okada, S., H. Nakauchi, K. Nagayoshi, S. Nishikawa, S. Nishikawa, Y. Miura, and T. Suda. 1991. 'Enrichment and characterization of murine hematopoietic stem cells that express c-kit molecule', *Blood*, 78: 1706-12.

Okano, M., D. W. Bell, D. A. Haber, and E. Li. 1999. 'DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development', *Cell*, 99: 247-57.

Okuda, T., J. van Deursen, S. W. Hiebert, G. Grosveld, and J. R. Downing. 1996. 'AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis', *Cell*, 84: 321-30.

Okuno, Y., G. Huang, F. Rosenbauer, E. K. Evans, H. S. Radomska, H. Iwasaki, K. Akashi, F. Moreau-Gachelin, Y. Li, P. Zhang, B. Gottgens, and D. G. Tenen. 2005. 'Potential autoregulation of transcription factor PU.1 by an upstream regulatory element', *Mol Cell Biol*, 25: 2832-45.

Osato, M. 2004. 'Point mutations in the RUNX1/AML1 gene: another actor in RUNX leukemia', *Oncogene*, 23: 4284-96.

Osato, M., N. Asou, E. Abdalla, K. Hoshino, H. Yamasaki, T. Okubo, H. Suzushima, K. Takatsuki, T. Kanno, K. Shigesada, and Y. Ito. 1999. 'Biallelic and heterozygous point mutations in the runt domain of the AML1/PEBP2alphaB gene associated with myeloblastic leukemias', *Blood*, 93: 1817-24.

Osawa, M., K. Hanada, H. Hamada, and H. Nakauchi. 1996. 'Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell', *Science*, 273: 242-5.

Ozeki, K., H. Kiyoi, Y. Hirose, M. Iwai, M. Ninomiya, Y. Kodera, S. Miyawaki, K. Kuriyama, C. Shimazaki, H. Akiyama, M. Nishimura, T. Motoji, K. Shinagawa, A. Takeshita, R. Ueda, R. Ohno, N. Emi, and T. Naoe. 2004. 'Biologic and clinical significance of the FLT3 transcript level in acute myeloid leukemia', *Blood*, 103: 1901-8.

Pagano, L., A. Pulsoni, M. Vignetti, M. E. Tosti, P. Falcucci, P. Fazi, L. Fianchi, A. Levis, A. Bosi, E. Angelucci, M. Bregni, A. Gabbas, A. Peta, P. Coser, F. Ricciuti, M. Morselli, M. Caira, R. Foa, S. Amadori, F. Mandelli, G. Leone, and Gimema. 2005. 'Secondary acute myeloid leukaemia: results of conventional treatments. Experience of GIMEMA trials', *Ann Oncol*, 16: 228-33.

Palma, C. A., D. Al Sheikha, T. K. Lim, A. Bryant, T. T. Vu, V. Jayaswal, and D. D. Ma. 2014. 'MicroRNA-155 as an inducer of apoptosis and cell differentiation in Acute Myeloid Leukaemia', *Mol Cancer*, 13: 79.

Papaemmanuil, E., M. Gerstung, L. Bullinger, V. I. Gaidzik, P. Paschka, N. D. Roberts, N. E. Potter, M. Heuser, F. Thol, N. Bolli, G. Gundem, P. Van Loo, I. Martincorena, P. Ganly, L. Mudie, S. McLaren, S. O'Meara, K. Raine, D. R. Jones, J. W. Teague, A. P. Butler, M. F. Greaves, A. Ganser, K. Dohner, R. F. Schlenk, H. Dohner, and P. J. Campbell. 2016. 'Genomic Classification and Prognosis in Acute Myeloid Leukemia', *N Engl J Med*, 374: 2209-21.

Park, I. K., D. Qian, M. Kiel, M. W. Becker, M. Pihalja, I. L. Weissman, S. J. Morrison, and M. F. Clarke. 2003. 'Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells', *Nature*, 423: 302-5.

Parkin, B., P. Ouillette, Y. Li, J. Keller, C. Lam, D. Roulston, C. Li, K. Shedden, and S. N. Malek. 2013. 'Clonal evolution and devolution after chemotherapy in adult acute myelogenous leukemia', *Blood*, 121: 369-77.

Pastor, W. A., L. Aravind, and A. Rao. 2013. 'TETonic shift: biological roles of TET proteins in DNA demethylation and transcription', *Nat Rev Mol Cell Biol*, 14: 341-56.

Patel, A., J. Anderson, D. Kraft, R. Finnon, P. Finnon, C. L. Scudamore, G. Manning, R. Bulman, N. Brown, S. Bouffler, P. O'Neill, and C. Badie. 2016. 'The Influence of the CTIP Polymorphism, Q418P, on Homologous Recombination and Predisposition to Radiation-Induced Tumorigenesis (mainly rAML) in Mice', *Radiat Res*, 186: 638-49.

Patel, J. P., M. Gonen, M. E. Figueroa, H. Fernandez, Z. Sun, J. Racevskis, P. Van Vlierberghe, I. Dolgalev, S. Thomas, O. Aminova, K. Huberman, J. Cheng, A. Viale, N. D. Socci, A. Heguy, A. Cherry, G. Vance, R. R. Higgins, R. P. Ketterling, R. E. Gallagher, M. Litzow, M. R. van den Brink, H. M. Lazarus, J. M. Rowe, S. Luger, A. Ferrando, E. Paietta, M. S. Tallman, A. Melnick, O. Abdel-Wahab, and R. L. Levine. 2012. 'Prognostic relevance of integrated genetic profiling in acute myeloid leukemia', *N Engl J Med*, 366: 1079-89.

Pedersen-Bjergaard, J., P. Philip, S. O. Larsen, G. Jensen, and K. Byrsting. 1990. 'Chromosome aberrations and prognostic factors in therapy-related myelodysplasia and acute nonlymphocytic leukemia', *Blood*, 76: 1083-91.

Peng, Y., N. Brown, R. Finnon, C. L. Warner, X. Liu, P. C. Genik, M. A. Callan, F. A. Ray, T. B. Borak, C. Badie, S. D. Bouffler, R. L. Ullrich, J. S. Bedford, and M. M. Weil. 2009. 'Radiation leukemogenesis in mice: loss of PU.1 on chromosome 2 in CBA and C57BL/6 mice after irradiation with 1 GeV/nucleon 56Fe ions, X rays or gamma rays. Part I. Experimental observations', *Radiat Res*, 171: 474-83.

Perou, C. M., S. S. Jeffrey, M. van de Rijn, C. A. Rees, M. B. Eisen, D. T. Ross, A. Pergamenschikov, C. F. Williams, S. X. Zhu, J. C. Lee, D. Lashkari, D. Shalon, P. O. Brown, and D. Botstein. 1999. 'Distinctive gene expression patterns in human mammary epithelial cells and breast cancers', *Proc Natl Acad Sci U S A*, 96: 9212-7.

Pierce, D. A., Y. Shimizu, D. L. Preston, M. Vaeth, and K. Mabuchi. 1996. 'Studies of the mortality of atomic bomb survivors. Report 12, Part I. Cancer: 1950-1990', *Radiat Res*, 146: 1-27.

Pitini, V., C. Arrigo, and G. Altavilla. 2008. 'Erlotinib in a patient with acute myelogenous leukemia and concomitant non-small-cell lung cancer', *J Clin Oncol*, 26: 3645-6.

Pogribny, I., I. Koturbash, V. Tryndyak, D. Hudson, S. M. Stevenson, O. Sedelnikova, W. Bonner, and O. Kovalchuk. 2005. 'Fractionated low-dose radiation exposure leads to accumulation of DNA damage and profound alterations in DNA and histone methylation in the murine thymus', *Mol Cancer Res*, 3: 553-61.

Preston, D. L., S. Kusumi, M. Tomonaga, S. Izumi, E. Ron, A. Kuramoto, N. Kamada, H. Dohy, T. Matsuo, T. Matsui, and et al. 1994. 'Cancer incidence in atomic bomb survivors. Part III. Leukemia, lymphoma and multiple myeloma, 1950-1987', *Radiat Res*, 137: S68-97.

Qian, Z., A. A. Fernald, L. A. Godley, R. A. Larson, and M. M. Le Beau. 2002. 'Expression profiling of CD34+ hematopoietic stem/ progenitor cells

reveals distinct subtypes of therapy-related acute myeloid leukemia', *Proc Natl Acad Sci U S A*, 99: 14925-30.

Ramana, C. V., M. Chatterjee-Kishore, H. Nguyen, and G. R. Stark. 2000. 'Complex roles of Stat1 in regulating gene expression', *Oncogene*, 19: 2619-27.

Rasmussen, K. D., and K. Helin. 2016. 'Role of TET enzymes in DNA methylation, development, and cancer', *Genes Dev*, 30: 733-50.

Rasmussen, K. D., G. Jia, J. V. Johansen, M. T. Pedersen, N. Rapin, F. O. Bagger, B. T. Porse, O. A. Bernard, J. Christensen, and K. Helin. 2015. 'Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis', *Genes Dev*, 29: 910-22.

Reya, T., A. W. Duncan, L. Ailles, J. Domen, D. C. Scherer, K. Willert, L. Hintz, R. Nusse, and I. L. Weissman. 2003. 'A role for Wnt signalling in self-renewal of haematopoietic stem cells', *Nature*, 423: 409-14.

Rithidech, K., V. P. Bond, E. P. Cronkite, M. H. Thompson, and J. E. Bullis. 1995. 'Hypermutability of mouse chromosome 2 during the development of x-ray-induced murine myeloid leukemia', *Proc Natl Acad Sci U S A*, 92: 1152-6.

Rithidech, K. N., J. J. Dunn, C. R. Gordon, E. P. Cronkite, and V. P. Bond. 1996. 'N-ras mutations in radiation-induced murine leukemic cells', *Blood Cells Mol Dis*, 22: 271-80.

Ron, E., D. L. Preston, K. Mabuchi, D. E. Thompson, and M. Soda. 1994. 'Cancer incidence in atomic bomb survivors. Part IV: Comparison of cancer incidence and mortality', *Radiat Res*, 137: S98-112.

Rosenbauer, F., S. Koschmieder, U. Steidl, and D. G. Tenen. 2005. 'Effect of transcription-factor concentrations on leukemic stem cells', *Blood*, 106: 1519-24.

Rosenbauer, F., K. Wagner, J. L. Kutok, H. Iwasaki, M. M. Le Beau, Y. Okuno, K. Akashi, S. Fiering, and D. G. Tenen. 2004. 'Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1', *Nat Genet*, 36: 624-30.

Rothenberg-Thurley, M., S. Amler, D. Goerlich, T. Kohnke, N. P. Konstandin, S. Schneider, M. C. Sauerland, T. Herold, M. Hubmann, B. Ksienzyk, E. Zellmeier, S. K. Bohlander, M. Subklewe, A. Faldum, W. Hiddemann, J. Braess, K. Spiekermann, and K. H. Metzeler. 2018. 'Persistence of pre-leukemic clones during first remission and risk of relapse in acute myeloid leukemia', *Leukemia*, 32: 1598-608.

Rothman, N., M. T. Smith, R. B. Hayes, R. D. Traver, B. Hoener, S. Campleman, G. L. Li, M. Dosemeci, M. Linet, L. Zhang, L. Xi, S. Wacholder, W. Lu, K. B. Meyer, N. Titenko-Holland, J. T. Stewart, S. Yin, and D. Ross. 1997. 'Benzene poisoning, a risk factor for hematological malignancy, is associated with the NQO1 609C-->T mutation and rapid fractional excretion of chlorzoxazone', *Cancer Res*, 57: 2839-42.

Rucker, F. G., R. F. Schlenk, L. Bullinger, S. Kayser, V. Teleanu, H. Kett, M. Habdank, C. M. Kugler, K. Holzmann, V. I. Gaidzik, P. Paschka, G. Held, M. von Lilienfeld-Toal, M. Lubbert, S. Frohling, T. Zenz, J. Krauter, B. Schlegelberger, A. Ganser, P. Lichter, K. Dohner, and H. Dohner. 2012. 'TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype, and dismal outcome', *Blood*, 119: 2114-21.

Saied, M. H., J. Marzec, S. Khalid, P. Smith, T. A. Down, V. K. Rakyan, G. Molloy, M. Raghavan, S. Debernardi, and B. D. Young. 2012. 'Genome wide analysis of acute myeloid leukemia reveal leukemia specific methylome and subtype specific hypomethylation of repeats', *PLoS One*, 7: e33213.

Salemi, D., G. Cammarata, C. Agueli, L. Augugliaro, C. Corrado, M. G. Bica, S. Raimondo, A. Marfia, V. Randazzo, P. Dragotto, F. Di Raimondo, R. Alessandro, F. Fabbiano, and A. Santoro. 2015. 'miR-155 regulative network in FLT3 mutated acute myeloid leukemia', *Leuk Res*, 39: 883-96.

Santillan, D. A., C. M. Theisler, A. S. Ryan, R. Popovic, T. Stuart, M. M. Zhou, S. Alkan, and N. J. Zeleznik-Le. 2006. 'Bromodomain and histone acetyltransferase domain specificities control mixed lineage leukemia phenotype', *Cancer Res*, 66: 10032-9.

Schlenk, R. F., K. Dohner, J. Krauter, S. Frohling, A. Corbacioglu, L. Bullinger, M. Habdank, D. Spath, M. Morgan, A. Benner, B. Schlegelberger, G. Heil, A. Ganser, H. Dohner, and Group German-Austrian Acute Myeloid Leukemia Study. 2008. 'Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia', *N Engl J Med*, 358: 1909-18.

Schnittger, S., F. Dicker, W. Kern, N. Wendland, J. Sundermann, T. Alpermann, C. Haferlach, and T. Haferlach. 2011. 'RUNX1 mutations are frequent in de novo AML with noncomplex karyotype and confer an unfavorable prognosis', *Blood*, 117: 2348-57.

Schoch, C., W. Kern, S. Schnittger, W. Hiddemann, and T. Haferlach. 2004. 'Karyotype is an independent prognostic parameter in therapy-related acute myeloid leukemia (t-AML): an analysis of 93 patients with t-AML in comparison to 1091 patients with de novo AML', *Leukemia*, 18: 120-5.

Schotte, D., J. C. Chau, G. Sylvester, G. Liu, C. Chen, V. H. van der Velden, M. J. Broekhuis, T. C. Peters, R. Pieters, and M. L. den Boer. 2009. 'Identification of new microRNA genes and aberrant microRNA profiles in childhood acute lymphoblastic leukemia', *Leukemia*, 23: 313-22.

Schroeder, T., H. Kohlhof, N. Rieber, and U. Just. 2003. 'Notch signaling induces multilineage myeloid differentiation and up-regulates PU.1 expression', *J Immunol*, 170: 5538-48.

Schuringa, J. J., A. T. Wierenga, W. Kruijer, and E. Vellenga. 2000. 'Constitutive Stat3, Tyr705, and Ser727 phosphorylation in acute myeloid leukemia cells caused by the autocrine secretion of interleukin-6', *Blood*, 95: 3765-70.

Scott, E. W., M. C. Simon, J. Anastasi, and H. Singh. 1994. 'Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages', *Science*, 265: 1573-7.

Shayevitch, R., D. Askayo, I. Keydar, and G. Ast. 2018. 'The importance of DNA methylation of exons on alternative splicing', *RNA*, 24: 1351-62.

Shearstone, J. R., R. Pop, C. Bock, P. Boyle, A. Meissner, and M. Socolovsky. 2011. 'Global DNA demethylation during mouse erythropoiesis in vivo', *Science*, 334: 799-802.

Shen, J. C., W. M. Rideout, 3rd, and P. A. Jones. 1994. 'The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA', *Nucleic Acids Res*, 22: 972-6.

Shen, Y., Y. M. Zhu, X. Fan, J. Y. Shi, Q. R. Wang, X. J. Yan, Z. H. Gu, Y. Y. Wang, B. Chen, C. L. Jiang, H. Yan, F. F. Chen, H. M. Chen, Z. Chen, J. Jin, and

S. J. Chen. 2011. 'Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia', *Blood*, 118: 5593-603.

Shlush, L. I., S. Zandi, A. Mitchell, W. C. Chen, J. M. Brandwein, V. Gupta, J. A. Kennedy, A. D. Schimmer, A. C. Schuh, K. W. Yee, J. L. McLeod, M. Doedens, J. J. Medeiros, R. Marke, H. J. Kim, K. Lee, J. D. McPherson, T. J. Hudson, Halt Pan-Leukemia Gene Panel Consortium, A. M. Brown, F. Yousif, Q. M. Trinh, L. D. Stein, M. D. Minden, J. C. Wang, and J. E. Dick. 2014. 'Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia', *Nature*, 506: 328-33.

Silver, A., J. Moody, R. Dunford, D. Clark, S. Ganz, R. Bulman, S. Bouffler, P. Finnon, E. Meijne, R. Huiskamp, and R. Cox. 1999. 'Molecular mapping of chromosome 2 deletions in murine radiation-induced AML localizes a putative tumor suppressor gene to a 1.0 cM region homologous to human chromosome segment 11p11-12', *Genes Chromosomes Cancer*, 24: 95-104.

Simonnet, A. J., J. Nehme, P. Vaigot, V. Barroca, P. Leboulch, and D. Tronik-Le Roux. 2009. 'Phenotypic and functional changes induced in hematopoietic stem/progenitor cells after gamma-ray radiation exposure', *Stem Cells*, 27: 1400-9.

Small, D. 2006. 'FLT3 mutations: biology and treatment', *Hematology Am Soc Hematol Educ Program*: 178-84.

Smith, L. T., S. Hohaus, D. A. Gonzalez, S. E. Dziennis, and D. G. Tenen. 1996. 'PU.1 (Spi-1) and C/EBP alpha regulate the granulocyte colony-stimulating factor receptor promoter in myeloid cells', *Blood*, 88: 1234-47.

Smith, R. E., J. Bryant, A. DeCillis, S. Anderson, Breast National Surgical Adjuvant, and Experience Bowel Project. 2003. 'Acute myeloid leukemia and myelodysplastic syndrome after doxorubicin-cyclophosphamide adjuvant therapy for operable breast cancer: the National Surgical Adjuvant Breast and Bowel Project Experience', *J Clin Oncol*, 21: 1195-204.

Smith, S. M., M. M. Le Beau, D. Huo, T. Karrison, R. M. Sobecks, J. Anastasi, J. W. Vardiman, J. D. Rowley, and R. A. Larson. 2003. 'Clinical-cytogenetic associations in 306 patients with therapy-related myelodysplasia and myeloid leukemia: the University of Chicago series', *Blood*, 102: 43-52.

Smits, E. L., S. Anguille, and Z. N. Berneman. 2013. 'Interferon alpha may be back on track to treat acute myeloid leukemia', *Oncoimmunology*, 2: e23619.

Snow, J. W., N. Abraham, M. C. Ma, N. W. Abbey, B. Herndier, and M. A. Goldsmith. 2002. 'STAT5 promotes multilineage hematolymphoid development in vivo through effects on early hematopoietic progenitor cells', *Blood*, 99: 95-101.

Sobulo, O. M., J. Borrow, R. Tomek, S. Reshmi, A. Harden, B. Schlegelberger, D. Housman, N. A. Doggett, J. D. Rowley, and N. J. Zeleznik-Le. 1997. 'MLL is fused to CBP, a histone acetyltransferase, in therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3)', *Proc Natl Acad Sci U S A*, 94: 8732-7.

Somerville, T. D., D. H. Wiseman, G. J. Spencer, X. Huang, J. T. Lynch, H. S. Leong, E. L. Williams, E. Cheesman, and T. C. Somervaille. 2015. 'Frequent Derepression of the Mesenchymal Transcription Factor Gene FOXC1 in Acute Myeloid Leukemia', *Cancer Cell*, 28: 329-42.

Song, B., Y. Long, D. Liu, W. Zhang, and C. Liu. 2017. 'MicroRNA-582 promotes tumorigenesis by targeting phosphatase and tensin homologue in colorectal cancer', *Int J Mol Med*, 40: 867-74.

Sonnet, M., R. Claus, N. Becker, M. Zucknick, J. Petersen, D. B. Lipka, C. C. Oakes, M. Andrulis, A. Lier, M. D. Milsom, T. Witte, L. Gu, S. Z. Kim-Wanner, P. Schirmacher, M. Wulfert, N. Gattermann, M. Lubbert, F. Rosenbauer, M. Rehli, L. Bullinger, D. Weichenhan, and C. Plass. 2014. 'Early aberrant DNA methylation events in a mouse model of acute myeloid leukemia', *Genome Med*, 6: 34.

Spangrude, G. J., and D. M. Brooks. 1993. 'Mouse strain variability in the expression of the hematopoietic stem cell antigen Ly-6A/E by bone marrow cells', *Blood*, 82: 3327-32.

Spangrude, G. J., S. Heimfeld, and I. L. Weissman. 1988. 'Purification and characterization of mouse hematopoietic stem cells', *Science*, 241: 58-62.

Spiekermann, K., S. Biethahn, S. Wilde, W. Hiddemann, and F. Alves. 2001. 'Constitutive activation of STAT transcription factors in acute myelogenous leukemia', *Eur J Haematol*, 67: 63-71.

Spiekermann, Karsten, Michael Pau, Ruth Schwab, Karin Schmieja, Sabine Franzrahe, and Wolfgang Hiddemann. 2002. 'Constitutive activation of STAT3 and STAT5 is induced by leukemic fusion proteins with protein tyrosine kinase activity and is sufficient for transformation of hematopoietic precursor cells', *Experimental Hematology*, 30: 262-71.

Staal, F. J., F. Famili, L. Garcia Perez, and K. Pike-Overzet. 2016. 'Aberrant Wnt Signaling in Leukemia', *Cancers (Basel)*, 8.

Staber, P. B., P. Zhang, M. Ye, R. S. Welner, E. Levantini, A. Di Ruscio, A. K. Ebralidze, C. Bach, H. Zhang, J. Zhang, K. Vanura, R. Delwel, H. Yang, G. Huang, and D. G. Tenen. 2014. 'The Runx-PU.1 pathway preserves normal and AML/ETO9a leukemic stem cells', *Blood*, 124: 2391-9.

Staber, P. B., P. Zhang, M. Ye, R. S. Welner, C. Nombela-Arrieta, C. Bach, M. Kerenyi, B. A. Bartholdy, H. Zhang, M. Alberich-Jorda, S. Lee, H. Yang, F. Ng, J. Zhang, M. Leddin, L. E. Silberstein, G. Hoefler, S. H. Orkin, B. Gottgens, F. Rosenbauer, G. Huang, and D. G. Tenen. 2013. 'Sustained PU.1 levels balance cell-cycle regulators to prevent exhaustion of adult hematopoietic stem cells', *Mol Cell*, 49: 934-46.

Stahlberg, A., J. Hakansson, X. Xian, H. Semb, and M. Kubista. 2004. 'Properties of the reverse transcription reaction in mRNA quantification', *Clin Chem*, 50: 509-15.

Stahlberg, A., and M. Kubista. 2018. 'Technical aspects and recommendations for single-cell qPCR', *Mol Aspects Med*, 59: 28-35.

Steidl, U., F. Rosenbauer, R. G. Verhaak, X. Gu, A. Ebralidze, H. H. Otu, S. Klippel, C. Steidl, I. Bruns, D. B. Costa, K. Wagner, M. Aivado, G. Kobbe, P. J. Valk, E. Passegue, T. A. Libermann, R. Delwel, and D. G. Tenen. 2006. 'Essential role of Jun family transcription factors in PU.1 knockdown-induced leukemic stem cells', *Nat Genet*, 38: 1269-77.

Steidl, U., C. Steidl, A. Ebralidze, B. Chapuy, H. J. Han, B. Will, F. Rosenbauer, A. Becker, K. Wagner, S. Koschmieder, S. Kobayashi, D. B. Costa, T. Schulz, K. B. O'Brien, R. G. Verhaak, R. Delwel, D. Haase, L. Trumper, J. Krauter, T. Kohwi-Shigematsu, F. Griesinger, and D. G. Tenen. 2007. 'A distal single nucleotide polymorphism alters long-range regulation of the PU.1 gene in acute myeloid leukemia', *J Clin Invest*, 117: 2611-20.

Stephanou, A., B. K. Brar, R. A. Knight, and D. S. Latchman. 2000. 'Opposing actions of STAT-1 and STAT-3 on the Bcl-2 and Bcl-x promoters', *Cell Death Differ*, 7: 329-30.

Stine, R. R., and E. L. Matunis. 2013. 'JAK-STAT signaling in stem cells', *Adv Exp Med Biol*, 786: 247-67.

Stirewalt, D. L., K. J. Kopecky, S. Meshinchi, F. R. Appelbaum, M. L. Slovak, C. L. Willman, and J. P. Radich. 2001. 'FLT3, RAS, and TP53 mutations in elderly patients with acute myeloid leukemia', *Blood*, 97: 3589-95.

Sun, D., M. Luo, M. Jeong, B. Rodriguez, Z. Xia, R. Hannah, H. Wang, T. Le, K. F. Faull, R. Chen, H. Gu, C. Bock, A. Meissner, B. Gottgens, G. J. Darlington, W. Li, and M. A. Goodell. 2014. 'Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal', *Cell Stem Cell*, 14: 673-88.

Sun, J. Z., Y. Lu, Y. Xu, F. Liu, F. Q. Li, Q. L. Wang, C. T. Wu, X. W. Hu, and H. F. Duan. 2012. 'Epidermal growth factor receptor expression in acute myelogenous leukaemia is associated with clinical prognosis', *Hematol Oncol*, 30: 89-97.

Suraweera, N., E. Meijne, J. Moody, L. G. Carvajal-Carmona, K. Yoshida, P. Pollard, J. Fitzgibbon, A. Riches, T. van Laar, R. Huiskamp, A. Rowan, I. P. Tomlinson, and A. Silver. 2005. 'Mutations of the PU.1 Ets domain are specifically associated with murine radiation-induced, but not human therapy-related, acute myeloid leukaemia', *Oncogene*, 24: 3678-83.

Swerdlow, A. J., A. J. Douglas, G. V. Hudson, B. V. Hudson, M. H. Bennett, and K. A. MacLennan. 1992. 'Risk of second primary cancers after Hodgkin's disease by type of treatment: analysis of 2846 patients in the British National Lymphoma Investigation', *BMJ*, 304: 1137-43.

Tatetsu, H., S. Ueno, H. Hata, Y. Yamada, M. Takeya, H. Mitsuya, D. G. Tenen, and Y. Okuno. 2007. 'Down-regulation of PU.1 by methylation of distal regulatory elements and the promoter is required for myeloma cell growth', *Cancer Res*, 67: 5328-36.

Thiede, C., C. Steudel, B. Mohr, M. Schaich, U. Schakel, U. Platzbecker, M. Wermke, M. Bornhauser, M. Ritter, A. Neubauer, G. Ehninger, and T. Illmer. 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: 4326-35.

Tischkowitz, M. D., N. V. Morgan, D. Grimwade, C. Eddy, S. Ball, I. Vorechovsky, S. Langabeer, R. Stoger, S. V. Hodgson, and C. G. Mathew. 2004. 'Deletion and reduced expression of the Fanconi anemia FANCA gene in sporadic acute myeloid leukemia', *Leukemia*, 18: 420-5.

Townsend, P. A., T. M. Scarabelli, S. M. Davidson, R. A. Knight, D. S. Latchman, and A. Stephanou. 2004. 'STAT-1 interacts with p53 to enhance DNA damage-induced apoptosis', *J Biol Chem*, 279: 5811-20.

Turkistany, S. A., and R. P. DeKoter. 2011. 'The transcription factor PU.1 is a critical regulator of cellular communication in the immune system', *Arch Immunol Ther Exp (Warsz)*, 59: 431-40.

Tyner, J. W., M. M. Loriaux, H. Erickson, C. A. Eide, J. Deininger, M. MacPartlin, S. G. Willis, T. Lange, B. J. Druker, T. Kovacsovics, R. Maziarz, N. Gattermann, and M. W. Deininger. 2009. 'High-throughput mutational screen of the tyrosine kinome in chronic myelomonocytic leukemia', *Leukemia*, 23: 406-9.

Upton, A. C., F. F. Wolff, J. Furth, and A. W. Kimball. 1958. 'A comparison of the induction of myeloid and lymphoid leukemias in x-radiated RF mice', *Cancer Res*, 18: 842-8.

Varnum-Finney, B., L. E. Purton, M. Yu, C. Brashem-Stein, D. Flowers, S. Staats, K. A. Moore, I. Le Roux, R. Mann, G. Gray, S. Artavanis-Tsakonas, and I. D. Bernstein. 1998. 'The Notch ligand, Jagged-1, influences the development of primitive hematopoietic precursor cells', *Blood*, 91: 4084-91.

Vassiliou, G. S., J. L. Cooper, R. Rad, J. Li, S. Rice, A. Uren, L. Rad, P. Ellis, R. Andrews, R. Banerjee, C. Grove, W. Wang, P. Liu, P. Wright, M. Arends, and A. Bradley. 2011. 'Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice', *Nat Genet*, 43: 470-5.

Vazquez, S. E., M. A. Inlay, and T. Serwold. 2015. 'CD201 and CD27 identify hematopoietic stem and progenitor cells across multiple murine strains independently of Kit and Sca-1', *Exp Hematol*, 43: 578-85.

Vegesna, V., S. Takeuchi, W. K. Hofmann, T. Ikezoe, S. Tavor, U. Krug, A. C. Fermin, A. Heaney, C. W. Miller, and H. P. Koeffler. 2002. 'C/EBP-beta, C/EBP-delta, PU.1, AML1 genes: mutational analysis in 381 samples of hematopoietic and solid malignancies', *Leuk Res*, 26: 451-7.

Verbiest, T., S. Bouffler, S. L. Nutt, and C. Badie. 2015. 'PU.1 downregulation in murine radiation-induced acute myeloid leukaemia (AML): from molecular mechanism to human AML', *Carcinogenesis*, 36: 413-9.

Verbiest, T., R. Finnon, N. Brown, L. Cruz-Garcia, P. Finnon, G. O'Brien, E. Ross, S. Bouffler, C. L. Scudamore, and C. Badie. 2018. 'Tracking preleukemic cells in vivo to reveal the sequence of molecular events in radiation leukemogenesis', *Leukemia*, 32: 1435-44.

Vigorito, E., K. L. Perks, C. Abreu-Goodger, S. Bunting, Z. Xiang, S. Kohlhaas, P. P. Das, E. A. Miska, A. Rodriguez, A. Bradley, K. G. Smith, C. Rada, A. J. Enright, K. M. Toellner, I. C. Maclennan, and M. Turner. 2007. 'microRNA-155 regulates the generation of immunoglobulin class-switched plasma cells', *Immunity*, 27: 847-59.

Wagers, A. J., R. C. Allsopp, and I. L. Weissman. 2002. 'Changes in integrin expression are associated with altered homing properties of Lin(-/lo)Thy1.1(lo)Sca-1(+)c-kit(+) hematopoietic stem cells following mobilization by cyclophosphamide/granulocyte colony-stimulating factor', *Exp Hematol*, 30: 176-85.

Wagers, A. J., and I. L. Weissman. 2006. 'Differential expression of alpha2 integrin separates long-term and short-term reconstituting Lin-/loThy1.1(lo)c-kit+ Sca-1+ hematopoietic stem cells', *Stem Cells*, 24: 1087-94.

Wajed, S. A., P. W. Laird, and T. R. DeMeester. 2001. 'DNA methylation: an alternative pathway to cancer', *Ann Surg*, 234: 10-20.

Walters, D. K., T. Mercher, T. L. Gu, T. O'Hare, J. W. Tyner, M. Loriaux, V. L. Goss, K. A. Lee, C. A. Eide, M. J. Wong, E. P. Stoffregen, L. McGreevey, J. Nardone, S. A. Moore, J. Crispino, T. J. Boggon, M. C. Heinrich, M. W. Deininger, R. D. Polakiewicz, D. G. Gilliland, and B. J. Druker. 2006. 'Activating alleles of JAK3 in acute megakaryoblastic leukemia', *Cancer Cell*, 10: 65-75.

Wang, Y., A. V. Krivtsov, A. U. Sinha, T. E. North, W. Goessling, Z. Feng, L. I. Zon, and S. A. Armstrong. 2010. 'The Wnt/beta-catenin pathway is required for the development of leukemia stem cells in AML', *Science*, 327: 1650-3.

Weber-Nordt, R. M., C. Egen, J. Wehinger, W. Ludwig, V. Gouilleux-Gruart, R. Mertelsmann, and J. Finke. 1996. 'Constitutive activation of STAT proteins in primary lymphoid and myeloid leukemia cells and in Epstein-Barr virus (EBV)-related lymphoma cell lines', *Blood*, 88: 809-16.

Weiss, H. A., S. C. Darby, and R. Doll. 1994. 'Cancer mortality following X-ray treatment for ankylosing spondylitis', *Int J Cancer*, 59: 327-38.

Weissmann, S., T. Alpermann, V. Grossmann, A. Kowarsch, N. Nadarajah, C. Eder, F. Dicker, A. Fasan, C. Haferlach, T. Haferlach, W. Kern, S. Schnittger, and A. Kohlmann. 2012. 'Landscape of TET2 mutations in acute myeloid leukemia', *Leukemia*, 26: 934-42.

Welch, J. S., T. J. Ley, D. C. Link, C. A. Miller, D. E. Larson, D. C. Koboldt, L. D. Wartman, T. L. Lamprecht, F. Liu, J. Xia, C. Kandoth, R. S. Fulton, M. D. McLellan, D. J. Dooling, J. W. Wallis, K. Chen, C. C. Harris, H. K. Schmidt, J. M. Kalicki-Veizer, C. Lu, Q. Zhang, L. Lin, M. D. O'Laughlin, J. F. McMichael, K. D. Delehaunty, L. A. Fulton, V. J. Magrini, S. D. McGrath, R. T. Demeter, T. L. Vickery, J. Hundal, L. L. Cook, G. W. Swift, J. P. Reed, P. A. Alldredge, T. N. Wylie, J. R. Walker, M. A. Watson, S. E. Heath, W. D. Shannon, N. Varghese, R. Nagarajan, J. E. Payton, J. D. Baty, S. Kulkarni, J. M. Klco, M. H. Tomasson, P. Westervelt, M. J. Walter, T. A. Graubert, J. F. DiPersio, L. Ding, E. R. Mardis, and R. K. Wilson. 2012. 'The origin and evolution of mutations in acute myeloid leukemia', *Cell*, 150: 264-78.

Wellcome Trust Case Control, Consortium. 2007. 'Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls', *Nature*, 447: 661-78.

Will, B., T. O. Vogler, S. Narayanagari, B. Bartholdy, T. I. Todorova, M. da Silva Ferreira, J. Chen, Y. Yu, J. Mayer, L. Barreyro, L. Carvajal, D. B. Neriah, M. Roth, J. van Oers, S. Schaetzlein, C. McMahon, W. Edelmann, A. Verma, and U. Steidl. 2015. 'Minimal PU.1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia', *Nat Med*, 21: 1172-81.

Wilson, A., E. Laurenti, G. Oser, R. C. van der Wath, W. Blanco-Bose, M. Jaworski, S. Offner, C. F. Dunant, L. Eshkind, E. Bockamp, P. Lio, H. R. Macdonald, and A. Trumpp. 2008. 'Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair', *Cell*, 135: 1118-29.

Wilson, A., G. M. Oser, M. Jaworski, W. E. Blanco-Bose, E. Laurenti, C. Adolphe, M. A. Essers, H. R. Macdonald, and A. Trumpp. 2007. 'Dormant and self-renewing hematopoietic stem cells and their niches', *Ann N Y Acad Sci*, 1106: 64-75.

Wong, T. N., G. Ramsingh, A. L. Young, C. A. Miller, W. Touma, J. S. Welch, T. L. Lamprecht, D. Shen, J. Hundal, R. S. Fulton, S. Heath, J. D. Baty, J. M. Klco, L. Ding, E. R. Mardis, P. Westervelt, J. F. DiPersio, M. J. Walter, T. A. Graubert, T. J. Ley, T. Druley, D. C. Link, and R. K. Wilson. 2015. 'Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia', *Nature*, 518: 552-55.

Xie, M., C. Lu, J. Wang, M. D. McLellan, K. J. Johnson, M. C. Wendl, J. F. McMichael, H. K. Schmidt, V. Yellapantula, C. A. Miller, B. A. Ozenberger,

J. S. Welch, D. C. Link, M. J. Walter, E. R. Mardis, J. F. Dipersio, F. Chen, R. K. Wilson, T. J. Ley, and L. Ding. 2014. 'Age-related mutations associated with clonal hematopoietic expansion and malignancies', *Nat Med*, 20: 1472-8.

Yamamoto, Y., H. Kiyoi, Y. Nakano, R. Suzuki, Y. Kodera, S. Miyawaki, N. Asou, K. Kuriyama, F. Yagasaki, C. Shimazaki, H. Akiyama, K. Saito, M. Nishimura, T. Motoji, K. Shinagawa, A. Takeshita, H. Saito, R. Ueda, R. Ohno, and T. Naoe. 2001. 'Activating mutation of D835 within the activation loop of FLT3 in human hematologic malignancies', *Blood*, 97: 2434-9.

Yamashita, Y., J. Yuan, I. Suetake, H. Suzuki, Y. Ishikawa, Y. L. Choi, T. Ueno, M. Soda, T. Hamada, H. Haruta, S. Takada, Y. Miyazaki, H. Kiyoi, E. Ito, T. Naoe, M. Tomonaga, M. Toyota, S. Tajima, A. Iwama, and H. Mano. 2010. 'Array-based genomic resequencing of human leukemia', *Oncogene*, 29: 3723-31.

Yan, X. J., J. Xu, Z. H. Gu, C. M. Pan, G. Lu, Y. Shen, J. Y. Shi, Y. M. Zhu, L. Tang, X. W. Zhang, W. X. Liang, J. Q. Mi, H. D. Song, K. Q. Li, Z. Chen, and S. J. Chen. 2011. 'Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia', *Nat Genet*, 43: 309-15.

Ye, F., W. Huang, and G. Guo. 2017. 'Studying hematopoiesis using single-cell technologies', *J Hematol Oncol*, 10: 27.

Yi, G., A. T. J. Wierenga, F. Petraglia, P. Narang, E. M. Janssen-Megens, A. Mandoli, A. Merkel, K. Berentsen, B. Kim, F. Matarese, A. A. Singh, E. Habibi, K. H. M. Prange, A. B. Mulder, J. H. Jansen, L. Clarke, S. Heath, B. A. van der Reijden, P. Flicek, M. L. Yaspo, I. Gut, C. Bock, J. J. Schuringa, L. Altucci, E. Vellenga, H. G. Stunnenberg, and J. H. A. Martens. 2019. 'Chromatin-Based Classification of Genetically Heterogeneous AMLs into Two Distinct Subtypes with Diverse Stemness Phenotypes', *Cell Rep*, 26: 1059-69 e6.

Yoshida, K., M. Sanada, Y. Shiraishi, D. Nowak, Y. Nagata, R. Yamamoto, Y. Sato, A. Sato-Otsubo, A. Kon, M. Nagasaki, G. Chalkidis, Y. Suzuki, M. Shiosaka, R. Kawahata, T. Yamaguchi, M. Otsu, N. Obara, M. Sakata-Yanagimoto, K. Ishiyama, H. Mori, F. Nolte, W. K. Hofmann, S. Miyawaki, S. Sugano, C. Haferlach, H. P. Koeffler, L. Y. Shih, T. Haferlach, S. Chiba, H. Nakauchi, S. Miyano, and S. Ogawa. 2011. 'Frequent pathway mutations of splicing machinery in myelodysplasia', *Nature*, 478: 64-9.

Yoshinaga, S., K. Mabuchi, A. J. Sigurdson, M. M. Doody, and E. Ron. 2004. 'Cancer risks among radiologists and radiologic technologists: review of epidemiologic studies', *Radiology*, 233: 313-21.

Yuan, Y., L. Zhou, T. Miyamoto, H. Iwasaki, N. Harakawa, C. J. Hetherington, S. A. Burel, E. Lagasse, I. L. Weissman, K. Akashi, and D. E. Zhang. 2001. 'AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations', *Proc Natl Acad Sci U S A*, 98: 10398-403.

Yuasa, K., and T. Hijikata. 2016. 'Distal regulatory element of the STAT1 gene potentially mediates positive feedback control of STAT1 expression', *Genes Cells*, 21: 25-40.

Yuki, H., S. Ueno, H. Tatetsu, H. Niiro, T. Iino, S. Endo, Y. Kawano, Y. Komohara, M. Takeya, H. Hata, S. Okada, T. Watanabe, K. Akashi, H. Mitsuya, and Y.

Okuno. 2013. 'PU.1 is a potent tumor suppressor in classical Hodgkin lymphoma cells', *Blood*, 121: 962-70.

Zhang, D. E., C. J. Hetherington, H. M. Chen, and D. G. Tenen. 1994. 'The macrophage transcription factor PU.1 directs tissue-specific expression of the macrophage colony-stimulating factor receptor', *Mol Cell Biol*, 14: 373-81.

Zhang, H., X. Q. Luo, P. Zhang, L. B. Huang, Y. S. Zheng, J. Wu, H. Zhou, L. H. Qu, L. Xu, and Y. Q. Chen. 2009. 'MicroRNA patterns associated with clinical prognostic parameters and CNS relapse prediction in pediatric acute leukemia', *PLoS One*, 4: e7826.

Zhong, J. F., Y. Zhao, S. Sutton, A. Su, Y. Zhan, L. Zhu, C. Yan, T. Gallaher, P. B. Johnston, W. F. Anderson, and M. P. Cooke. 2005. 'Gene expression profile of murine long-term reconstituting vs. short-term reconstituting hematopoietic stem cells', *Proc Natl Acad Sci U S A*, 102: 2448-53.

Zhou, J., J. Wu, B. Li, D. Liu, J. Yu, X. Yan, S. Zheng, J. Wang, L. Zhang, L. Zhang, F. He, Q. Li, A. Chen, Y. Zhang, X. Zhao, Y. Guan, X. Zhao, J. Yan, J. Ni, M. A. Nobrega, B. Lowenberg, R. Delwel, P. J. Valk, A. Kumar, L. Xie, D. G. Tenen, G. Huang, and Q. F. Wang. 2014. 'PU.1 is essential for MLL leukemia partially via crosstalk with the MEIS/HOX pathway', *Leukemia*, 28: 1436-48.

Appendices

Appendix A
Primers sequences for human DNA sanger sequencing

Gene	Name	Sequence 5'->3'
SETBP1	F primer	GAGTAGCGCAGACAAAGAG
	R primer	TACGTTCCAAAGCCTTCATAG
TET2	F primer	GCGAGTTCGAGACTCATAAT
	R primer	ACTTCTGCTCCTGTTCTTG
SRSF2	F primer	GTGGACAACCTGACCTAC
	R primer	GCATCCATAGCGTCCTC
KRAS	F primer	CGATGGAGGAGTTTGTAAATGAA
	R primer	TTCGTCCACAAAATGATTCTGA
EZH2	F primer	TGGACCAAATGCTAAATCTGTTCA
	R primer	CACTTACGATGTAGGAAGCAGTCA
DNMT3A	F primer	GACTTCTTTGCCAAGTTCAC
	R primer	TATCCAAGGAGGAAGCCTAT
RUNX1	F primer	GGTATAGCATCCTGGGTAATC
	R primer	CGTAGTACAGGTGGTAGGA
PU.1 DRU(Bonadies et al.2010)	F primer	AGAGGAAACTGAGGCCAAGTG
	R primer	TGGCAGTCCTCACTGAGGCCATTG
PU.1 PRU(Bonadies et al. 2010)	F primer	CAATGGCCTCAGTGAGGACTGCCA
	R primer	TCTTGGCGGAAGCTGTTAGGGAAG

Appendix B

Primers sequences for mouse DNA sequencing

Gene	Name	Sequence 5'->3'
Npm1	F primer	TAAATAGGGCTGACCCACAG
	R primer	ACCAAGTAAAGGGTGGAGTT
Sfpi1	F primer	GTGGACAAGGACAAAGGTA
	R primer	CCATAGCATTAACCCGTCG
Flt3 (Finnon et al. 2012)	F primer	GCAATTTAGGTACGAGAGTCAGC
	R primer	CTTTTAGCATCTTCACCGCCACC
Idh1	F primer	CTGTCTTCAGGGAAGCTATTAT
	R primer	GGAGCTAAAGGTCTGTGAAA
Idh2	F primer	TGGTGGGTCTATTGTACCT
	R primer	AATCTGTGGCCTTGTACTG
Dnmt3a	F primer	AACTAACATCCGCCATCAC
	R primer	TCATCACTACTTCAGTTTGCC
Kras	F primer	CAACAAAGAATACCGCAAGGGT
	R primer	AGAGCAGCGTTACCTCTATCG
Nras	F primer	GCAAGGAATGCTATGTTTCTG
	R primer	CCTCTATGGTGGGATCATATTC
Cebpa	F primer	CTATAGACATCAGCGCCTAC
	R primer	GCTCTTGTTTGATCACCAG
C Kit	F primer	TCGGAGAGCTGAAATGAATG
	R primer	GTGACATTACAAGGTAGGAGTT

Appendix C

Primers sequences for human DNA sequencing by pyrosequencing

Gene	Name	Sequence 5'->3'
ASXL1	F primer	ATCATCCCCACCACGGAGT
	R primer	TGGTGGCCGCCTCTCTATG
	Pyrosequencing primer	AAAGCCCGTGCTCTG
SRSF2	F primer	ACCGCTACACCAAGGAGTCCC
	R primer	GGCGGCTGTGGTGTGAGT
	Pyrosequencing primer	CGCTACGGCCGCCCC
RUNX1	F primer	TGTGATGGCTGGCAATGA
	R primer	CGACAAACCTGAGGTCATTAAAT
	Pyrosequencing primer	GGCTGCGGTAGCATTT

Appendix D

Primers sequences for mouse DNA sequencing by pyrosequencing

Gene	Name	Sequence 5'->3'
Sfpi1	F primer	GCATCCAGAAGGGCAACC
	R primer	TCGCCTGTCTTGCCGTAGT
	Pyrosequencing primer	CCTGTCTTGCCGTAGT
Kras	F primer	GTAAGGCCTGCTGAAAATGACTGA
	R primer	TATCGTCAAGGCGCTCTTGC
	Pyrosequencing primer	CGCTCTTGCCTACGC

Appendix E

Primers sequences for human DNA methylation by pyrosequencing

Gene	Name	Sequence 5'->3'
PU.1 Promoter	F primer	GGTTTTTGTAGTTTAGGGGGTAG
	R primer	ACCCTAACTTCCCACTAATAAC
	Pyrosequencing primer	GTAGTTTAGGGGGTAGGT
PU.1 URE	F primer	GGGATTGAGTTGAGAGTTTAGAAGAAG
	R primer	AACTACAACTACCCCTATTTCCA
	Pyrosequencing primer	ATTTTTTTGTAGGTTTGGTTTA

Appendix F
Primers sequences for mouse DNA methylation by pyrosequencing

Gene	Name	Sequence 5'->3'
Sfpi1 Promoter	F primer	AGGGTTTATAGGAAGAGTTAAGT
	R primer	AAATAATCCACTATTCTTTTAACCTAA
	Pyrosequencing primer	AAATTTATTTTTAAAATTAGGGA
Sfpi1 DRU Upper	F primer	TGGGTGTTTTAGGTTGTTGTTT
	R primer	ACCTAAAAAAACCTATATTCCTTCAAC
	Pyrosequencing primer	TTGTTGTTTGGTAGGT
Sfpi1 DRU Lower	F primer	AAGGTAGGGTATGGGGATTAG
	R primer	ACCTACTTTACCCTCTATCCA
	Pyrosequencing primer	GGTATGGGGATTAGG
Sfpi1 PRU	F primer	TGGGGAGGTAGAGTATATATGTTTTT
	R primer	CACTCCCTTCTAAACAAAATCAAAAT
	Pyrosequencing primer	AGTATATATGTTTTTTGTGGTG

Appendix G

Primers sequences for human MQRT-PCR

Gene	Name	Sequence 5'->3'
HPRT (F Pallier)	F primer	GGACAGGACTGAAAGACTTG
	R primer	TAATCCAGCAGGTCAGCAAA
	Probe	CCCTTGAGCACACAGAGGGCCACA
PU.1	F primer	CCCTATGACACGGATCTATAC
	R primer	CCCAGTAATGGTCGCTATG
	Probe	AACGCCAAACGCACGAGTATTACC

Appendix H

Primers and probe sequences for mouse MQRT-QPCR

Gene	Name	Sequence 5'->3'
Hprt (F Pallier)	F primer	GGACAGGACTGAAAGACTTG
	R primer	TAATCCAGCAGGTCAGCAAA
	Probe	CCCTTGAGCACACAGAGGGCCACA
Sfpi1	F primer	AGAAGCTGATGGCTTGGAGC
	R primer	GCGAATCTTTTTCTTGCTGCC
	Probe	TGGGCCAGGTCTTCTGCACGG
Flt3	F primer	GCAGCTACTTTGAGATGAGTA
	R primer	CAGGTGTAATATCCGGTGTC
		CCTACTCCACAAACAGGACCATG
	Probe	A
Hoxb5	F primer	CTCGAGCACAGCCAGAG
	R primer	CTGGCCCAGTCATATCGT
	Probe	CCTGGATGAGGAAGCTTCACATCA
Sesn2 (F Pallier)	F primer	CGTTTTGAGCTGGAGAAGTCA
	R primer	GTGGAGAAGGCTCCAGGATA
	Probe	AGCCTGCTGGTGACCCCCTCAGC
Stat1	F primer	CGAGAACATACCAGAGAATCC
	R primer	GGTTCTGGTGCTTCCTTT
	Probe	ATAATACTTCCCAAAGGCGTGGTC
Jak3	F primer	TGATGGGAATCCACCTTTC
	R primer	TCTGTCGGTGAGCATTTC
	Probe	ATTAAGCTGAGTGATCCTGGTGTC