The Role of WNT Transcription Factor TCF7L2 in Acute Myeloid Leukaemia

by

Siti Sarah Daud

2014

A thesis presented to Cardiff University in partial fulfilment of the requirement for the degree of Doctor of Philosophy

Department of Haematology, Institute of Cancer and Genetics,
School of Medicine, Cardiff University,
CF14 4XN Cardiff, United Kingdom

Supervisors: Dr. Alex Tonks and Dr. Richard Darley

Acknowledgements

I would like to express my gratitude and greatest appreciation to my supervisors; Dr. Alex Tonks and Dr. Richard Darley for their generous guidance and support throughout this study.

This study has been fully supported by the National Science Fellowship (NSF) from Malaysia. I am very thankful and deeply grateful to Ministry of Science, Technology and Innovation Malaysia (MOSTI) for the financial support they have provided.

I would like to thank all staffs of the Department of Haematology; it has been a real pleasure to work with them. Importantly;

Professor Alan Burnett for allowing this research to be conducted within Department of Haematology.

Amanda Gilkes for her contribution to RNA samples from primary AML cohort and for lending her experience in microarray processing.

Sara Pumford for her kind help during the construction of retroviruses constructs.

Lorna Pearn and Dr. Marie Gilmour for their endless technical laboratory support.

Dr. Nader Omidvar, Dr. Rhys Morgan and Dr. Paul Hole for their scientific insights.

Dr. Paul White and the NCRI-MRC AML trials team for providing the primary AML samples.

Dr. Robert Hills for his assistance with statistical analyses.

Dr. Derrick Bowen for his guidance with the exon-splicing experiment.

Dr. Anna Evans for sharing her bioinformatics expertise.

I am especially indebted to my truly beloved husband, Mr. Jaime Castillo Garcia, who has constantly encouraged this path. It is again to him that this Ph.D is dedicated.

Last but not least, I wish to thank my beloved parents; Mr. Hj. Daud and Mrs. Hjh. Noor Azizan, and my family who has been the fundamental base to every success I have achieved in life.

Foreword

Full supplementary materials are provided as electronic documents on the accompanying CD (see back sleeve).

The accompanying CD also contains an electronic version of this thesis.

Especially dedicated to my beloved husband,

Jaime Castillo Garcia.

Presentations

Conference proceedings:

Identification of the Wnt Signalling Protein, TCF7L2 as a Significantly Overexpressed Transcription Factor in AML. **Daud SS**, Pumford SL, Gilmour MN, Gilkes AF, Burnett AK, Hills R, Tonks A, Darley RL. *Blood*, Volume 120, (November 2012) pp.1281 (see below *)

Large-Scale Integration of Gene Profiling Identifies TCF7L2/TCF4 as the Most Frequently Dysregulated Wnt Signalling Component in AML. **Daud SS**, Burnett AK, Darley RL, Tonks A. *Blood*, Volume 116, (November 2010) pp.1029 (see below **)

Poster presentations:

* November 2012: American Society of Hematology Abstract Achievement Award for 54th Annual Meeting poster presentation, Atlanta, Georgia, U.S.A. Presenter and award recipient.

**December 2010: American Society of Hematology Travel Award for 52nd Annual Meeting poster presentation, Orlando, Florida, U.S.A. Presenter and award recipient.

Oral presentations:

November 2011: Identification of novel splice variants of TCF7L2, lacking the Groucho repressor domain in Acute Myeloid Leukaemia patients. 26th Annual Postgraduate Research Day, Graduate School of Biomedical and Life Sciences, Cardiff University, U.K.

October 2011: Interview with Feinstein Kean Healthcare, New York; for The cancer Biomedical Informatics Grid (caBIG) Milestones newsletter a story on "Identification of Dysregulated Networks in Acute Myeloid Leukemia Using caBIG® Distance Weighted Discrimination".

September 2011: Aberrant gene expression in Acute Myeloid Leukaemia : Dysregulation of T-cell factor 4 (TCF4/TCF7L2). Cancer IRG Ph.D Seminar Day; Medical Genetics. Henry Wellcome Building, Cardiff University, U.K.

May 2011: The flow of "Wnt" signals in cancer cells: from Probes to Proteins. Speaking of Science Conference, Graduate Centre,Park Place, Cardiff University, U.K.

June 2010: Role of Wnt Signaling in Acute Myeloid Leukemia. Cancer IRG Ph.D Seminar Day; Medical Genetics. Henry Wellcome Building, Cardiff University, U.K.

Abstract

Acute myeloid leukaemia (AML) is a heterogeneous clonal disorder that affects the development of haematopoietic cells and has an incidence of 2300 cases / year in the UK. Whilst treatment with conventional cytotoxic agents has significantly increased cure rates in AML, only 35-50% patients under the age of 60 years will be long term survivors. There is a need to develop novel agents targeting molecular lesions or dysregulated pathways to improve clinical responses. This study identified commonly dysregulated pathways and genes using Affymetrix gene expression profiling of AML patients. Analysis of gene expression data using Metacore^{TM} online gene ontology pathway analysis identified WNT proteins to be one of the most frequently dysregulated processes associated with AML. *TCF7L2* was the most significantly dysregulated WNT transcription factor gene and was subsequently examined in greater detail. The expression of *TCF7L2* was validated by RQ-PCR. These data significantly correlated with the normalised array expression data ($R=0.748$; $P<0.01$) which in turn correlated with overall protein expression determined and quantified by western blotting. *TCF7L2* mRNA overexpression was found to be independently prognostic for reduced complete remission rate ($P<0.05$), OR=5.19 [95% C.I.=1.39 - 19.39]. The *TCF7L2* gene undergoes exon splicing which yields multiple isoforms which have been reported to yield functionally distinct proteins. *TCF7L2* mRNA isoform expression in AML patients was compared with that in normal human bone marrow and a human cord blood derived from $CD34^+$ haematopoeitic progenitor cells. Extensive variability of *TCF7L2* splicing at the 3' end of the gene (exons 13-18) was identified. AML patients were heterogeneous in the isoform expression pattern, but aberrant exon composition (compared to normal cells) was not observed. At the protein level, expression of the 58 kDa isoform (exons 1-14 and 18) and 56 kDa isoform (exons 1-13 and 18) were detected. In both normal and AML cells, expression of the 56 kDa and 58 kDa isoforms were dominant. To determine the functional significance of TCF7L2 overexpression lentiviral shRNA vectors targeting TCF7L2 was transduced in leukaemic cell lines co-expressing a TCF-GFP reporter which enabled simultaneous analysis of the effect on TCF-dependent transcription. Reporter activity was inhibited by shRNA vectors targeting TCF7L2, and the cells became non-responsive to WNT agonists (WNT3A and BIO) demonstrating that canonical WNT signalling is dependent on TCF7L2 expression in these cells. Phenotypically, shRNA expression was found to strongly inhibit proliferation and to promote apoptosis indicating that TCF7L2 expression is required for the proliferation and survival of myeloid leukaemia cells. Paradoxically, overexpression of individual TCF7L2 isoforms (72 kDa, 58 kDa and 56 kDa) suppressed WNT agonist responses in myeloid leukaemia cell lines but not in epithelial cells. Overexpression of TCF7L2 in normal $CD34^+$ cells was found to promote monocytic differentiation. In summary, this study presents novel data identifying WNT signalling as the most commonly dysregulated process in AML. *TCF7L2*, a WNT transcription factor was found to be significantly overexpressed in AML and associated with poor clinical outcome. This protein was found to be required to maintain the proliferation and viability of myeloid leukaemia cells suggesting that targeting TCF7L2 maybe a valid approach in AML therapy.

List of Abbreviations

A	adenine
ABL	Abelson
ADAPT	A Database of Affymetrix Probesets and Transcripts
ALL	acute lymphoblastic leukaemia
AML	acute myeloid leukaemia
ANOVA	analysis of variance
APL	acute promyelocytic leukaemia
APC	allophycocyanin
Ara-C	cytosine arabinoside
ATRA	all-trans-retinoic acid
BIO	(2'Z,3'E)-6-Bromoindirubin-3'-oxime
BLAST	Basic Local Alignment Search Tool
BM	bone marrow
bp	base pair
C	cytosine
C-terminus	carboxyl (COOH)-terminus
C-clamp	cysteines-clamp
CB	cord blood
CBP	CREB-binding protein
cDNA	complementary deoxyribonucleic acid
CD	cluster of differentiation
.CDF	chip description file
CEB	cytosolic extraction buffer
.CEL	cell intensity file
CFU	colony-forming unit
.CHP	chip file
CI	confidence interval
CML	chronic myeloid leukaemia
CMP	common myeloid progenitor
CMV	cytomegalovirus immediate early promoter
CR	complete remission
cRNA	complementary ribonucleic acid
CSF	colony stimulating factor
Ct	cycle threshold
CtBP	C-terminal binding protein
.DAT	data file
DMEM	Dulbecco's Modified Eagles's Medium
DMSO	dimethyl sulfoxide
dn	dominant negative
DNA	deoxyribonucleic acid
DNAse	deooxyribonuclease
dNTP	deoxynucleotide triphosphate
dsRED	Discosoma sp. Red Fluorescent Protein
DTT	dithiothreitol
DWD	Distance Weighted Discrimination
ECL	enhanced chemiluminescence
ECOG	Eastern Cooperative Oncology Group

EDTA	ethylenediamine tetraacetic acid
EPO	erythropoietin
EST	Expressed Sequence Tags
EV	empty vector
FAB	French-American-British
FACS	fluorescence-activated cell sorting
FCS	foetal calf serum
FDR	false discovery rate
FITC	fluorescein isothiocyanate
FLT3	fms-like tyrosine kinase-3
FSC	forward scatter
G	guanine
gag	group-specific antigen
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GC	guanine cytosine
GC-RMA	Guanine Cytosine-Robust Multiarray Average
GCOS	GeneChip® Operating Software
G-CSF	granulocyte-colony stimulating factor
GEO	Gene Expression Omnibus
GEP	gene expression profiling
GFP	green fluorescent protein
GM-CSF	granulocyte monocyte-colony stimulating factor
GMP	granulocyte monocyte progenitor
GO	gene ontology
HBSS	Hank's balanced salt solution
HEPES	4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid
HG-	Human Genome-
HGNC	HUGO Gene Nomenclature Committee
HMG	high mobility group
HRP	horseradish peroxidase
HSC	haematopoietic stem cell
HPC	haematopoietic progenitor cell
HUGO	Human Genome Organization
IMDM	Iscove's Modified Dulbecco's Medium
inv	inversion
IL	interleukin
IVT	in vitro transcription
kb	kilo base
kDa	kilo Dalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Luria-Bertani
LDS	lithium dodecyl sulphate
LEF	lymphoid ehancer factor
LSC	leukaemic stem cell
LTR	long terminal repeat
M	molarity (moles per L)
MACS	magnetic-activated cell sorting
MCS	multiple cloning site
M-CSF	macrophage colony-stimulating factor
MILE	**M**icroarray **I**nnovations in **LE**ukemia
MM	mismatch

MNC	mononuclear cells
mRNA	messenger ribonucleic acid
MDS	myelodysplastic syndrome
M-MLV-RT	moloney murine leukemia virus-reverse transcriptase
MLL	myeloid/lymphoid , mixed lineage leukaemia
MMTV	mouse mammary virus tumor
MOPS	3-(N-morpholino)propanesulfonic acid
MPPs	multipotent progenitors
MRC	medical research council
MTG	myeloid translocation gene
MW	molecular weight
NCBI	National Centre for Biotechnology Information
NEB	nuclear extraction buffer
NLS	nuclear localisation signal
N-terminus	amino (NH2)-terminus
OR	odds ratio
OS	overall survival
p(A)	polyadenylation signal
PAGE	polyacrylamide gel electrophoresis
PB	peripheral blood
PBS	phosphate-buffered saline
PCA	principal component analysis
PCR	polymerase chain reaction
PE	R-phycoerythrin
PerCP	Peridinin chlorophyll protein
PLM	probe level model
PM	perfect match
pol	polyprotein viral reverse transcriptase gene
QC	quality control
qRT-PCR	quantitative reverse transcriptase-polymerase chain reaction
RAEB	refractory anaemia with excess blasts
RARα	retinoic acid receptor-alpha
RefSeq	Reference Sequence (from NCBI)
RIN	RNA integrity number
RMA	Robust Multiarray Average
RNA	ribonucleic acid
rpm	revolutions per minute
RPMI	Roswell Park Memorial Institute
RT	room temperature
RT-PCR	reverse transcriptase-polymerase chain reaction
SSC	side scatter
SD	standard deviation
SE	standard error
SDS	sodium dodecyl sulphate
SG	SYBR-Green 1
shRNA	small hairpin ribonucleic acid
SOC	super-optimal broth catabolite-repressing
T	thymine
TAE	tris acetate EDTA
TBE	tris borate EDTA
TBS	tris-buffered saline

TCF	T-cell Factor
TCF7L2	Transcription Factor 7 Like 2
TD	touchdown
TD-PCR	touchdown polymerase chain reaction
TE	tris-EDTA
TF	transcription factor
TLE	transducin-like enhancer
Tm	melting temperature
U	uracil
UniProt	Universal Protein Resources
UTR	untranslated region
WBC	white blood cell
WNT	wingless-type MMTV integration site family

Table of Contents

List of Figures

List of Tables

1 - General Introduction

1.1 THE HAEMATOPOIETIC SYSTEM

1.1.1 Overview of normal haematopoiesis

Haematopoiesis results in the continuous production of separate lineages of mature blood cells including erythrocytes, monocytes, macrophages, neutrophils, platelets and lymphocytes. In foetal life this process starts in the yolk sac and later migrates to the foetal liver and spleen (Medvinsky *et al*, 1993; Muller *et al*, 1994; Godin *et al*, 1995). Following birth, normal haematopoiesis is restricted to the bone marrow (BM), but also can be found in umbilical cord blood (CB) and peripheral blood (PB). Approximately 1 in every 10^4 BM cells is thought to be a haematopoietic stem cell (HSC) (Spangrude *et al*, 1988 ; Morrison *et al*, 1996). The ability to commit and to self-renew are defining properties of HSCs (Till & McCulloch, 1961) and as they differentiate and become committed to certain lineages, they express end-stage markers representative of each cell type. Differentiation of HSC is associated with a loss of self-renewal capacity or multipotency, thus requiring HSCs to self-renew to maintain the HSC pool. The ability of HSC to balance self-renewal and commitment to differentiation is accompanied by long periods of quiescence (G_0). Once commitment to differentiation takes place, active cell division occurs (Suda *et al*, 1983; Ogawa, 1993).

Cytokines play an important role in regulating haematopoiesis. While self-renewal of HSC appears to be a stochastic process (Nakahata *et al*, 1982; Tsuji & Nakahata, 1989), survival and proliferation and, to a certain, extent lineage fate of blood cells is regulated by growth factors and cytokines. The commitment of bipotential granulocyte-macrophage (GM) progenitor cells is regulated by interleukin-3 (IL-3), granulocyte-macrophage colony stimulating factor (GM-CSF), and interleukin-4 (IL-4) after they exit from G_0 and begin active cell proliferation (Metcalf, 1980; Metcalf & Burgess, 1982). GM-CSF can also interact synergistically with granulocyte CSF (G-CSF) on HSC progenitors to produce distinct colonies (McNiece *et al*, 1989). Instructive cytokine signalling can alter the balance

of endogenous lineage determining transcription factor (TF). Lineage conversion by ectopic overexpression of cytokine receptor shows strong context dependence. For example, expression of the GM-CSF receptor can instruct myeloid lineage conversion in common lymphoid progenitors (CLP) and pro-T cells but not in pro-B cells or megakaryocyte-erythroid progenitors (Sarrazin & Sieweke, 2011). Besides a direct instructive effect on an uncommitted cell, a permissive role of cytokine supporting the selective survival or proliferation of a daughter cell could also be derived from a cytokine independent commitment division. In support of this, the development of monocytes / macrophages has been reported previously in macrophage CSF (M-CSF) deficient mice (Lagasse & Weissman, 1997).

Proliferation and maturation of committed HSC is stimulated by several late-acting lineage specific cytokines such as erythropoietin, M-CSF, G-CSF, and interleukin-5 (IL-5) (Ogawa, 1993). For example, G-CSF stimulates the formation of colonies of granulocytes (Burgess & Metcalf, 1980) and IL-1, IL-3, IL-6, IL-11 and stem cell factor (SCF) have been shown to stimulate production of megakaryocytes and to increase platelet production (Bruno *et al*, 1991; Briddell *et al*, 1992). For erythrocyte production, erythropoietin is a an absolute requirement, otherwise the erythroid progenitor cells fail to proliferate and will undergo apoptosis (Silva *et al*, 1996).

The proliferation of quiescent HSC can also be activated in response to injury or injection of cytokines, such as G-CSF that induce the exit of HSCs from their niches and mobilisation of the cells into the circulation. In addition, the self renewal and maintenance of HSC progenitors have been linked to extracellular signals such as the Notch and Wingless-type MMTV integration site family (WNT). This was demonstrated by the inhibition of Notch signalling diminishing the capacity of HSCs to maintain an undifferentiated state but allowing normal proliferation and survival that must be intact for WNT proteins to enhance HSC renewal (Duncan *et al*, 2005).

1.1.2 Characteristics of HSC

HSCs can be classified into long-term, short-term and multipotential progenitor cells (MPPs), based on the extent of their self-renewal abilities and specific cell surface markers as shown in Figure 1.1. It has been reported that only 25% of the non-cycling or quiescent HSCs are long-term MPPs (Osawa et al, 1996; Muller-Sieburg et al, 2002), and others are transient or short-term (Morrison & Weissman, 1994). Although 75% of long-term HSCs are quiescent at any one time, over 90% of long-term HSCs actually enter the cell cycle at least once every 30 days to generate additional HSCs to retain the HSC pool (Cheshier et al, 1999).

Both the long-term or short-term HSCs cannot be recognised morphologically as they resemble lymphocytes. In mice HSCs can be identified by cell surface markers and Lin^- $Sca\text{-}1^+$ $c\text{-}kit^+$ has been generally used as a canonical marker set for HSC enrichment (Spangrude *et al*, 1988 ; Ogawa *et al*, 1991). However studies have suggested that the HSC compartment in adult BM is heterogeneous since different HSCs show clearly distinguishable repopulation patterns and distinct intrinsic properties of clonality (Muller-Sieburg *et al*, 2002; Dykstra *et al*, 2007; Morita *et al*, 2010). Studies also suggest that the Lin^- $Sca\text{-}1^+$ $c\text{-}kit^+$ $CD34^-$ $CD150^+$ $CD48^-$ fraction in adult BM contains truly dormant HSCs and this population is the most primitive long-term HSC, whereas Lin^- $Sca\text{-}1^+$ $c\text{-}kit^+$ $CD34^+$ cells revealed early but only short-term reconstitution. These dormant cells could reversibly switch from dormancy to self-renewal stage and have been reported to divide approximately five to six times during the murine life span (Wilson *et al*, 2008). In humans, overwhelming evidence suggests a primitive population of stem cells expressing Lin^- $CD34^-$ $CD38^-$ HSC in human BM with long-term multilineage repopulating ability (Osawa *et al*, 1996; Goodell *et al*, 1997; Nakamura *et al*, 1999).

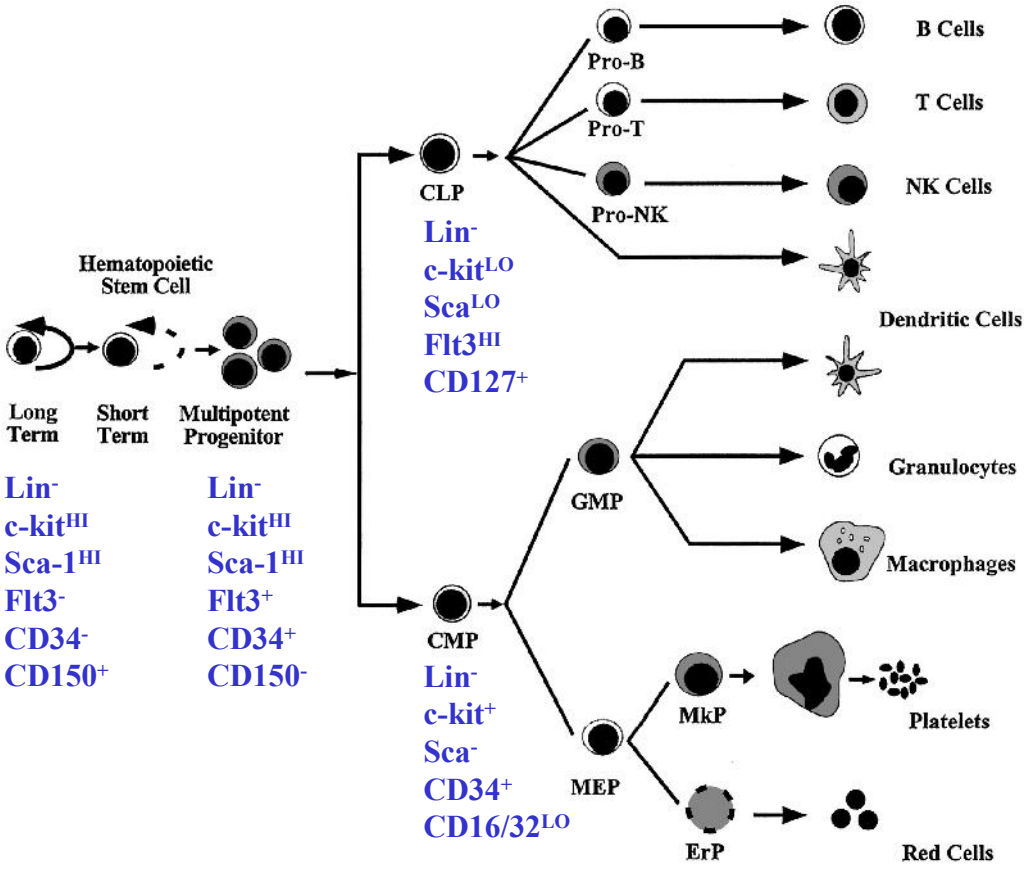


Figure 1.1: Haematopoietic development in the mouse. Classification of HSC into long-term, short-term, and MPP, based on surface expression markers. Shown are Common lymphocyte progenitor (CLP), common myeloid progenitor (CMP), granulocyte-monocyte progenitor (GMP), megakaryocyte-erythrocyte progenitor (MEP), natural killer (NK) cells. Adapted from (Reya, 2003) and (Kokolus & Nemeth, 2010).

1.1.3 HSC microenvironment

Self-renewal and the developmental capacity of HSCs can be controlled in a non-autonomous manner by their cellular microenvironment. Such a microenvironment is usually referred to as a stem cell niche (Nagasawa, 2006; Hsu & Fuchs, 2012). HSCs populate the BM niches that consist of adherent cells known as stromal cells which also control HSC dormancy and the balance between HSC self-renewal and differentiation. Although HSCs are mostly quiescent, they can be mobilised from their niche to proliferate and differentiate. As shown in Figure 1.2, the dormant long-term HSCs are mostly located close to the osteoblast lining or the endosteal niche. This niche contains osteoblastic cells and reticular cells expressing adhesion molecules such as CXC chemokine ligand-12 (CXCL12). These stromal cells of the BM containing fibroblasts, endothelial cells and CXCL12 react with corresponding ligands on HSC and maintain their viability (Wilson & Trumpp, 2006; Nagasawa, 2006). It has been recently determined that chemokine stimulation of HSCs by CXCL12 leads to an enhancement in stromal migration via activation of adhesion molecules, in addition to its well-known ability to stimulate motility (Greenbaum *et al*, 2013) and also modulates adhesion receptor function via CD44 (Avigdor *et al*, 2004). In human $CD34^+$ cells, CXCL12 promotes a rapid increase in the affinity of β1 integrin for its ligands, vascular cell adhesion molecule (VCAM1) and fibronectin (Hidalgo *et al*, 2001).

The HSCs that reside in the BM can be found in endosteal and perivascular niches. HSCs located on the endosteal side tend to be more quiescent, whereas HSCs located at the perivascular side are more active and the most dormant HSCs have been reported to locate near osteoblast progenitor cells. With the exception of osteoclasts, all of the cellular components in the BM niches have been reported to modulate HSC behaviour (Sugiyama *et al*, 2006; Nagasawa, 2007).

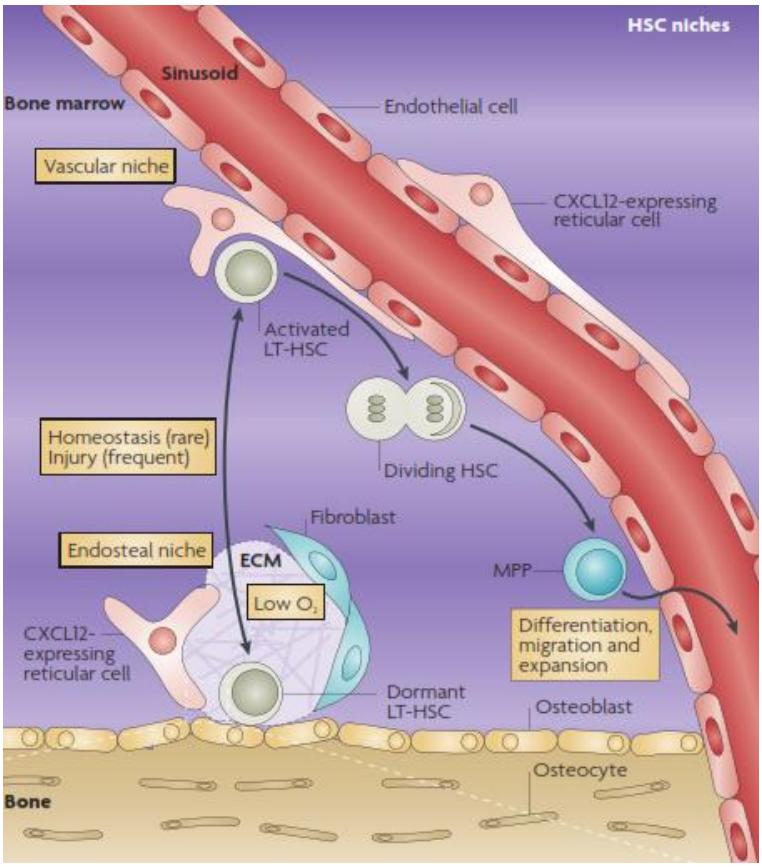


Figure 1.2: Overview of the complex environment of the HSC niche and the molecules that mediate interactions between HSCs, osteoblasts and reticular cells. Available from http://www.nature.com/nri/posters/hsc/nri0707_hsc_poster.pdf. Information was adapted from (Nagasawa, 2006) and (Hsu & Fuchs, 2012).

1.1.4 Transcription factors in haematopoiesis

A large number of transcription factors (TFs) are involved in the control of blood cell generation and development, for instance Stem Cell Leukaemia (SCL / TAL1), GATA binding protein GATA1, GATA2, Runt-related transcription factor 1 (RUNX1), CCAAT/enhancer-binding protein alpha (C/EBPα) and PU.1 (Scott *et al*, 1997; Lichtinger *et al*, 2012). The transcription factors PU.1 and C/EBPα are responsible for normal myeloid differentiation from stem cells to monocytes or granulocytes as shown in Figure 1.3. In particular, PU.1 induces expression of the M-CSF receptor and the development of monocytes, whereas C/EBPα increases the expression of the G-CSF receptor and leads to mature granulocytes (Behre *et al*, 1999). Dual expression of PU.1 and GATA1 leads development of HSCs to CMPs, but dominant expression of PU.1 is restricted to GMPs (Zhu & Emerson, 2002). During multilineage development of $CD34^+$ cells, C/EBP proteins appear to be expressed at high levels in neutophilic but not monocytic or erythroid cells (Tenen *et al*, 1997).

Some TFs regulate HSC self-renewal such as Homeobox protein B4 (HOXB4), a key TF that has been reported to induce the *in vitro* expansion of HSCs via self-renewal (Beslu *et al*, 2004; Sharma *et al*, 2006). B-lymphoma Mo-MLV insertion region 1 homolog (Bmi1) on the other hand is important for long-term maintenance of adult HSCs but not for foetal HSCs (Park *et al*, 2003). HOXB4 overexpression triggered the expansion of short-term repopulating HSCs in the absence of Bmi1 showing independent function in both TF and HSC activities (Faubert *et al*, 2008). Other HOX genes such as HOXB3 and HOXB1 are expressed at higher levels in populations containing immature HSCs. Other TFs are known to influence lineage fate. For instance, in the absence of GATA1 or SCL / TAL1, embryonic erythropoiesis was impaired suggesting crucial roles of these TF to establish lineage decisions as reviewed previously (Orkin, 1995).

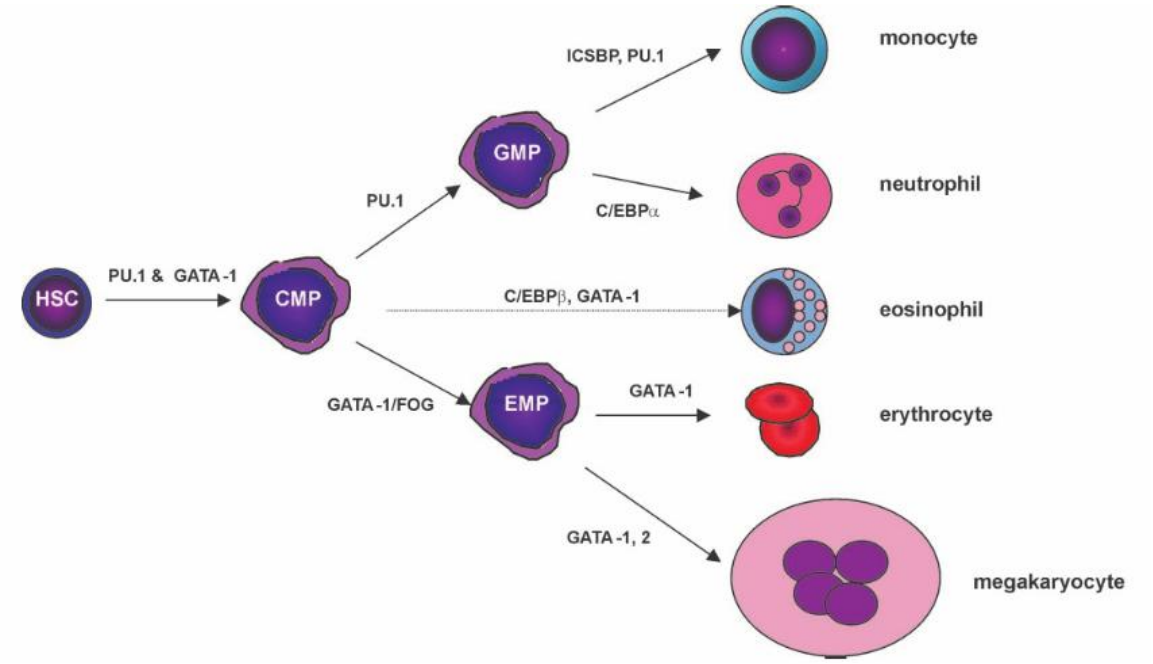


Figure 1.3: Important TFs for haematopoietic development of myeloid lineage. Shown are transcriptional regulation of CMP, GMPs and erythroid and megakaryocytic lineages (EMPs). Adapted from (Zhu & Emerson, 2002).

1.2 ACUTE MYELOID LEUKAEMIA

1.2.1 Pathophysiology of Acute Myeloid Leukaemia (AML)

AML is an aggressive malignancy of the BM. It is a clonal disorder characterised by the inhibition of differentiation resulting in the subsequent accumulation of cells at various stages of incomplete maturation. The pathophysiology of AML can be explained by acquired genetic changes in HSC that cause a complete or partial block in normal HSC maturation. AML is a heterogeneous disease and its heterogeneity is evident from variations in morphology, immunophenotype, cytogenetics and molecular abnormalities (Schoch *et al*, 2002). Genomic and functional studies have demonstrated that multiple mutations are necessary to transform normal cells into a leukemic clone (Ley *et al*, 2008; Mardis *et al*, 2009). Since HSCs are the only self-renewing cells among BM progenitors, a model has been proposed that mutations must sequentially accumulate within distinct clones of HSC over time. In the case of AML, genome sequencing suggests that up to 10 mutations are serially acquired in a single cell lineage that ultimately generates a dominant leukaemic clone and from 750 point mutations, only small subsets are relevant for the development of AML (Jan *et al*, 2012; Jan & Majeti, 2013), in which most AML mutations are probably background events in HSCs (Welch *et al*, 2012).

The molecular pathogenesis of AML has not yet been completely defined. Recurrent chromosomal structural variations (e.g. t(15;17), t(8;21), inv(16), del5, del7 and others) and mutations leading to activation of the receptor tyrosine kinase (e.g. Fms-like tyrosine kinase 3 (FLT3) and c-KIT) signalling are established diagnostic markers, suggesting that these acquired genetic abnormalities play an essential role in leukaemogenesis (Betz & Hess, 2010). Expression profiling studies have yielded signatures that correlate with specific cytogenetic subtypes of AML, but have not yet suggested new initiating mutations. A two-hit model of AML has been described in which the critical events are an activating mutation of a kinase combined with a mutation that alters the function of a haematopoietic TF (Dash & Gilliland, 2001). This model is partially supported by data from AML samples in which these two types of mutations have been documented (Schaub *et al*, 2010; Ding *et al*, 2012). Recent studies suggest that a third complementation group of mutations in the epigenome should be added to the two hit model

(Figueroa *et al*, 2010; Chen *et al*, 2013). AML displays global DNA hypermethylation and a specific hypermethylation signature.

AML can be viewed a disease which retains some of the hierarchical characteristics of normal haematopoiesis where repopulating activity is restricted to a subset of cells known as leukaemic stem cells (LSC) (Lowenberg & Terpstra, 1998). In some cases the immunophenotype of the cells can be similar to that of normal HSCs (e.g. CD34, CD38, CD71 and HLA-DR), some antigens display leukaemia-specific characteristics (CD90, CD117 and CD123) (Terpstra *et al*, 1996; Blair *et al*, 1998; Blair & Sutherland, 2000). The heterogeneity among leukaemic blasts with respect to their capacity to proliferate (McCulloch et al, 1988), supports a hierarchy within leukaemic blast populations in patients with AML (Sutherland et al, 1996). However in AML, LSCs are not necessarily rare which challenges the hierarchical model. Other complications with the LSC model include the lack (as yet) of any consistent phenotype for LSCs, the reliance on immunocompromised mice for its definition and the evidence for polyclonality within LSC populations, restricting their potential as therapeutic tagets (Lutz *et al*, 2013).

In leukaemogenesis, it is presumed that certain oncoproteins could either inappropriately activate a program of growth or interfere with terminal differentiation (Kvinlaug *et al*, 2011). Normal primitive cells, rather than committed progenitor cells, are the target for most leukaemic transformation (Bonnet & Dick, 1997), though AML may also arise from more committed progenitor cells caused by mutations or selective expression of genes that enhance their otherwise limited self-renewal capabilities. In either case the consequence is a developmental block giving rise to an accumulation of leukaemic blasts (Figure 1.4). This impacts on normal haematopoiesis with a reduction in erythrocytes, granulocytes and platelets due to overcrowding in the BM. The disease is clinically heterogeneous with biologically distinct subtypes as discussed further in section 31.2.3. The term ‘acute’ is applied because the disease is usually aggressive and characterised by large numbers of very immature, undifferentiated cells that, if untreated, can lead to the rapid death of the patient. These immature leucocytes are called blasts and they represent an early phase of the normal differentiation process that occurs in the BM.

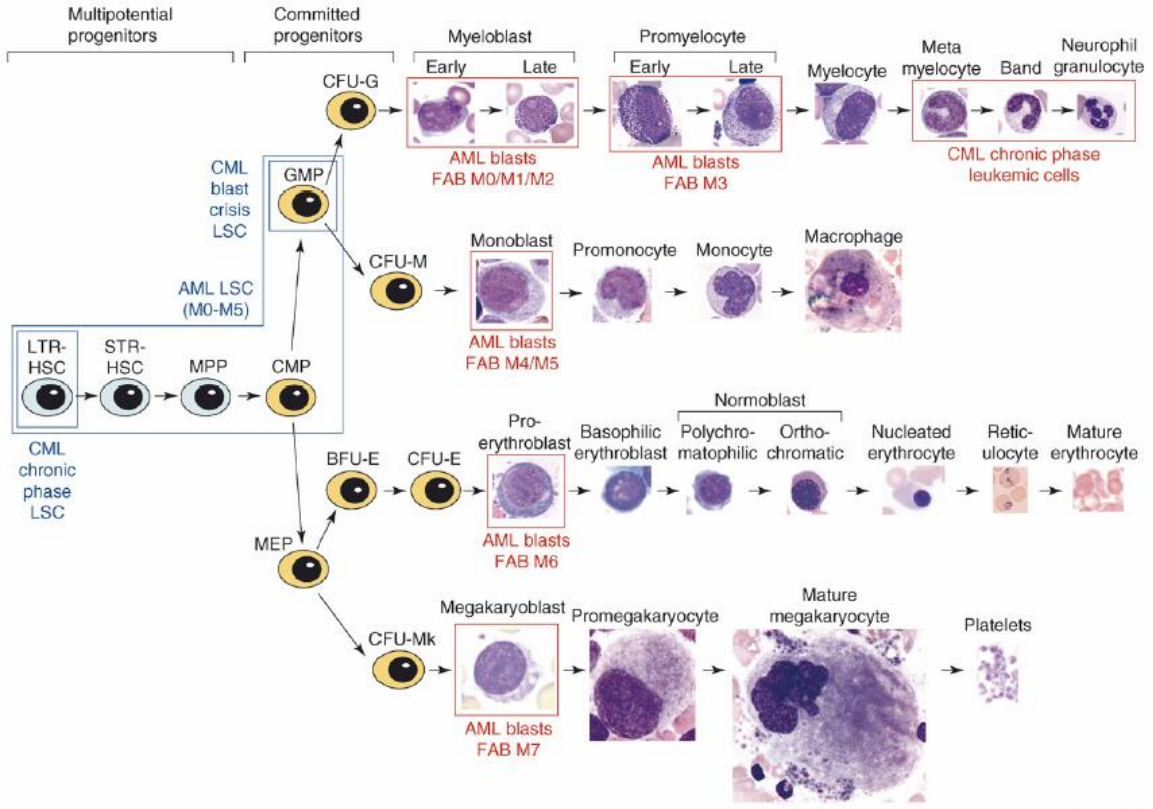


Figure 1.4: Diagram of normal myeloid development and its relationship to both leukaemic cells and LSCs. The LSCs for AML are restricted to multipotential and committed progenitor cells, as indicated by the blue boxes. Multipotential stem and progenitor cells, including long-term repopulating and short-term repopulating HSCs, are depicted in blue on the left. Differentiating myeloid cells, recognisable by their distinct morphology, are shown on the right. The malignant cells in AML are indicated by red boxes; leukaemic blasts for the different French American British (FAB) subclasses of AML (M0 through M7) correspond approximately to the different normal blasts in each lineage (Krause & Van Etten, 2007).

1.2.2 Diagnosis and incidence of adult AML

Patients with AML have excess numbers of blast cells in the BM and usually in the PB as well. The presence of over 20 – 30% myeloblasts in the BM at clinical presentation is the basis of defining AML (Harris *et al*, 1999). AML incidence in the UK is around 2,500 people each year and the risk of developing AML increases with age (Shah *et al*, 2013). AML is the most common variant of acute leukemia occurring in adults, comprising approximately 80 to 85% of cases of acute leukemia diagnosed in individuals greater than 20 years of age. It is most common in people over 65 years old, AML accounts for 10 to 15% of newly diagnosed cases of childhood leukaemias. Historically, paediatric AML patients have a worse prognosis than Acute Lymphoblastic Leukamia (ALL) patients. Adults with AML remain at increased risk of induction failure (Burnett, 2012), early relapse, and isolated central nervous system relapse (Meshinchi & Arceci, 2007).

Diagnosis of AML requires the examination of BM morphology, cytochemistry and immunophenotyping. Full blood count investigation usually shows features such as anaemia, thrombocytopenia and variable leucocyte count with or without the presence of blast cells in the PB. About one-quarter to one-third of cases begin with a low white blood cell (WBC) count, while about half of patients show some degree of leucocytosis. The most common presenting symptoms are pallor and fatigue secondary to anaemia, bleeding problems secondary to thrombocytopenia, infection due to neutropenia and hepatomegaly and splenomegaly due to infiltration by blasts (Estey & Dohner, 2006). Disseminated intravascular coagulation occurs more commonly in Acute Promyelocytic Leukaemia (APL).

Immunophenotyping using flow cytometry and panels of monoclonal antibodies is very useful to assign lineage to confirm AML diagnosis. This technique can classify the subtypes of AML or biphenotypic leukaemia if there is evidence of both lymphoid and myeloid markers. Monoclonal antibodies directed against antigens (defined within the cluster of differentiation (CD) system) against CD11b, CD13, CD14, and CD33 are considered to be restricted to cells committed to myeloid differentiation and are used for diagnostic distinction (Bradstock *et al*, 1994; Harris *et al*, 1999).

1.2.3 Classification

AML is a highly heterogeneous disease and the subclassification of AML depends on morphological, immunophenotyping, cytogenetic and molecular criteria. Originally AML was subdivided based on leukaemic blast morphology according to the FAB classification in 1976 (Table 1.1). The FAB classification system divides AML into eight subtypes, M0 through to M7, based on its degree of maturity and the type of cell from which the leukaemia developed (Bennett *et al*, 1976). This has now been largely superseded by a World Health Organization (WHO) classification (Vardiman *et al*, 2002). This WHO classification incorporates recurring cytogenetic and molecular genetic abnormalities as shown in Table 1.2. Molecular characterisation of leukaemia cells enables more precise diagnosis and prognosis than is possible with the FAB classification.

FAB classification of AML	% of total	Prominent features
M0	2	Large and granular blasts with minimal myeloid differentiation, expression of at least one myeloid antigen either CD13 or CD33
M1	10-18	The cells in the BM show some evidence of granulocytic differentiation. Poorly differentiation myeloblasts with occasional Auer rods (elongated clumps of azurophilic granule material found in the cytoplasm of leukaemic blasts)
M2	27-29	More than 50% of the BM cells are myeloblasts and promyelocytes
M3	5-10	Hypergranular abnormal promyelocytes with Auer rods
M4	16-25	Both granulocytic and monocytic differentiation are present in varying proportions in the BM and PB. M4 resembles M2 in all respects except that the proportion of promonocytes and monocytes exceeds 20% of the nucleated cells in the BM
M5	13-22	Monoblastic differentiation
M6	1-3	Myeloblastic leukaemia with megablastoid features or leukaemia with erythroblastic differentiation
M7	4-8	Megakaryoblastic differentiation with frequent BM fibrosis

Table 1.1: AML classification based on FAB criteria. Adapted from (Bennett *et al*, 1976).

WHO classification of AML
AML with recurrent genetic abnormalities:
AML with t(8;21)(q22;q22), (RUNX1/ETO)
AML with abnormal BM eosinophils and inv(16)(p13q22) or t(16;16)(p13;q22), (CBFβ/MYH11)
Acute promyelocytic leukaemia (APL) with t(15;17)(q22;q12), (PML/RARα) and variants
AML with 11q23 (MLL) abnormalities
AML with multilineage dysplasia:
Following MDS or MDS/ myeloproliferative disease (MPD)
Without antecedent MDS or MDS/MPD, but with dysplasia in at least 50% of cells in two or more myeloid lineages
AML and myelodysplastic syndromes, therapy related:
Alkylating agent/radiation–related type
Topoisomerase II inhibitor–related type (some may be lymphoid)
AML, not otherwise categorised:
AML, minimally differentiated
AML without maturation
AML with maturation
Acute myelomonocytic leukemia
Acute monoblastic/acute monocytic leukemia
Acute erythroid leukemia (erythroid/myeloid and pure erythroleukemia)
Acute megakaryoblastic leukemia
Acute basophilic leukemia
Acute panmyelosis with myelofibrosis
Myeloid sarcoma

Table 1.2: AML classification based on WHO. Adapted from (Vardiman *et al*, 2002).

1.2.4 Prognostic factors

Prognosis is the expected course of a disease and the patient's chance of recovery. The prognosis predicts the outcome of a disease and therefore the future for the AML patient and different intensity of treatment is assigned according to the risk groups. Cytogenetic analysis provides the most powerful independent predictor of disease outcome in AML allowing the assignment of patients to groups with favourable or unfavourable (high risk) prognoses as listed in Table 1.3 (Grimwade & Hills, 2009 ; Grimwade *et al*, 2010).

Favourable prognostic factors have been associated with the presence of the cytogenetic translocations t(8;21), t(15;17) and inv(16), (Fenaux & Detourmignies, 1994; Meshinchi & Arceci, 2007). In contrast, patients with 11q23 abnormalities, monosomy 5, monosomy 7 or complex karyotypes that involve the presence of more than three abnormalities carry a poor prognosis. The remainder of AML patients are assigned to a prognostically intermediate group. This latter group is very heterogeneous because it includes patients with a normal karyotype as well as those with rare chromosome aberrations. Currently, risk stratification of AML may be further refined by consideration of Nucleophosmin1 (*NPM1), C/EBPα,* Fms-like tyrosine kinase 3-internal tandem duplication (*FLT3-ITD*) and mixed-lineage leukemia gene-partial tandem duplication *(MLL-PTD)* mutations (Foran, 2010). Significantly worse outcome with regard to overall survival (OS) is seen in patients harbouring an *NPM1* mutation together with higher *FLT3-ITD* expression as compared to those with mutated *NPM1* and a low *FLT3-ITD* expression. This is an independent subgroup with a unfavourable prognosis (Schneider *et al*, 2012). Chou *et al* showed that a median 2-log decline in *NPM1* mutant copy number was obtained following induction therapy. Any increase of mutant numbers was always accompanied by three-fold increased risk of relapse compared to patients with persistently low signals (Chou *et al*, 2007). In addition to cytogenetics, WBC count at diagnosis of more or equal to 30 x 10^9 /L has also been reported as a significant poor prognostic factor, predicting a lower OS and is independent of the treatment used.

Risk status	Cytogenetics	Molecular abnormalities
Favourable	t(8;21)(q22;q22) inv(16)(p13.q22), t(16;16)(p13.q22) t(15;17)	Normal cytogenetics with NPM1 mutation or C/EBPα mutation in absence of FLT3-ITD
Intermediate	Normal cytogenetics +8 t(3;5)4 t(9;11)(p22q23) Entities not classified as favourable or adverse	c-KIT mutation with: t(8;21)(q22;q22), or inv(16)(p13.q22), t(16;16)(p13.q22)
Adverse	Complex (≥ 4 unrelated abnormalities) abn(3q) [excluding t(3;5)(q21 ~ 25;q31 ~ 35)], inv(3)(q21q26)/t(3;3)(q21;q26), add(5q), del(5q), -5, −7, add(7q)/del(7q), t(6;11)(q27;q23), t(10;11)(p11 ~ 13;q23), Other t(11q23) [excluding t(9;11)(p21–22;q23) and t(11;19)(q23;p13)] t(9;22)(q34;q11), −17/abn(17p)	High Ecotropic Viral Integration Site 1 (EVI1) expression (with or without 3q26 cytogenetic lesion) Normal cytogenetics with FLT3-ITD in the absence of NPM1 mutation

Table 1.3: Prognostic subgroups of AML based upon presenting cytogenetics and genetic lesions. The revised Medical Research Council (MRC) cytogenetic risk group was based on multivariable analysis conducted in 5876 adults (16–59 years old) treated in the MRC AML10, 12 and 15 trials (Grimwade & Hills, 2009 ; Grimwade *et al*, 2010 ; Smith *et al*, 2011).

1.2.5 AML treatment and outcome

1.2.5.1 Conventional therapy

Treatment for AML consists of remission induction chemotherapy followed by post-remission chemotherapy with or without BM transplantation. Achievement of complete remission (CR) is the essential first step in effective treatment of AML. For most patients, remission induction regimes usually comprise a combination of anthracycline, (e.g. daunorubicin), cytosine arabinoside with either etoposide or thioguanine (Stone & Mayer, 1993). This regimen has been capable of inducing CR rates of 65-75% in adults aged 18-60 years (Tallman *et al*, 2005). Delivery of high dose anthracycline-cytosine arabinoside (Ara-C) to younger AML patients aged 15-24 years will provide a 40 - 45 % chance of cure (Figure 1.5) but there is little evidence for increasing cure rates in older patients (Figure 1.6) (Burnett, 2012). Cure could not be estimated for patients over 70 years, because survival was consistently low (<5%) (Shah *et al*, 2013) and 90% of older adults still die of their disease (Rowe & Tallman, 2010). Whilst remission will be achieved in up to 80% of those receiving intensive chemotherapy, the main variables precluding cure are treatment-related mortality and relapse rates. Only around half of older AML patients with age above 60 years will enter CR and around 85% will relapse within two years (Burnett, 2012). For patients under 60 years, a remission rate of 80% has been reported, and about half of patients will survive (Newland, 2002). Younger AML patients will have around 40% chance of cure by allogeneic transplantation as an alternative to consolidation therapy (Koreth *et al*, 2009). For an optimal treatment approach in AML, both a precise diagnosis and prognostic parameters that determine response to therapy and survival are needed.

1.2.5.2 Targeted therapy of AML

Over the years, treatment outcomes in younger patients with AML have improved, but optimisation and new combinations of drug therapy are needed. Patients with APL (or AML with M3 subtype) are given all-trans retinoic acid (ATRA), a vitamin A derivative, to induce differentiation of promyelocytes (Tallman *et al*, 1997; Tallman *et al*, 2002). Remission induction invariably follows a period of 15 to 25 days of severe BM hypoplasia and a stable remission must be associated with less than 5% blasts in BM with evidence of recovery of normal cellular elements. With this regimen, about 75% to 85% of patients will enter CR. APL patients especially those with a low WBC count have substantial benefit

from extended ATRA treatment indicated by a high remission rate due to fewer early and induction deaths and less resistant disease (Burnett *et al*, 1999). CR rates greater than 90% can be achieved in newly diagnosed APL by adding Ara C to ATRA from the onset of treatment or when leucocyte counts rapidly increases.

In contrast, chromosome aberrations with an unfavourable clinical course including -5/del(5q), −7/del(7q), inv(3)/t(3;3) and complex aberrant karyotypes show cure rates of less than 10%. CR rates and OS were not improved even by using combined drug treatments (e.g. multidrug resistance gene-1 (MDR-1) modulator valspodar) as compared to chemotherapy alone in AML patients with poor-risk AML (Greenberg *et al*, 2004). Trials combining FLT3 inhibitors (e.g. sorafenib) with chemotherapy in which *FLT3* mutant AML patients are enrolled, have shown highly promising activity in early trials with the ability to achieve more sustained in vivo inhibition of FLT3 (Sato *et al*, 2011; Knapper, 2011). The addition of nucleoside analogues appeared to be beneficial in AML with adverse cytogenetics. The inclusion of cladaribine but not fludaribine improved remission rate and OS (Burnett, 2012). Other targeted approaches for AML include immunoconjugate or antibody-directed chemotherapy such as Gemtuzumab Ozogamicin (Mylotarg), which is directed to CD33 expressed in most leukemic blast cells but also in normal haematopoietic cells (Wheatley *et al*, 1999; Newland, 2002). This has gained approval for treatment of older patients in relapse who are considered unsuitable for intensive therapy (Larson *et al*, 2002; Cheson *et al*, 2003). Randomised trials have found a benefit for Gemtuzumab Ozogamicin in newly diagnosed patients with favourable-risk AML, as reviewed previously (Ravandi *et al*, 2012). Since AML differs widely both clinically and in molecular genetic heterogeneity, optimal management of AML may eventually encompass a combination of cytotoxic and targeted therapies.

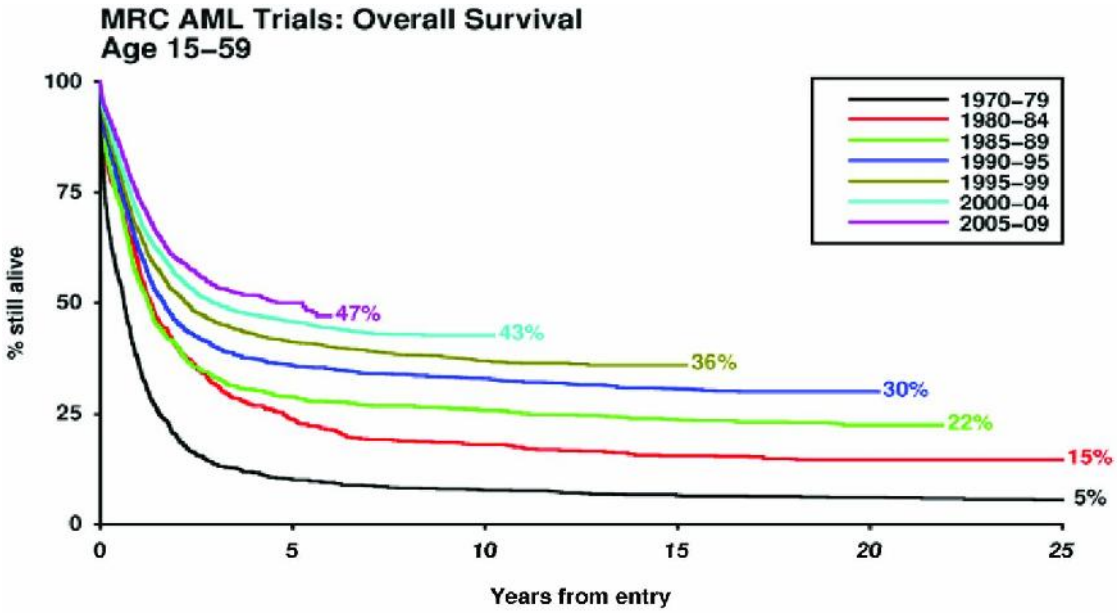


Figure 1.5 : Survival for younger AML patients. Adapted from (Burnett, 2012).

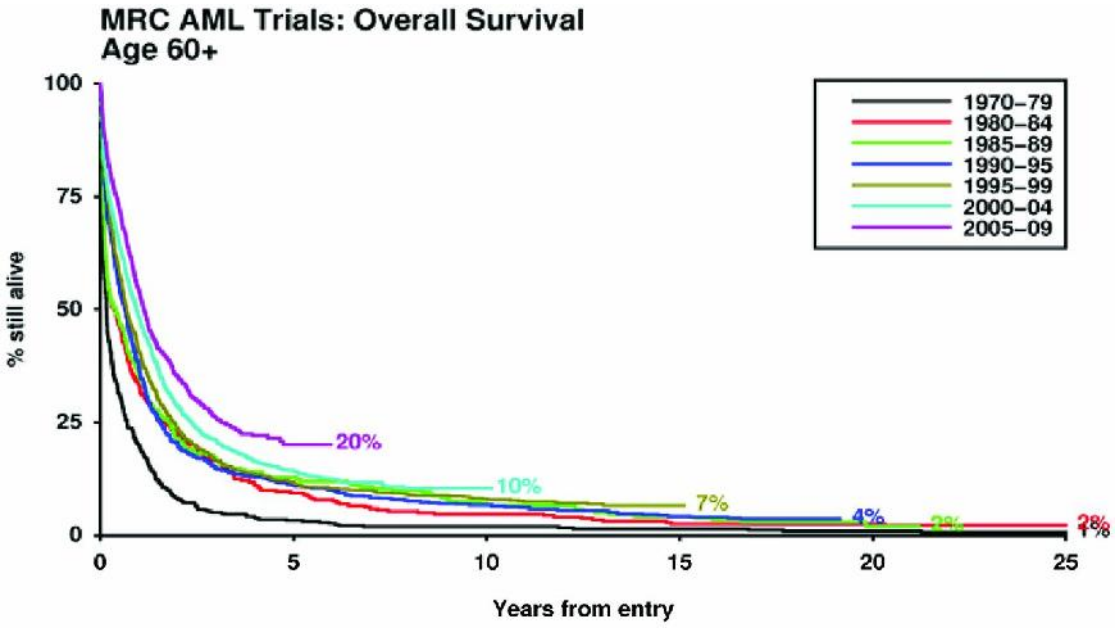


Figure 1.6 : Survival for older AML patients. Adapted from (Burnett, 2012).

1.3 WNT SIGNALLING

1.3.1 Overview of canonical and non-canonical WNT signalling pathways

WNT signalling plays a critical role in the control of cell proliferation and cell differentiation in many contexts. There are two main pathways of WNT signalling known as the 'canonical' and 'non-canonical' (refer **Figure 1.7**). The main function of the canonical WNT pathway is its association with Catenin (Cadherin-Associated Protein) Beta-1 or CTNNB-1 hereafter referred to as β-catenin. Canonical and non-canonical pathways exhibit different signalling events; the canonical WNT pathway employs β-catenin to activate WNT target gene expression through binding to lymphoid enhancer-binding factor (LEF) / T-cell factor (TCF) TFs which is the primary focus of this thesis; whereas the non-canonical signalling constitutes the WNT / Ca^{2+} pathway and the WNT / planar cell polarity (PCP) pathway involving activity of other downstream genes particularly calmodulin kinase II (Cam-KII), protein kinase C (PKC) and c-jun kinase (JNK). In some cases, the non-

canonical WNT signalling cascade can also inhibit nuclear β-catenin activity via activation of WNT5A (Ishitani *et al*, 2003).

Canonical WNT signalling (also known as WNT / β-catenin, or TCF / β-catenin pathway) is a tightly regulated pathway in developmental processes. The most crucial event in canonical WNT signalling is the cytoplasmic accumulation of β-catenin and its subsequent nuclear translocation followed by its binding to TCF or LEF proteins (Clevers & van de Wetering, 1997; Cavallo *et al*, 1998 ; Novak & Dedhar, 1999). Under non-stimulated conditions, a β-catenin destruction complex formed by AXIN, adenomatous polyposis coli (APC), and glycogen synthase kinase 3 beta (GSK3β) will keep cytoplasmic levels of β-catenin low through phosphorylation by GSK3β (Eastman & Grosschedl, 1999). Following WNT agonist binding to a receptor complex such as by WNT family, frizzled (FZD) family or low density lipoprotein receptor-related protein (LRP) (Cong *et al*, 2004), the AXIN-APC-GSK3β complex is inhibited, leading to β-catenin accumulation in the cytoplasm (Rubinfeld *et al*, 1996; Eastman & Grosschedl, 1999; Gordon & Nusse, 2006; MacDonald *et al*, 2009). When β-catenin enters the nucleus, it replaces repressor proteins such as Groucho, Transducin-like enhancer protein 1 (TLE) or histone deacetylases from TCF proteins and converts the complex into a transcriptional activator, thereby activating transcription of WNT target genes (Aberle *et al*, 1997; Cavallo *et al*, 1998 ; Clevers, 2006). As nuclear effectors, they act as multimeric transcription complexes that either repress or activate WNT target gene expression.

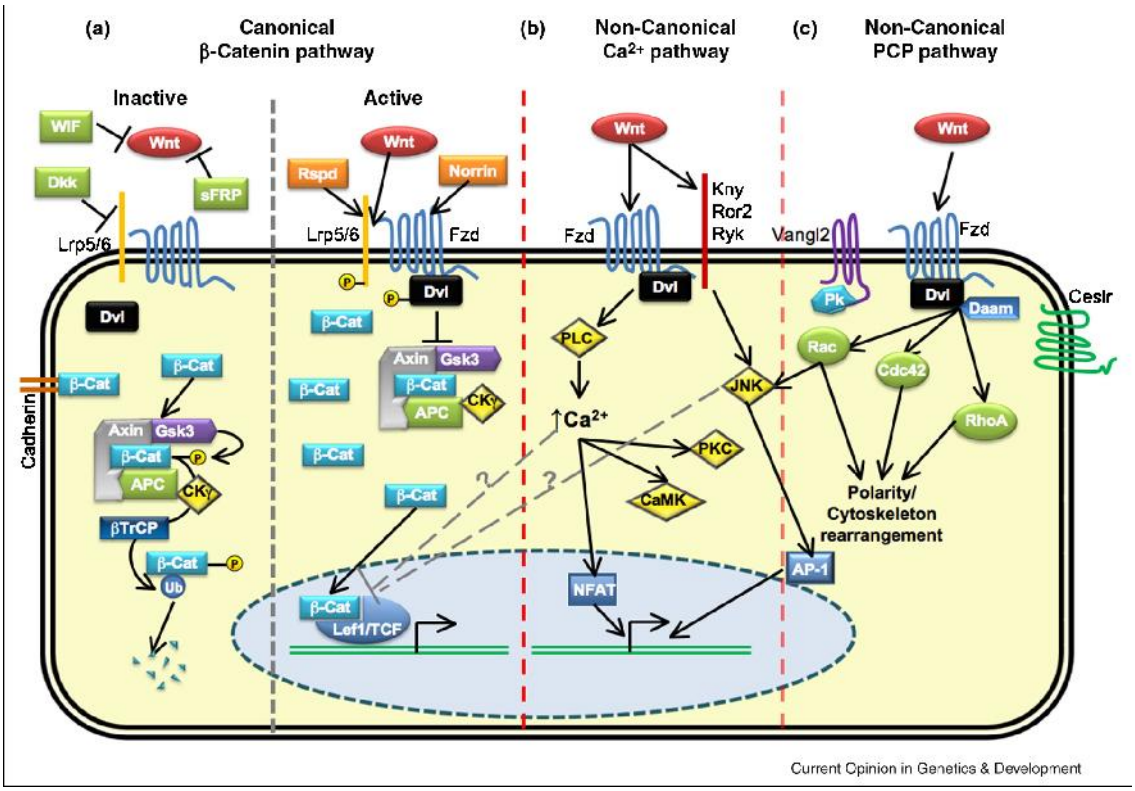


Figure 1.7: The three types of WNT-dependent pathway. (A) The canonical WNT/β-catenin pathway. Inactive-state: in the absence of WNT signal, β-catenin is phosphorylated by the destruction complex composed of Axin, APC, and GSK3β. Phosphorylated β-catenin is then targeted for degradation. Active-state: WNT binding to the Frizzled and LRP receptors induces phosphorylation of LRP and recruitment of Axin and the Axin-APC-GSK3β complex is inhibited, leading to accumulation of cytosolic β-catenin. **(B and C)** the non-canonical WNT / Ca^{2+} and WNT / PCP pathways (Franco *et al*, 2009).

1.3.2 Roles of WNT signalling in haematopoiesis and normal development

The requirement of WNT signalling activity in HSC self-renewal and BM repopulation has been indicated by the positive effect of WNT activation on HSC recovery in transplantation studies (Reya, 2003; Congdon *et al*, 2008) and that WNT activation through TCF / β-catenin signalling was necessary for optimal HSC formation (Goessling *et al*, 2009) and HSC integrity (Malhotra & Kincade, 2009). WNT5A, WNT2B or WNT10B from transduced stromal cells, stimulated expansion of human $CD34^+$ progenitors (Van Den Berg *et al*, 1998). Homeodomain protein TFs such as HOXA9 and HOXA10 can regulate WNT10B expression in human $CD34^+$ progenitors implying a role in early haematopoietic differentiation (Ferrell *et al*, 2005). A number of recent studies suggest that WNT signalling may play a role in vascular endothelial growth control and differentiation in which increased accumulation of β-catenin by WNT1 ligand binding causes increased endothelial proliferation (Wright *et al*, 1999; Franco *et al*, 2009). WNT signalling also regulates a wide range of processes including cell differentiation and promotion of growth and development during early stages of embryogenesis (Fleming *et al*, 2008). In vertebrates, WNT is not only a crucial factor influencing body axis and polarity during early development (St Amand & Klymkowsky, 2001; Petersen & Reddien, 2009) but also in directing early precursor cells in multiple organs such as for neuronal (Hirabayashi *et al*, 2004) and forebrain development (Vacik *et al*, 2011), liver regeneration (Goessling *et al*, 2008) and thymocyte differentiation (Verbeek *et al*, 1995; Castrop *et al*, 1995). During normal development, APC proteins play important roles for cytoskeletal functions regulating morphogenesis and cell-cell adhesion (Barth *et al*, 1997; Penman *et al*, 2005; Nathke, 2006).

1.3.3 Alteration of WNT signalling in leukaemias and other malignancies

Opposing effects of WNT signalling on haematopoiesis have been reported. Stabilised forms of β-catenin could result in immature phenotype or exhaustion of the HSC pool. These differences might be explained by different levels of WNT activation (Cobas *et al*, 2004; Kirstetter *et al*, 2006). Dysregulation of WNT signalling has been implicated in many cancers (Cadigan & Nusse, 1997; Reya & Clevers, 2005) including the progression of AML and other haematological malignancies (Lu *et al*, 2004; Simon *et al*, 2005). While β-catenin itself appears to be important for normal haematopoiesis (Cobas *et al*, 2004; Reya &

Clevers, 2005), some AML patients' cells strongly overexpress β-catenin (Tsutsui *et al*, 1996; Chung *et al*, 2002; Hwang *et al*, 2002). In models of T-Acute Lymphoblastic Leukaemia (ALL), thymus specific expression of activated β-catenin leads to development of thymic lymphoma (Guo *et al*, 2007). Different percentages and WNT signaling levels on a per cell basis have been reported. Foetal and leukaemic stem cells and thymocyte development may require higher WNT activity than normal adult HSCs (Luis *et al*, 2012). Precise amounts of WNT signals are required to maintain HSC integrity. Interactions between HSCs and niche cells may also alter the secretion of WNT proteins, and changes in balances between these factors may lead to leukaemia or immunodeficiencies (Sengupta *et al*, 2007). Studies have also described a correlation of β-catenin expression with poor prognosis of AML patients (Ysebaert *et al*, 2006; Xu *et al*, 2008). Besides β-catenin and the GSK3β complex, relatively little is known about the role of other WNT signalling molecules such as the FZD receptors in the haematopoietic system or in AML. A number of studies have demonstrated dysregulation of WNT signalling in AML in terms of epigenetic alterations particularly of the extracellular WNT antagonists, including secreted frizzled-related proteins (sFRPs), WNT5a, Dickkopf-1 (DKK1) and WNT inhibitory factor 1 (WIF1) gene promoter hypermethylation (Suzuki *et al*, 2007; Jost *et al*, 2008; Valencia *et al*, 2009; Martin *et al*, 2010). In particular, constitutive upregulation of WNT target gene, *cyclin D1*, has been shown to result from methylation of *WNT5A* (Martin *et al*, 2010). Epigenetic downregulation of WNT antagonists, such as DKK1 and WIF1 by hypermethylation has also been implicated in chronic lymphocytic leukaemia (CLL) (Bennett *et al*, 2010) and progressive neoplasia of the lung (Licchesi *et al*, 2008).

Besides dysregulation of WNT stimuli, the accumulation of β-catenin may also be the result of other factors such as the mutation of either APC (Korinek *et al*, 1997; Rosin-Arbesfeld *et al*, 2000), conductin (Behrens *et al*, 1998) or β-catenin itself (Morin *et al*, 1997 ; Sparks *et al*, 1998). WNT signalling is known to be crucial for intestinal stem cell activation (Korinek *et al*, 1998a; He *et al*, 2004) and mutations in APC are a cause of colorectal cancer (van Es *et al*, 2001), medulloblastoma and ovarian cancer. In colorectal cancer progression, besides loss of APC, the interaction of APC with the cytoskeleton might also contribute to cancer initiation and progression (Nathke, 2006).

1.4 T-CELL FACTOR (TCF) GENES

1.4.1 Overview of the TCF family

Members of the TCF family of high mobility group (HMG) DNA-binding proteins are responsible for transcriptional regulation of WNT target genes upon β-catenin stabilisation and entry into the nucleus. Humans have four types of TCF proteins namely Transcription factor 7 (TCF7), Transcription factor 7 like 1 (TCF7L1), Transcription factor 7 like 2 (TCF7L2), and LEF1 (Hurlstone & Clevers, 2002), as illustrated in Figure 1.8. TCF7, TCF7L1 and TCF7L2 were previously known as T cell factor (TCF), TCF1, TCF3 and TCF4 respectively. TCF proteins are thought to be bimodal regulators of WNT signalling that function mainly in the nucleus to activate WNT target genes (e.g. *cyclin D1, SP1 CD44, c-MYC*) via β-catenin transactivation. As shown in Figure 1.9, in the absence of WNT signals, TCF acts with corepressors to keep WNT target genes silenced. β-catenin binding to TCF protein in the nucleus antagonises this repression and recruits additional coactivators, inducing target gene expression. Thus, TCF is considered as a transcriptional switch that requires β-catenin to convert its repression into activation (Willert & Jones, 2006). This mechanism could occur through sequence-specific DNA binding on TCFs and also a low-affinity binding site on LEF1 (Daniels & Weis, 2005).

The TCF family including LEF1 are known to produce isoforms through alternative splicing generating proteins with similar basic domain structures (β-catenin-binding, HMG and nuclear localisation signals (NLS) domains) (Hoppler & Kavanagh, 2007). An internal exon in the context-dependent regulatory domain (CRD) is alternative in all members except for TCF7L1, and the exon is flanked by small amino acid motifs (LVPQ, SxxSS; only in TCF7L1 and TCF7L2) also created by alternative splicing, and part of the CRD is alternatively spliced in LEF1 (Pukrop *et al*, 2001). Alternative splicing also creates different C-termini; the CRARF domain is a 33 amino acid motif in the C-tail-E conserved among all non-vertebrates but not present in vertebrates LEF1 and TCF7L1. The C-tail-E also contains two C-terminal binding protein (CtBP)-binding motifs found in TCF7L1 and TCF7L2 (Arce *et al*, 2006; Atcha *et al*, 2007).

TCF7L1 appears to have unique functions in early embryonic development (Merrill *et al*, 2004). Within the TCF family, TCF7 has been reported to be a feedback repressor of WNT signalling (Roose *et al*, 1999). TCF7 and LEF1 produce full-length isoforms that

mediate WNT signals and their truncated dominant negative (dn) isoforms actually limit WNT signals and may function as growth suppressors (Waterman, 2002). Expression of TCF7 is most abundantly expressed in T-lymphocytes, but is also expressed in HSC (Novershtern *et al*, 2011) and NK cell development (Held *et al*, 2003). Similar to TCF7L2 and LEF1, TCF7 has been previously studied to be a partner with nuclear β-catenin, which serves as a downstream transcriptional activator in response to external WNT stimulation. All TCFs can bind to plakoglobin (aka γ-catenin) a homolog of β-catenin, and phosphorylation of serine60 in TCF7L2, but not in other TCFs, inhibits plakoglobin binding (Castano *et al*, 2002). LEF1 and TCF7L1 are both expressed during hair and skin development, however only LEF1 knockout causes loss of hair development indicating dominant roles in different cell types (Reya & Clevers, 2005). In relation to TCF7L2, LEF1 is preferentially expressed by differentiated proliferative phenotype cells in melanoma whereas TCF7L2 is preferentially expressed by invasive phenotype cells, this disease showing differential behaviour of TCF / LEF (Eichhoff *et al*, 2011).

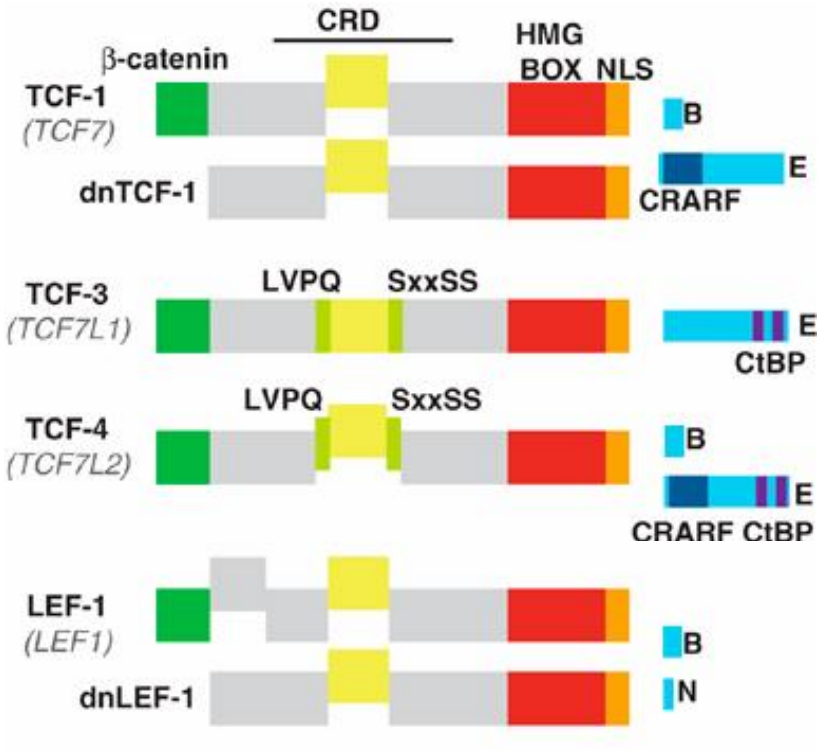


Figure 1.8 : Diversity of splicing variants and alternative promoters generated by TCFs / LEF. The β-catenin binding domain (green) is absent in dominant negative isoforms (dnTCF7, dnLEF1). A context-dependent regulatory domain (CRD; grey/yellow) separates the β-catenin-binding domain from the HMG and NLS. HUGO nomenclature for the mammalian genes is indicated in parentheses (Arce *et al*, 2006).

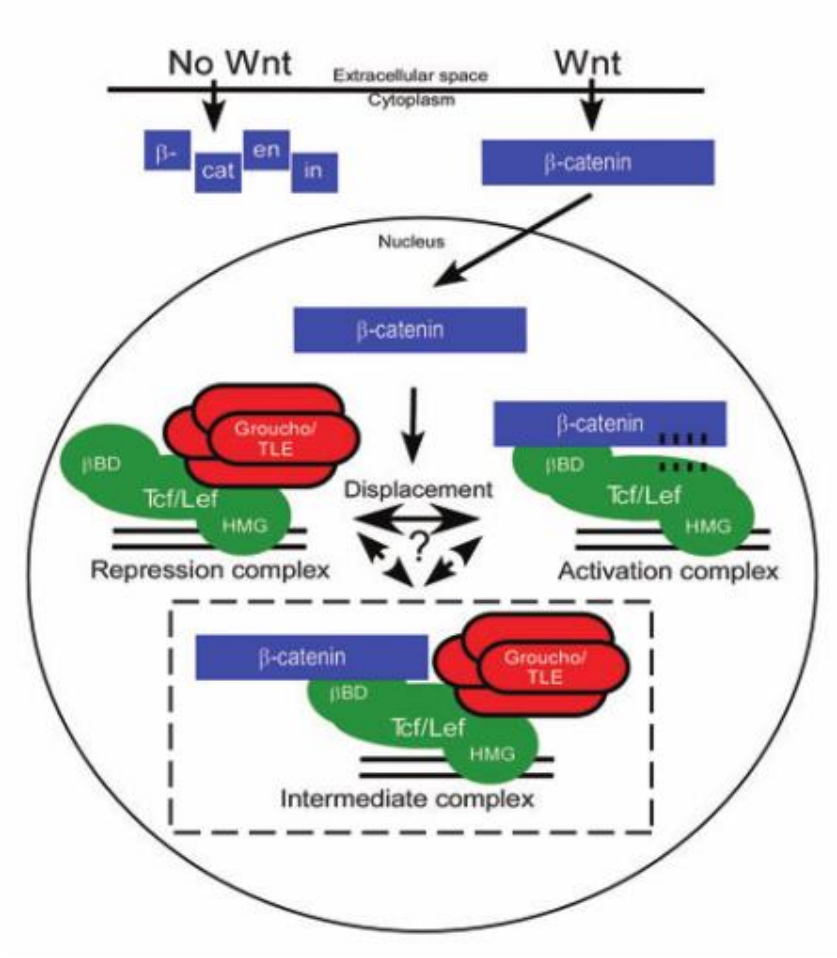


Figure 1.9: Transcriptional activation by β-catenin in the nucleus. Adapted from (Daniels & Weis, 2005).

1.4.2 Structure and function of TCF7L2

The *TCF7L2* is a complex gene of which several different isoforms have been reported, including isoforms that lack the β-catenin binding domain (Korinek *et al*, 1997). *TCF7L2* gene consists of 18 identified exons encoding different functional domains (illustrated in Figure 1.10). This protein is known to display a complex pattern of splicing producing several alternative exons. Tissue-specific splicing giving rise to several *TCF7L2* variants has been detected in pancreatic islets (Prokunina-Olsson *et al*, 2009), vertebrate central nervous system (Nazwar *et al*, 2009), colorectal cancers (Duval *et al*, 2000; Cuilliere-Dartigues *et al*, 2006) and hepatocarcinoma (Tsedensodnom *et al*, 2011).

Frameshift mutations causing microsatellite genetic instability have also been reported in leukaemia models (Chang *et al*, 2006) and in several solid tumours (Duval *et al*, 1999).

TCF7L2 isoforms provide a DNA binding moiety through the HMG domain, and β-catenin contributes the transactivation domain, within a TF complex. Both the β-catenin binding domain and HMG are among important domains that exist within coding regions of all TCF7L2 isoforms used in this current study. The N-terminus region of TCF7L2 is crucial for β-catenin binding, whereas for β-catenin, its core armadillo repeat region interacts, not only with TCF7L2, but also with other TCFs (TCF7, TCF7L1), cadherins and APC (Poy *et al*, 2001; Young *et al*, 2002). The middle region of the TCF7L2 transcript (exon 11- exon 12) is where the HMG DNA binding domain is located. The ability of TCF7L2 to bind DNA can be inhibited by other proteins such as MAD2B (Hong *et al*, 2009). Pioneering work by Korinek *et al* has demonstrated blockage of WNT signalling caused by a dnTCF7L2 that lacks exon 1 sequence (Korinek *et al*, 1997), and upon WNT stimulation, this dominant-negative form is unable to bind β-catenin. This illustrates the importance of TCF7L2 interaction with β-catenin in the nucleus. For transcriptional regulation, TCF7L2-dependent gene responses to WNT signals are primarily mediated by a Cysteine-clamp (C-clamp) containing TCF7L2 with C-tail type E splice variants. TCF7L2 isoforms with C-tail type B are much less potent in transactivation compared with C-tail type E (Wallmen *et al*, 2012). These data suggest that the diversity of TCF7L2 isoforms may exist to enable fine control of WNT target gene expression.

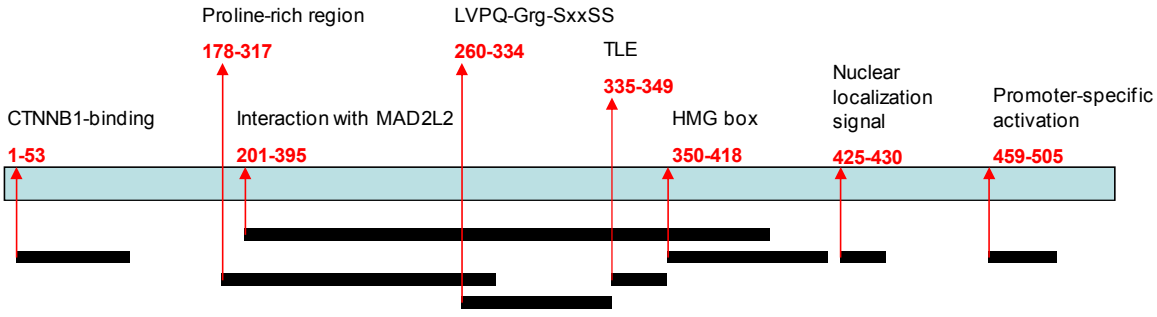


Figure 1.10: Structural view of the full length TCF7L2 protein and the corresponding functional domains. Protein sequence annotation for TCF7L2 (Entry # Q9NQB0) was adapted from The UniProt Consortium accessible from http://www.uniprot.org/uniprot/Q9NQB0 (Magrane & Consortium, 2011).

1.4.3 Roles of TCF7L2 in disease and cancer development

Single nucleotide polymorphisms (SNPs) of *TCF7L2* have been widely associated with risk of Type 2 diabetes (Grant, 2012; Boj *et al*, 2012; Greenawalt *et al*, 2012; Peng *et al*, 2013). Emerging reports also show that *TCF7L2* polymorphisms are associated with an increased risk of gestational diabetes mellitus (Zhang *et al*, 2013), diabetic retinopathy (Luo *et al*, 2013), breast cancer (Connor *et al*, 2012) and colon cancer recurrence (Paez *et al*, 2013). Aberrant activation of TCF7L2 by the accumulation of β-catenin in intestinal crypts has been implicated in colorectal cancer (van der Flier *et al*, 2007), and mice models lacking TCF7L2 show loss of the proliferative compartment in the crypt region (Korinek *et al*, 1997; Korinek *et al* 1998a), indicating that constitutive activation of TCF7L2 is important for maintaining the malignant phenotype. Activation of the TCF7L2 / β-catenin transcriptional complex in colorectal disease arises mainly due to activating point mutations in β-catenin (Morin *et al*, 1997). It has been reported in renal cell carcinoma that imbalances in TCF7L2 splicing are associated with inhibition of apoptosis (Shiina *et al*, 2003). TCF7L2 chromatin occupancy sites have also been validated in colon carcinoma cells, breast cancer cells and glioma cells by chromatin immunoprecipitation (ChIP) coupled to PCR, demonstrating direct transcriptional regulation of miRNA by TCF7L2 in epithelial cells and roles in cancer cell proliferation and invasion (Lan *et al*, 2012). The precise role of TCF7L2 and TCF / LEF family members in the maintenance of cell-type-specific WNT signalling responses is still unknown in the context of AML and haematopoiesis.

1.5 GENE EXPRESSION STUDIES IN AML

Gene expression profiling (GEP) has been incorporated into many studies given its versatility in high throughput identification of disease classification signatures and the elucidation of new biomarkers for clinical development. Microarray testing of differential gene expression signatures in leukaemia subtypes has demonstrated that this approach is not only useful to confirm diagnosis (e.g. classification of molecular signatures; see below) but may detect novel associations between aberrant gene expression and gene mutations that affect the same biological processes. The genetic variations of the different AML subtypes

often lead to distinct changes in gene expression across more than 30,000 genes, which is comprehensively analysed by DNA microarrays provided by several manufacturers including Affymetrix. This study utilised GEP data generated using the Affymetrix GeneChip® Human platform shown in Table 1.4.

Previous multi-center studies have identified minimal numbers of genes needed to identify prognostically important clusters with a high degree of accuracy. Gene expression profiles of AML patients were used to delineate differential gene expression signatures corresponding to *NPM1* gene mutation (Kohlmann *et al*, 2010). GEP studies showed that analysis with whole-genome DNA microarrays leads to a prediction accuracy of 96% with respect to the classical methods (eg: flow cytometry), and allows further distinction of AML subtypes. Molecular features unique for *NPM1* (Verhaak *et al*, 2009), *C/EBP*α (Valk *et al*, 2004; Wouters *et al*, 2007) and distinct subtypes AML M2 with t(8;21), AML M3 with t(15;17), or AML M4 with inv(16) have been reported (Schoch *et al*, 2002). Unique molecular and prognostic features have been demonstrated by GEP studies in AML with t(8;16) and are able to discriminate this subtype from AML with t(11q23) (Haferlach *et al*, 2009).

In acute leukaemias, GEP studies not only provide molecular distinction of subtypes within AML (Kohlmann *et al*, 2003; Schoch *et al*, 2005) and ALL (Yeoh *et al*, 2002), but also able to distinguished ALL from AML (Kohlmann *et al*, 2003) and this has become a tool to further characterise MDS from AML patients (Mills *et al*, 2009; Theilgaard-Monch *et al*, 2011). Clinically-relevant gene changes associated with the granulocytic lineage due to aberrant transcriptional activity of RUNX1-RUNX1T1 were identified by GEP studies (Tonks *et al*, 2007b) and have previously been validated in a primary AML cohort (Tonks *et al*, 2007a; Morgan *et al*, 2013). Taken together, it is evident that GEP studies carry high potential for future integrative studies with other genomic technologies, which will continue to improve our understanding of malignant transformation particularly in myeloid malignancies and thereby contribute to individualised risk-adapted treatment strategies in AML.

Description	Human Genome U133 Plus 2.0 array	Human Genome U133A array
Number of transcripts	~ 47,400	~18,400
Number of genes	>38,500	>14,500
Number of probe sets	>54,000	>22,000
Oligonucleotide probe length	25-mer	25-mer
Probe pairs / sequence	11	11
Control sequences included:		
Hybridisation controls	*bioB, bioC, bioD, cre*	*bioB, bioC, bioD, cre*
Poly-A controls	*dap, lys, phe, thr*	*dap, lys, phe, thr*
Normalisation control set	100 probe sets	100 probe sets
Housekeeping / control genes	*GAPDH, β-Actin, ISGF-3 (STAT1)*	*GAPDH, β-Actin, ISGF-3 (STAT1)*

Table 1.4: Specifications for Affymetrix GeneChip® Human Genome U133 cartridge format. Affymetrix, the HG-U133 Plus2.0 set contains more than 50,000 different oligonucleotide probe sets representing ~48,000 transcripts on the GeneChip. Both arrays also contain 100 probe sets for normalisation to control and housekeeping genes. These controls were selected to reflect a variety of different expression levels and so can be used as quality control to determine the sensitivity of each hybridization reaction. Adapted from www.affymetrix.com/analysis/index.affx.

1.6 AIMS OF THE STUDY

The initial aim of this study was to identify dysregulated pathways in AML. Subsequent aims focussed on understanding of the contribution of WNT signalling to the pathogenesis of AML.

Specific aims were:-

1) To identify dysregulated pathways and the corresponding aberrantly regulated genes in primary AML samples. This was achieved by comparing GEP of AML blasts to normal haematopoietic progenitor cells (HPC) using a comprehensive genome wide microarray analysis.

2) To characterise the selected dysregulated gene (i.e. *TCF7L2*) involved in WNT signalling at the transcript and protein level in AML patients and normal samples.

3) To determine the functional significance of TCF7L2 using myeloid leukaemia lines and normal human haematopoietic progenitor cells.

2 - General Materials and Methods

2.1 CHEMICALS, REAGENTS AND SUPPLIERS

All chemicals used in this study were Analar grade or tissue culture (TC) tested unless otherwise stated. Chemicals used are listed below grouped by supplier in an alphabetical order:

Reagents	Source
Foetal calf serum (FCS)	BioSera Labtech, Ringmer, UK
Nuclear / Cytosol Fractionation kit	BioVision, California, USA
DNA primers and synthetic genes were synthesized *de novo* as ordered, and subcloned into a standard delivery vector (pBS II SK(+) or pEX-A). *TCF7L2* cDNA sequences were submitted online to http://www.eurofinsgenomics.eu	Eurofins MWG Operon, Ebersberg, Germany
Ultra-purified water was obtained from PureLab® Ultra water purification unit	Elga LabWater, High Wycombe, UK
Ethanol, hydrochloric acid (HCl), methanol (CH_3OH) and sodium chloride (NaCl)	Fisher Scientific, Loughborough, UK
Sterile water for irrigation	Fresenius Kabi, Cheshire, UK
Ficoll-Paque PLUS®, Amersham ECL advance western blot detection kit, ECL advances blocking agent	GE Healthcare, Buckinghamshire, UK
Hank's Balanced Salt Solution (HBSS), penicillin-streptomycin	Gibco®, Paisley, UK
Sterile water TC grade	Hameln Pharmaceuticals, Gloucester, UK

Dulbecco's Phosphate Buffered Saline (dPBS), human IL-3, human IL-6, human G-CSF, human GM-CSF, human SCF, L-glutamine, NuPAGE® 3-(N-morpholino)propanesulfonic acid (MOPS)-SDS Running Buffer, NuPAGE® Transfer Buffer, OneShot® Stbl3 and OneShot®Top10 chemically competent *E. coli*, sodium bicarbonate ($NaHCO_3$) 7.5%, Super-Optimal broth with Catabolite repression (SOC) medium, Trizol®, Trypsin-EDTA 0.05%, Tris (tris(hydroxymethyl)aminomethane) ultra pure	Invitrogen™, Paisley, UK
SeaKem® GTG™ Agarose	Lonza, Basel, Switzerland
Benzonase 25 U/µl	Novagen®, Nottingham, UK
Human Fms-like tyrosine kinase-3 ligand (FLT3L) and M-CSF	Peprotech, London, UK
Recombinant human WNT-3A	R&D Systems, Abingdon, UK
Human transferrin	Roche Diagnostics, Burgess Hill, UK
Gentamycin (Cidomycin®)	Manufactured by Sanofi-Aventis, Guildford, UK and supplied by UHW Pharmacy
7-Aminoactinomycin D (7-AAD), ampicillin, bovine serum albumin (BSA) fraction V, Bradford reagent, calcium chloride ($CaCl_2$), chloroform molecular grade, chloroquine (N'-(7-chloroquinolin-4-yl)-N,N-diethyl-pentane-1,4-diamine), deoxyribonuclease I (DNAse I), dimethyl sulfoxide (DMSO), Dulbecco's Modified Eagle's Medium (DMEM), ethanol molecular grade, ethidium bromide (EtBr), ethylenediamine-tetraacetic acid (EDTA), glycerol, isopropanol, 4-	Sigma-Aldrich®, Poole, UK

(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid, N-(2-Hydroxyethyl) piperazine-N′-(2-ethanesulfonic acid) HEPES buffered saline (HeBs), Iscove's Modified Dulbecco's Medium (IMDM), Luria Bertani (LB) agar and LB broth, magnesium chloride ($MgCl_2$), magnesium sulphate ($MgSO_4$), 2-mercaptoethanol (2-ME), MISSION® shRNA, phosphatase inhibitor cocktail, Ponceau S solution, protease inhibitor cocktail, protein standard (1 mg bovine serum albumin /ml in 0.15 M NaCl), puromycin, Roswell Park Memorial Institute 1640 culture medium (RPMI-1640), sodium azide (NaN_3), sodium orthovanadate (Na_3VO_4), sodium dodecyl sulfate (SDS), Triethylammonium bicarbonate buffer 1.0 M (TEAB), Tris-Borate-EDTA (TBE) buffer 10X concentrate, Tris-Acetate-EDTA (TAE) buffer 10X concentrate, Triton X-100, Tween-20, xylene cyanol	
RetroNectin®	Takara, Shiga, Japan
(2'Z,3'E)-6-Bromoindirubin-3'-oxime (BIO)	Tocris, Bristol, UK

2.2 BUFFER AND SOLUTION COMPOSITIONS

The compositions of the most frequently used buffers are listed as below:

Buffer	Recipe
DNA gel loading buffer 6X	0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol and 30% (v/v) glycerol in distilled water
ECL Advance western blotting detection	Solution A containing Tris buffer in 3.2% (v/v) ethanol and Solution B containing proprietary substrate in Tris buffer
Flow cytometry staining buffer	1X PBS containing 1% (v/v) BSA, 3.1 mM NaN_3 in TC grade water
Freezing mix	30% (v/v) FCS, 20% (v/v) DMSO, 50% (v/v) growth media and 0.45 µM filtered
HBSS	25 mM HEPES, 100 µg/ml gentamycin and 1 U/ml heparin
HeBS 2X	50 mM HEPES, 280 mM NaCl, 1.5 mM Na_2HPO_4, pH 7.05
IMDM	1% (w/v) BSA fraction V, 20% (v/v) FCS, 45 µM 2-ME, 360 µg/ml 30% human transferrin, 100 U/ml penicillin, 100 µg/ml streptomycin
LB-broth	10 g Bacto-Tryptone, 5 g Bacto-yeast extract, 10 g NaCl dissolved in I L distilled water, pH 7.5
Magnetic activated cell sorting (MACS) buffer	1X PBS, 0.5% (v/v) BSA, 2 mM EDTA, 5 mM $MgCl_2$, pH 7.2, 0.45 µm filtered

MOPS SDS running buffer 20X	50 mM MOPS, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.7
Ponceau S solution	1.3 mM (0.1% w/v) Ponceau S in 5% (v/v) glacial acetic acid
SOC medium	2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 10 mM magnesium chloride, 10 mM magnesium sulfate, 20 mM glucose
TAE buffer 10X	400 mM Tris-acetate and 10 mM EDTA, pH 8.3
TBE buffer 10X	890 mM Tris, 890 mM boric acid, 20 mM EDTA, pH 8.3
Tris buffered saline (TBS)	20 mM Tris HCl, 135 mM NaCl, pH 7.6
TBS Tween-20 (TBS-T)	TBS with 0.1% (v/v) Tween-20
Tris-EDTA (TE)	10 mM Tris-HCl, 1 mM EDTA, pH 7.5
Transfer buffer	25 mM Bicine, 25 mM Bis-Tris, 1 mM EDTA, pH7.2

2.3 GENERAL CELL CULTURE

2.3.1 Subculture of suspension cell lines

Each cell line was processed and cultured according to its own biohazard risk and standard cell culture technique. Separate bottles of reagents and growth media were used for each type of cell line to minimise cross-contamination risk. Each cell line was handled separately under containment level II and the flow cabinet was decontaminated thoroughly by wiping the internal surfaces using 70% (v/v) ethanol. Cell culture growth medium and solutions were pre-warmed to room temperature (RT) prior to subculturing. The culture was first examined for signs of high density growth, infection and medium colour change. A sample of cells was removed and diluted 1 in 10 for counting in appropriate growth medium. The media used to culture each cell line are listed in Table 2.1. In order to maintain the growth of the suspension cells, excess cells were removed directly from the culture flask by pipetting them off into a waste bottle containing 1% (v/v) chlorine and new fresh growth medium was added. All cultures were incubated at 37°C with 5% CO_2 humidified atmosphere for 16-24 hours.

2.3.2 Subculture of adherent cell lines

The cultures were examined microscopically for their level of confluence before subculturing. The existing growth medium was decanted and 1-2 ml / 25 cm^2 trypsin-EDTA was added directly to the flask followed by tilting (to cover the adherent layer) for about 3-5 minutes at RT. Subsequently, an equal volume of growth medium was added to neutralise the trypsin activity and the cells were transferred into a fresh tube. If required, cells were counted before centrifugation for 10 minutes at 200 x g. Supernatant was discarded and the cell pellet was resuspended with fresh growth medium.

2.3.3 Cell quantification

Estimation of cell number was determined using haemocytometer counting chambers with improved Neubauer ruling (Hawksley, Brighton, UK). Briefly, 8 µl of cell suspension was aliquoted from culture and directly pipetted under a counting chamber cover slip. The cellularity of the sample in cells / ml was given by the average number of cells per cuboid multiplied by 1×10^4, as counted using Eclipse TS100 light transmission microscope (Nikon, Surrey, UK).

2.3.4 Cryopreservation and resuscitation of cells.

To freeze cells for continued culture at a later date, between $1 – 10 \times 10^6$ cells were collected by centrifugation at 200 x g for 5 minutes and resuspended in the relevant growth medium into 1.8 ml cryopreservation vials (Nunc). An equal volume of freezing mix (50% growth medium (RPMI or DMEM), 30% (v/v) FCS, 20% (v/v) DMSO was added drop-wise to the cell suspension and the tube was immediately placed inside a Mr. Frosty freezing container that have been filled with 70% (v/v) isopropanol. The freezing container was placed at –80°C overnight and the tubes were transferred to liquid nitrogen (LN_2) for long-term storage.

Prototype/Name of cell line	Cell phenotype	Seeding density cells / ml	Source reference	Growth medium
K562	Erythrocytic cell derived from chronic myeloid leukaemia	1×10^5	(Lozzio & Lozzio, 1975)	RPMI-1640, 10% (v/v) FCS, 2mM L-glutamine, 20 µg/ml gentamycin
THP-1	Monocytic cell derived from AML	1×10^5	(Tsuchiya *et al*, 1982)	RPMI-1640, 10% (v/v) FCS, 2mM L-glutamine, 20 µg/ml gentamycin
HL-60	Promyelocytic cell derived from AML	1×10^5	(Collins *et al*, 1978)	RPMI-1640, 10% (v/v) FCS, 2mM L-glutamine, 20 µg/ml gentamycin
NB-4	Promyelocytic cell derived from AML t(15;17)	1×10^5	(Lanotte *et al*, 1991)	RPMI-1640, 10% (v/v) FCS, 2mM L-glutamine, 20 µg/ml gentamycin
U937	Monocytic cell derived from histiocytic lymphoma	1×10^5	(Sundstrom & Nilsson, 1976)	RPMI-1640, 10% (v/v) FCS, 2mM L-glutamine, 20 µg/ml gentamycin
Kasumi-1	Myeloblastic cell derived from AML t(8,21)	3×10^5	(Asou *et al*, 1991)	RPMI-1640, 10% (v/v) FCS, 2mM L-glutamine, 20 µg/ml gentamycin
MV4-11	Biphenotypic B-myelomonocytic leukaemia	1×10^5	(Lange *et al*, 1987)	RPMI-1640, 10% (v/v) FCS, 2mM L-glutamine, 20 µg/ml gentamycin
Phoenix & HEK293T	Amphotropic virus-packaging cell line derived from Human embryonic kidney (HEK293T)	2.5×10^6	(Kinsella & Nolan, 1996)	DMEM containing 10% (v/v) FCS, 20 µg/ml gentamicin, 2 mM L-glutamine
HeLa	Epithelial cell derived from cervical cancer	2.5×10^6	(Masters, 2002)	DMEM containing 10% (v/v) FCS, 20 µg/ml gentamicin, 2 mM L-glutamine

Table 2.1: Cell lines used in this study and their growth condition requirements.

2.4 ISOLATION OF HAEMATOPOIETIC MONONUCLEAR AND PROGENITOR CELLS

2.4.1 Patient samples

The samples taken for this study were approved by the local and multicentre ethical committees in the UK. *De novo* disease (AML presentation BM / PB samples) were collected with informed ethical consent for research purposes in accordance with the Declaration of Helsinki.

2.4.2 Ficoll density gradient

For the isolation of mononuclear cells from CB or BM, density centrifugation over FicollPaque® solution was used. The CB or BM samples were diluted in 1:1 ratio with HBSS (see 2.2). The diluted sample (8 ml) was then gently layered over FicollPaque® (5ml) in a 50 ml falcon tube before centrifuging at 400 x g for 40 minutes with slow acceleration. Differential migration during centrifugation results in the formation of layers containing different cell types. The bottom layer contained erythrocytes which have been aggregated by the FicollPaque® solution. The mononuclear leukocytes were found at the interface between the plasma and the FicollPaque® solution because of their lower density. The cell layer was aspirated by gentle aspiration in circular motions to ensure collection of all cells at the edge of tube and was transferred into a new conical tube containing 15 ml of growth medium (RPMI-1640, 10% FCS, 2mM L-glutamine, 20 μg/ml gentamycin) and centrifuged at 200 x g for 10 minutes. Additional washing was performed in the same way until the supernatant was free of platelet contamination as determined by microscopic inspection. Mononuclear cells isolated were counted before resuspending in growth medium (RPMI-1640, 10% FCS, 2mM L-glutamine, 20 μg/ml gentamycin) at 50 x 10^6 cells per vial for cryopreservation.

2.4.3 Immunoselection of normal human haematopoietic CD34$^+$ progenitor cells

Normal human CD34$^+$ HPCs were isolated and purified from the total mononuclear cells by using a magnetic-activated cell separation (miniMACS™) kit (Miltenyi Biotec, Bisley, UK). Freshly isolated cells or thawed cells from LN_2 were resuspended in 150 µl MACS buffer per 1 x 10^8 mononuclear cells. Cells were incubated at 4˚C for 15 minutes with 50 µl hapten-conjugated monoclonal CD34 antibody (clone QBEND/10) per 10^8 cells in the presence of FcR blocking agent. Cells were then washed with 5 ml of MACS buffer to stop the reaction and pelleted at 200 x g for 5 minutes. Washed cells were resuspended in MACS buffer and incubated at 4°C with the addition of 50 µl of anti-hapten microbeads (per 10^8 cells) for 15 minutes. In order to isolate the magnetically-labelled CD34 cells, mononuclear cells were resuspended in MACS buffer and applied to a magnetised column. The magnetically-labelled CD34$^+$ cells were eluted with 1 ml MACS buffer and flowed through a second column to maximise enrichment. To assess the CD34 purity, 1 x 10^4 cells were resuspended in flow cytometry staining buffer combined with 2.5 µg/ml R-phycoerythrin conjugated anti-human CD34 monoclonal antibody (Clone 8G12, BD Biosciences). After incubation for 30 minutes at 4˚C, cells were washed using 1 ml of flow cytometry staining buffer and pelleted by centrifugation at 200 x g for 5 minutes and analysed by flow cytometry as described in section 2.9.

2.5 NUCLEIC ACID ANALYSIS

2.5.1 Isolation of total RNA using Trizol

This method utilised a monophasic solution of phenol and guanidine isothiocyanate which is a Trizol reagent with Phase Lock Gel-Heavy (2 ml) tubes. Between 1-2 x 10^6 cells were transferred into a sterile tube and washed with 1X PBS buffer and pelleted by centrifugation before adding 1 ml of Trizol reagent (Sigma). The Phase Lock Gel-Heavy (2 ml) tubes were centrifuged briefly to collect gel on the tube bottoms by centrifugation at 1500 x g, 30 seconds. The Trizol cell mixture was then transferred into the pre-spun Phase Lock Gel-Heavy tubes and incubated for 5 minutes at RT. 200 µl of chloroform (1 ml Trizol : 200 µl chloroform) was added into each microfuge tube. In a vigorous manner, the mixture of Trizol suspension and chloroform was mixed for 15 seconds and incubated at RT for 2 minutes. Tubes were centrifuged at 12 000 x g at 4°C for 10 minutes using a

refrigerated tabletop centrifuge. After centrifugation, two phases were obtained and care was taken not to disturb the interface layer which contained DNA. During this step, RNA was separated into the upper aqueous phase of the supernatant and protein was separated into the lower phase. The upper aqueous layer was carefully transferred into a new 2 ml RNase free microtube before adding 600 µl of isopropanol. The tube was gently inverted to allow adequate mixing. The mixture was incubated at RT for 10 minutes and centrifuged at 12 000 x g, 4°C for 10 minutes. In order to precipitate RNA, 900 µl of 80% (v/v) ice cold ethanol was added and gently mixed. Next, the tube was centrifuged again but at lower speed which is at 7500 x g, 4°C for 5 minutes. Then, all the ethanol solution was removed by pipetting. This step was done carefully to prevent loss of the RNA pellet present in the tube. The tube was air dried on ice for 30 - 45 minutes. Finally, the RNA pellet was eluted in an appropriate amount of RNase-free water (20 µl – 35 µl) and treated with DNase I (refer 2.5.2). The RNA sample was stored in -20°C for immediate use or at -80°C for long term storage until further use.

2.5.2 Total RNA clean-up by DNase I digestion

RNA isolated above (2.5.1) was subjected to further DNAse treatment. The volume of RNA sample was adjusted to 100 µl by using sterile water. 350 µl Buffer RLT was added to the sample and thoroughly mixed before adding 250 µl absolute ethanol to the lysate. 700 µl of sample mix was applied to an RNeasy mini spin column (Qiagen®, West Sussex, UK) placed in a 2 ml collection microtube and the column was incubated for 5 minutes. The tube was then centrifuged for 15 seconds at 10 000 x g. The sample was reloaded again on the RNeasy column a second time to increase binding of RNA to the RNeasy membrane and centrifuged using the same method. The RNeasy column was transferred into a new 2 ml collection tube. 350 µl of Buffer RW1 was pipetted onto the column and incubated for 1 minute before centrifuging at 10 000 xg for 15 seconds. The flow through was discarded and inside a separate fresh tube, 10 µl DNase I (Qiagen®) stock solution was mixed with 70 µl Buffer RDD. Subsequently, 80 µl DNase I mix was pipetted directly onto the column and incubated at RT for 15 minutes. Following that, 350 µl of Buffer RW1 was pipetted onto the same column and centrifuged at 10 000 x g for 15 seconds. The flow through was discarded and the wash was repeated twice using 350 µl Buffer RW1. After the final wash, 500 µl Buffer RPE was pipetted onto RNeasy column and incubated for 5 minutes and centrifuged at 10 000 x g for 15 seconds. Flow through

was discarded and 500 µl RPE buffer was added onto the RNeasy column and centrifuged for 2 minutes at maximum speed to dry the RNeasy membrane. The RNeasy column was placed in a new 2 ml collection tube with the lid opened for 5 to 10 minutes to completely evaporate the ethanol. Finally, 40 µl of sterile water was pipetted directly onto the center of the RNeasy membrane and incubated for 5 minutes before centrifuging for 1 minute at 10 000 x g to elute the RNA. This step was repeated a second time by pipetting the flow through back into the same collection tube (~36 µl) onto the RNeasy membrane. After incubation for 2 minutes, tube was centrifuged at 10 000 x g for 1 minute. The eluate containing DNase treated RNA was kept at -80°C.

2.5.3 Reverse-transcriptase Polymerase Chain Reaction (RT-PCR)

For complementary DNA (cDNA) synthesis by RT-PCR, 500 ng of total RNA was reversed transcribed in a 20 µl reaction mix containing 2.5 U of MuLV RTase, 2.5 µM random hexamers, 1 U of RNAse inhibitor, 4 mM dNTP mix, 1X PCR buffer II and 5 mM $MgCl_2$. All reagents were purchased from GeneAmp® RNA PCR Kit (Applied Biosystems®, Paisley, UK). The RT reaction was carried out in an Applied BioSystem thermal cycler (Applied Biosystems) using the following cycling parameters; 10 minutes at 25°C followed by 30 minutes at 42°C and 5 minutes at 95°C. RT products were stored at -20°C.

2.5.4 Quantitative RT-PCR (qRT-PCR)

For each sample, qRT-PCR was performed in duplicate in a 10 µl reaction containing 1 µl cDNA, 1µl FastStart DNA SYBR Green I master mix (Roche), 4 mM $MgCl_2$, 0.5 µM forward and reverse *TCF7L2* and *ABL* (Abelson murine leukaemia viral oncogene homolog 1) primers (Table 2.2). The 45-cycle qRT-PCR programme was performed using a LightCycler 2.0 (Roche Diagnostics, Burgess Hill, UK). Amplification was performed using the following cycling parameters; pre-incubation at 95ºC for 10 minutes, followed by 40 cycles of [95°C for 3 seconds, 60°C for 5 seconds, 72°C for 12 seconds], then 1 cycle of melting curve step of [95°C for 0 second, 65ºC for 15 seconds and 95ºC for 0 second] and a final 1 cycle of cooling at 40ºC for 30 seconds.

Name	Orientation	Exon	Sequence (5' to 3')	Tm (ºC)
TCF7L2-exon 7/8	Forward	7/8	CGTAGACCCCAAAACAGGAA	58.4
TCF7L2-exon 9	Reverse	9	TCCTGTCGTGATTGGGTACA	58.4
TCF7L2-exon 18	Forward	18	TGCGTTCGCTACATACAAGG	57.3
TCF7L2-exon 18	Reverse	18	TGGGTCTGCTCAGTCTGTGA	59.4
ABL-A2N	Forward	2	CCCAACCTTTTCGTTGCACTGT	60.3
ABL-NA4	Reverse	4	CGGCTCTCGGAGGAGACGTAGA	65.8

Table 2.2: Primer pairs used for qRT-PCR analysis in this study.

2.5.5 PCR product purification using QIAquick PCR Purification Columns

The QIAquick silica membrane (Qiagen®) was used to purify DNA from aqueous solutions (PCR sample), and up to 10 µg DNA can bind to each QIAquick column. For the sample mix, 5 volumes of Buffer PB (500 µl) were added to 1 volume (100µl) of each PCR sample. QIAquick spin column was placed in a provided 2 ml collection tube and the sample mixture was applied to the QIAquick column and centrifuged for 30 – 60 seconds at 10 000 x g to bind DNA. Flow-through was discarded and the same QIAquick column was placed back into the same tube. To wash, 750 µl Buffer PE was added to the QIAquick column and centrifuged for an additional 30 – 60 seconds. Flow-through was discarded and the QIAquick column was placed back in the same tube and centrifuged for another minute to remove the residual ethanol. The QIAquick column was then placed in a clean 1.5 ml microcentrifuge tube. To elute DNA, 50 µl Buffer EB (10 mM Tris HCl, pH 8.5) was added to the center of the QIAquick membrane, and following 1 minute incubation at RT, the column was centrifuged for 1 minute. All centrifugation steps were performed at 10, 000 x g. The eluate containing purified DNA sample was collected and stored at -20 °C.

2.5.6 Direct sequencing

The plasmid DNA or purified PCR products (2.5.5) were sequenced using the BigDye Terminator v3.1 protocol. The Sequencing Buffer (5X) is supplied at a 5X concentration and was adjusted to a final concentration of 1X in the reaction volume. For each reaction, a final volume of 20 µl was used containing 1X final concentration of the following reagents; Ready Reaction premix (4µl), BigDye Sequencing Buffer (2µl), forward and reverse primer (3.2 pmol), DNA template (50 ng – 500 ng) and sterile water. The tubes were placed in the thermal cycler and the following temperature cycling protocol was used: 96 ºC for 1 minute followed by 25 cycles of [96°C for 10 seconds, 50°C for 5 seconds, 60°C for 4 minutes], then ramp to 4°C. Subsequently, the extension products (sequencing reactions) were purified by using the DyeEx 2.0 Spin Kit (Qiagen®) according to manufacturer's recommendations prior to BigDye Terminator v3.1 Cycle Sequencing.

2.6 PREPARATION OF RECOMBINANT PLASMID DNA

2.6.1 Restriction enzyme (RE) conditions for creation of TCF7L2 DNA constructs

RE digest for the creation of TCF7L2 expression plasmids (5.3.3.1) was performed in a total reaction volume of 100 µl using 5 – 10 µg of plasmid DNA containing appropriate NEB (New England Biolabs Ltd., Hitchin, UK) buffer, RE enzymes, BSA and water. For *Xba*1 digestion, 100 U enzyme was used with a compatible NEB4 buffer. For *Bgl*II digestion, 100 U enzyme was used with a compatible NEB3 buffer. For *Bam*H1 and *EcoR*1 digestion, 50 U of each enzyme was used with a compatible NEB4 buffer. All reactions were mixed gently by pipetting and incubated at 37°C in a water bath for 1 hour. Aliquot of digested DNA was taken from the digested product and electrophoresed using 0.8% (w/v) agarose gel at 80 V for 45 minutes and visualised using EtBr.

2.6.2 Agarose gel purification of insert DNA

To purify DNA inserts generated by RE or PCR, preparative agarose gel was prepared by creating a 0.8% (w/v) Seakem agarose solution in 1X TAE buffer (or 3% (w/v) agarose for PCR products). DNA samples for loading contained 5 - 10 µg digested DNA (or ~ 1 µg for PCR products), 200 µM bromophenol blue and 120 µM xylene cyanol per lane. Samples were electrophoresed in 1X TAE buffer at 40 V for 2 hours. The size of DNA fragments or plasmids was estimated using a 1 kb DNA Ladder. Upon completion of electrophoresis, agarose gel was stained in EtBr (500 µg/mL) for 20 minutes. Gel was destained using water for 20 minutes and DNA was visualised by long wavelength UV transillumination using a LAS-3000 digital imaging device (Fujifilm UK Ltd., Bedford, UK) before excising the desired DNA band by using a clean scalpel. All apparatus used was initially wiped with 70% (v/v) ethanol. The excised gel was then placed inside a sterile 15 ml tube and subjected to further purification using the QIAquick Gel Extraction kit protocol (Qiagen®).

2.6.3 Ligation

For subcloning, vector and insert DNA fragments were purified according to 2.6.2 prior to ligation. A final volume of 10 µl ligation reaction was used, containing insert DNA combined with vector DNA using a 5:1 ratio of insert:vector ligation, 200 cohesive end units/µl T4 DNA ligase (1 µl) and 1X T4 DNA ligase reaction buffer (1 µl) (New England Biolabs Ltd., Hitchin, UK). In order to determine the level of background ligation due to self-ligation of the linearised vector plasmid, a background reaction tube was also set up as above, but lacking insert DNA. All reaction tubes were incubated for between 4 - 6 hours at 16˚C then stored at -20˚C before performing the transformation into competent *E.coli*.

2.6.4 Transformation of competent cells with plasmid DNA

Each plasmid DNA to be propagated by transformation was thawed on ice alongside one 50 µl vial of OneShot® Stbl3 or OneShot® Top10 chemically competent *E. coli* (Invitrogen™). For each reaction, 6 µl of DNA sample was aseptically transferred directly into the respective vial of competent cells and tapped gently to mix, followed by 30 minute incubation on ice. Following this, the cells were further incubated at 42˚C for 30 seconds (or 45 seconds for OneShot® Stbl3 cells), without agitation, before replacing on ice and

adding 250 µl of pre-warmed SOC medium (Invitrogen™) to each vial to maximise transformation efficiency. All vials were secured in a shaking incubator at 37°C for 1 hour at 225 revolutions per minute (rpm). Once the incubation had elapsed, between 50 µl to 100 µl of transformation mixture was spread onto pre-warmed LB-agar plates containing 100 µg/ml ampicillin (or kanamycin as in Table 5.1), using sterile plastic spreaders to generate a film of bacterial growth, and incubated at 37°C overnight. Transformed colonies were selected on the following day for large scale purification of plasmid DNA as described below.

2.6.5 Large scale purification of plasmid DNA

Plasmid DNA was prepared from an *E.coli* strain using HiSpeed TM Maxi kit (Qiagen®) based on manufacturer's recommendations. Firstly, colonies of interest were picked from LB-agar plates and a single colony was inoculated into 5 mL of LB-broth containing 5µl of ampicillin or kanamycin (stock 100 mg/ml) by using a sterile inoculation loop and incubated in 37°C incubator with 225 rpm shaker for 8 hours. Following that, the starter culture was diluted 1 in 500 into LB-broth with appropriate antibiotic and further incubated in 37°C with 225 rpm shaker for 16 hours. After overnight incubation, the bacterial pellet was harvested from 150 ml culture by centrifugation at 6000 x g for 15 minutes at 4°C. Briefly, pelleted bacteria were resuspended in 10 ml Buffer P1 before mixing with 10ml of Buffer P2, and tube was allowed to stand for 5 minutes at RT. Next, 10 ml of Buffer P3 was added followed by incubation on ice for 15 minutes, and centrifugation at 20,000 x g at 4°C for 30 minutes. Supernatant containing plasmid DNA was removed and centrifuged again, as above, to remove insoluble material. During this process, QIAGENtip-500 was equilibrated by adding 10 ml Buffer QBT onto the tip. The bacterial supernatant was transferred to the equilibrated QIAGEN-tip where plasmid DNA was bound to the column and washed twice with 30ml Buffer QC. Subsequently, DNA was eluted from the column using 15 ml Buffer QF and precipitated by the addition of 10.5 ml isopropanol and centrifugation at 15,000 x g for 30 minutes at 4˚C. The DNA pellet was finally washed using 5 ml of 70% (v/v) ethanol before final centrifugation at 15,000 x g for 10 minutes. Finally, the supernatant was decanted with care, and the purified DNA pellet was air-dried for 10 minutes before resuspending in 1 ml of TE Buffer and concentration of DNA was quantified using the NanoDrop® and stored at -20˚C.

2.7 RETROVIRAL TRANSDUCTION OF HAEMATOPOIETIC CELLS

2.7.1 Production of retroviruses by transfection

The principle of this method is based on the formation of a precipitate containing calcium phosphate and DNA. The procedure used is based on Calcium Phosphate Transfection Kit, Sigma-Aldrich® according to the manufacturer's guidelines. The quality of DNA is crucial and is purified using protocols described in section 2.6.5. Briefly, inside a sterile 1.5 ml microcentrifuge tube with a total volume of 900 µl reaction, the following reagents were added in the order listed: 45 µg of plasmid DNA (Figure 2.1) was added to sterile molecular grade water to a volume of 405 µl followed by 45 µl of 2.5 M $CaCl_2$. Then 450 µl of 2X HeBS pH 7.05 was added drop-wise while bubbling the solution to form precipitate. This was mixed by vortexing and allowed to incubate at RT for 20 minutes before addition to the adherent cell culture (Phoenix cells, see Table 2.1). These cells transfect best at higher density and optimal density is around 7.5 x 10^6 cells, seeded the night before. This achieves approximately 60 - 70% cell confluence on the day of the transfection. Chloroquine (15 µl of 25 mM stock in 15 ml culture) was added to the cell culture before the last 5 minutes of the incubation. The culture was immediately returned to 33°C incubator supplied with 5% CO_2 for 16 hours. The precipitate then adhered to the cell surface and was internalised. The following day, the chloroquine-containing growth media (DMEM containing 10% (v/v) FCS, 20 µg/ml gentamicin, 2 mM L-glutamine DMEM) was replaced using half volume of fresh growth media. The culture flask was gassed with CO_2 and incubated overnight at 33˚C for optimal retrovirus production.

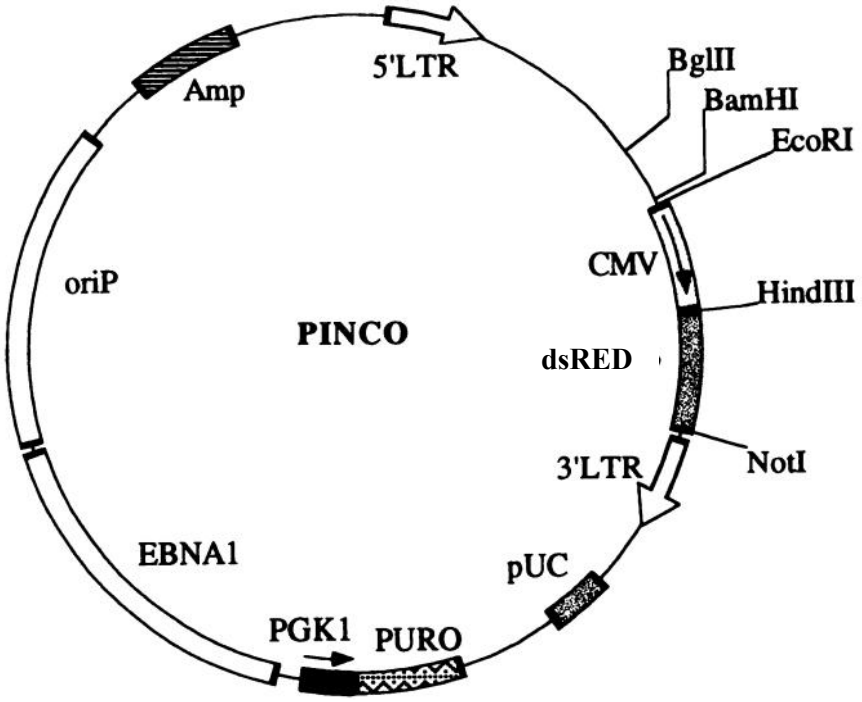


Figure 2.1: Plasmid map of the PINCO retroviral expression vector used in this study. All PINCO vectors contain a cytomegalovirus immediate early promoter (CMV) promoter-driven DsRed expression cassette. Gel-purified insert DNA sequence was cloned into unique *Bam*HI and *Eco*RI sites in the PINCO vector for retroviral long terminal repeats (LTR)-driven expression and the sequence verified before use (Grignani *et al*, 1998).

2.7.2 Production of lentiviruses by transfection

Lentivirus particles were generated and packaged using the HEK293T cell line by co-transfection of three plasmids: the lentiviral vector pLKO.1-puro that contains TCF7L2 shRNA (Figure 2.2), the psL3 [pMD2] as envelope plasmid (Addgene, Cambridge, USA) and psPAX2 as packaging plasmid (Addgene, Cambridge, USA) with *gag, pol*, and *rev* genes. The HEK293T cells transfect best at higher density and 1 x 10^7 cells were plated the night before. For transfection by the calcium phosphate protocol, inside a sterile 1.5 ml microcentrifuge tube with a total volume of 900 µl reaction, the following reagents were added in the order listed: 20 µg of transfer vector DNA, 8 µg of pMD2 DNA and 15 µg of psPAX2 DNA were added to sterile molecular grade water to a volume of 405 µl followed by 45 µl of 2.5M $CaCl_2$. Then 450 µl of 2X HeBS pH 7.05 was added drop-wise while bubbling the solution to form a precipitate. This was mixed by vortexing and allowed to

incubate at RT for 20 minutes before addition to the adherent cell culture (HEK293T cells). Chloroquine was added to the cell culture before the last 5 minutes of the incubation. The culture was immediately returned to a 37°C incubator supplied with 5% CO_2 for 16 hours. The precipitate then adhered to the cell surface and was internalised. The following day, the growth media was replaced using half volume of fresh growth media. The culture flask was gassed with CO_2 and incubated overnight at 37°C for optimal lentivirus production.

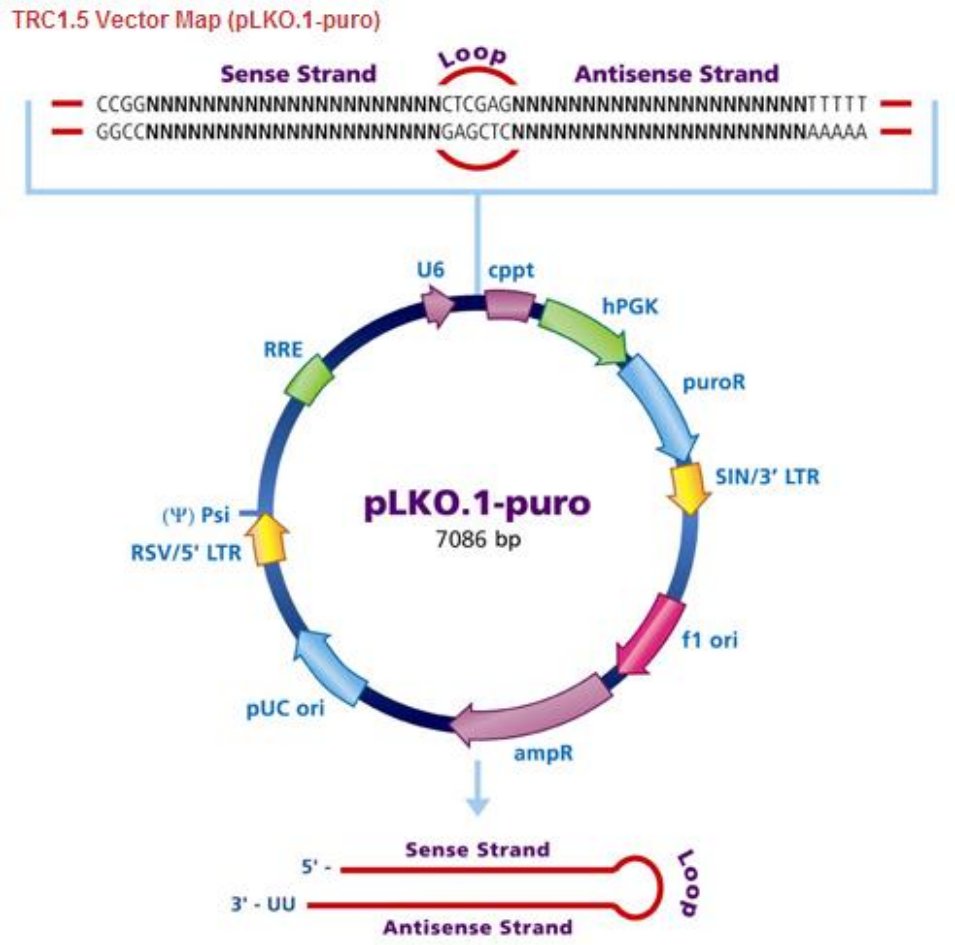


Figure 2.2: Structure of the pLKO.1 puro lentiviral vector backbone for shRNA study. This image was adapted from http://www.sigmaaldrich.com/life-science/functional-genomics-and-rnai/shrna/library-information/vector-map.html#pLKO

2.7.3 Harvesting of retroviral / lentiviral particles

The collection of infective particles was performed on days 2 and 3 following transfection, and the procedure was carefully handled under class II conditions. The retroviral / lentiviral supernatants were centrifuged at 200 x g for 10 minutes at RT. The supernantants were aliquoted into cryovial tubes and briefly snap frozen in LN_2 to lyse any remaining packaging cells and virus stocks were stored at -80°C.

2.7.4 Retroviral / lentiviral transduction of cell lines and haematopoietic cells

RetroNectin from Takara is a recombinant polypeptide of human fibronectin fragments binding viruses and cell surface proteins. When coated on the surface of culture plates, it significantly enhances retrovirus or lentivirus mediated gene transduction into mammalian cells (Tonks *et al*, 2005). RetroNectin-assisted transduction of cells was performed by coating a 24-well plate (non TC treated) with 25 µg of RetroNectin (Takara-Bio, Shiga, Japan) into each well. The culture plate was incubated for a minimum of 2 hours with agitation at RT. Following the incubation, RetroNectin solution was discarded and 250 µl of sterile 1X PBS containing 1% BSA was added into each well for blocking and the plate was left at RT for 30 minutes. During the incubation, each virus stock was thawed rapidly in a water bath set at 37°C. At the end of incubation, 1X PBS 1% (v/v) BSA buffer was removed from each well and immediately replaced with 1 ml of virus. The plate was double sealed with a secure carrier and centrifuged for 90 minutes at 3200 x g. During this process, the target reporter cells (refer 5.3.3.2) were counted and diluted to a density of 4 x 10^5 /ml in the specified culture media. Upon completion of centrifugation, the viral containing supernatant was removed from each well and 2 x 10^5 target cells were added to each well before any drying of the pre-coated well takes place, followed by overnight incubation at 37°C in 5% CO_2.

2.8 PREPARATION OF PROTEIN FOR WESTERN BLOTTING

2.8.1 Cytosolic and nuclear protein isolation

For protein extraction, samples harvested from 2 x 10^6 cell lines or 1 x 10^7 $CD34^+$ HPC were fractionated using a nuclear / cytosol fractionation kit (Biovision, California, USA). Initially, cells were taken from the culture flask and transferred into a conical tube and washed twice with 20 ml of TBS followed by centrifugation at 200 x g for 10 minutes. Cell pellets were fully resuspended in 200 µl CEB-A (Cytosol Extraction Buffer) containing 1 mM Dithiothreitol (DTT) and 1X protease inhibitor cocktail (PIC) and then transferred into a pre-chilled 1.5 ml microtube. Cell mixtures were vortexed for 15 seconds and returned to ice for 10 minutes. Following the incubation, 11 µl of CEB-B buffer was added to the cell mixture followed by a brief 5 second vortex before resuming incubation on ice for another 1 minute. Cells were again vortexed for 5 seconds before undergoing centrifugation at 16 000 x g for 8 minutes in microfuge set at 4°C. The supernatant containing cytoplasmic protein extract was carefully pipetted into a fresh 1.5 ml microtube and held on ice. The residual pellet containing mostly nuclear protein was washed by carefully applying 500 µl of 1X PBS containing 5 mM $MgCl_2$ and centrifuged as above for 3 minutes. Supernatant containing cytoplasmic contaminants was removed and nuclear cell pellets were snap frozen in LN_2 for nuclear disruption. Nuclear protein was extracted by thawing the pellet on ice in the presence of 1 mg/ml DNase and 100 µl of NEB (Nuclear Extraction Buffer) containing 1 mM DTT and 1X PIC. The mixture was vortexed for 15 seconds and a similar process was repeated every 10 minutes during a 40 minute incubation on ice. Finally the mixture was centrifuged for 15 minutes at 16 000 x g in a 4°C microfuge to pellet detergent-insoluble material before isolating the supernatant containing nuclear protein. Both cytosolic and nuclear protein lysates were stored at -80°C.

2.8.2 Determination of protein concentration

The concentration of nuclear and cytosolic protein from cell lysates was determined using the Bradford's reagent protein assay (Sigma-Aldrich®). Initially, protein calibration standards were prepared by diluting BSA stock ranging from 0, 10, 40, 70 and 100 µg/ml BSA in lysis buffer. Before aliquoting the protein extracts, Bradford's reagent was diluted with an equal volume of water to create sufficient diluted reagent (190 µl per test well). An

aliquot of 10 µl of each protein standard, and 10 µl of each cell lysate was placed into the well of a 96-well microtitre plate (in duplicate) and 190 µl of diluted Bradford's working solution was mixed with each sample. Absorbance of the samples was measured at 590 nm using an ASYS Hitech Expert plus spectrophotometer (Biochrom, Cambridge, UK). BSA standard solutions were assayed in duplicate alongside protein extract samples to create a standard curve, which enabled calculation of protein extract concentrations.

2.8.3 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) and electroblotting

Protein electrophoresis for western blotting was performed using the NuPAGE® X-Cell II® electrophoresis and electroblotting systems (Invitrogen™). A maximum of 20 µg of protein extracts were denatured by incubation for 10 minutes at 70°C in the presence of 1X LDS sample buffer and 50 mM DTT. In order to estimate the molecular weight (MW) of proteins, outer lanes of SDS-PAGE gels were loaded with 10 µl of 1X LDS buffer containing a 20 fold dilution of MagicMark XP® Protein Standard (Invitrogen™). Denatured protein samples (20 µl per lane between 10 – 20 µg) were loaded onto 4 – 12% (w/v) Bis-Tris gels (Invitrogen™) in the presence of antioxidant using MOPS-SDS running buffer and electrophoresed at 200 V for 50 minutes inside the X-Cell II® Mini-Cell apparatus. Upon completion of electrophoresis, the gel was removed from the tank and one surface of the gel was moistened with transfer buffer containing 10% (v/v) methanol, 1X NuPAGE® Transfer Buffer and 1 ml antioxidant. During this process, a piece of blotting pad (or sponge) was soaked inside the similar transfer buffer and was then carefully layered onto the moistened gel avoiding any air bubbles. The other side of the SDS-PAGE gel was also moistened with transfer buffer, before layering a pre-soaked nitrocellulose membrane (0.45 µm pore size) ensuring all trapped air bubbles were removed. Sandwiched between pre-soaked blotting pads, the SDS-PAGE gel and nitrocellulose membrane (Figure 2.3) was then inserted into an XCell II™ Blot Module. The blot module was filled to the top of the pads with transfer buffer and the outside of the tank filled with distilled water and protein transfer was performed over 1 hour at 30 V (400 mA).

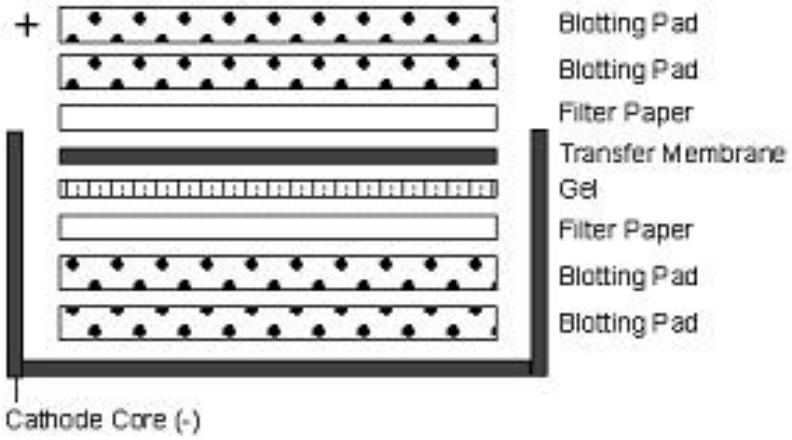


Figure 2.3: Assembly of nitrocellulose transfer membrane and SDS-PAGE gel for protein transfer by using the using the X-Cell II® Blot Module. Image was adapted from http://www.invitrogen.com/site/us/en/home/References/protocols/

2.8.4 Protein detection and chemiluminescent analysis

Upon completion of electrophoresis and electroblotting, the nitrocellulose membrane was washed twice with 10 ml of distilled water for 5 minutes at RT and incubated with Ponceau S solution for 30 seconds to check for equal protein loading and transfer before blocking the membrane. The membrane was rinsed with distilled water for 5 minutes to remove Ponceau S stain and incubated with 10 ml of TBS-T containing 5% (w/v) membrane blocking agent ECL Advance® (GE Healthcare, Amersham, UK) for 1 hour at RT. Blocked membranes were then rinsed with 10 ml TBS-T for 15 minutes and three subsequent 5 minute washes. Following washing, blocked membranes were incubated for 16 hours at 4°C with shaking in the presence of diluted primary antibody (Table 2.3) in TBS-T containing 5% (w/v) membrane blocking agent. Following primary antibody incubation, washed membranes were incubated with diluted HRP-conjugated anti-mouse or anti-rabbit secondary antibody in TBS-T containing 2% (w/v) membrane blocking agent for 1 hour at RT. Finally, membranes electroblotted with proteins labelled with both primary and secondary antibodies were visualised using ECL Advance® Western Blotting Detection Kit (GE Healthcare). Each immunolabelled membrane was incubated for 5 minutes at RT with an equal volume of Solution A containing Tris buffer in 3.2% (v/v) ethanol and

Solution B containing proprietary substrate in Tris buffer. Membranes were placed between two sheets of clean acetate and ECL Advance® chemiluminescence was detected using LAS3000 digital imaging device (Fujifilm) and images were processed and analysed using AIDA Image Analyzer v4.19 (Fujifilm).

Monoclonal antibody (mAb) reagent	Manufacturer	Dilution used
Anti-mouse HRP-linked whole Ab (from sheep) - NA931V	GE Healthcare, Buckinghamshire, UK.	1:50k
Anti-rabbit HRP-linked whole Ab (from donkey) - NA934V	GE Healthcare, Buckinghamshire, UK.	1:50k
Mouse anti human Histone H1 mAb - 4974-78	AbD Serotech, Kidlington, UK	1:2k
Mouse anti human GAPDH mAb - 6C5	Santa Cruz Biotechnology Inc., Middlesex, UK	1:10k
Rabbit anti human TCF7L2 mAb – C48H11 (anti exon 11)	Cell Signaling, Hitchin, UK	1:2k
Rabbit anti human TCF7L2 mAb - EP2033Y (anti exon 1)	Abcam plc, Cambridge, UK	1:20k

Table 2.3: List of primary and secondary antibodies used for western blotting.

2.9 FLOW CYTOMETRIC PROCEDURE

The principle of flow cytometry is to deliver a single cell to a point of measurement: a beam of light (usually a laser). The sample is injected into the centre of a stream of liquid known as the sheath fluid and this sample stream is focused. Combination of fluorescence-labelled antibodies or chemicals with fluorescent properties, and measurement of scattered light provide a number of measured parameter including size, protein expression and cell viability. The basic instrument consists of a source of light, a flow cell, optical components to focus light of different colours on to the detectors and amplifier, and a computer to

process the resulting signals (Figure 2.4). The system used was a benchtop model supplied by Accuri C6 Flow Cytometer™ (Accuri Cytometers, Ann Arbor, MI, USA) equipped with a 'blue' and a 'red' laser, two scatter detectors, and four fluorescence detectors with interference filters as described in Table 2.4.

2.9.1 Sample preparation

Single cells must be suspended at a density of 1 x 10^5 –10^7 cells / ml to prevent the flow cytometer capillary from clogging up. Cells were harvested and washed with 1X PBS and stained according to section 5.3.5.2 depending on the experiment requirements. Stained cells were analysed directly using the Accuri flow cytometer using the appropriate optical filters. For sample preparation, appropriate volume of cells was recovered from growth medium by centrifugation at 180 x g for 5 minutes. Cell pellet of approximately 1 x 10^6 cells were washed in 1 ml of flow cytometry staining buffer (refer 2.2) and centrifuged as above. Before data acquisition, cells to be analysed were resuspended in a maximum of 100 µl flow cytometry staining buffer. Data were acquired using a minimum of 15,000 collected events. Debris was excluded by gating live cells using forward scatter (FSC) and side scatter (SSC) characteristics. FSC intensity was used to detect the size of a particle and to distinguish between debris (events with low FSC signal) and live cells (with higher FSC signal). Appropriate compensation data were applied to the analysis workflow according to the dye used. Events were acquired using the Accuri C6 cytometer (Accuri Cytometers, Ann Arbor, MI, USA) and raw data analysis was performed using FCS Express version 4 (De Novo Software, Los Angeles, CA, USA).

Parameter	Description
Laser Excitation	488 nm 640 nm
Laser Profile	10 x 75 µm
Light Scatter Detection	Forward (0°, ±13°) Side (90°, ±13°)
Emission Detection / Optical filters (4 colours)	FL1 533/30 nm (eg, FITC / GFP) FL2 585/40 nm (eg, PE / PI, bio) FL3 > 670 nm (eg, 7-AAD, PerCP,BIO PerCP-Cy5.5, PE-Cy™7) FL4 675/25 nm (eg, APC)

Table 2.4: Optical systems in the Accuri C6 Flow Cytometer™.

2.9.2 Data analysis

2.9.2.1 Histogram analysis for BIO /WNT3A stimulated samples

To determine the percentage of cells showing reporter response in terms of green flurescent protein (GFP) expression between stimulated (BIO or WNT-3A) and non-stimulated (vehicle) cells, the histogram Overton subtraction tool was used (Overton, 1988). The GFP fluorescence from the non-stimulated sample was therefore subtracted from the stimulated sample. The positive difference value was subsequently normalised to the control sample. Cultures transduced with PINCO-DsRed constructs, were first gated for DsRed positivity. The autofluorescence signals from parental cells were used as a fluorescence threshold for DsRed gating.

2.9.2.2 Histogram analysis for surface marker staining

Histogram data analysis was performed similarly to 2.9.2.1 except that each histogram include the FITC conjugated IgG isotype control for the CD14 and CD15 surface marker antibodies to account for any background effect of the analysis. In order to calculate positive differences in each sample, the FITC fluorescence from isotype stained was subtracted from the lineage marker stained by Overton subtraction. To calculate relative positive stained cells, the positive difference in empty vector (EV) transduced control was divided by the positive difference in TCF7L2 transduced samples.

2.9.3 Statistical analyses

To determine if a difference exists between two groups (target *vs* control), significance testing of difference using the 2-tailed Independent Student's T-test was performed using SPSS for Windows version 11.5 (SPSS Inc., Chicago, USA).

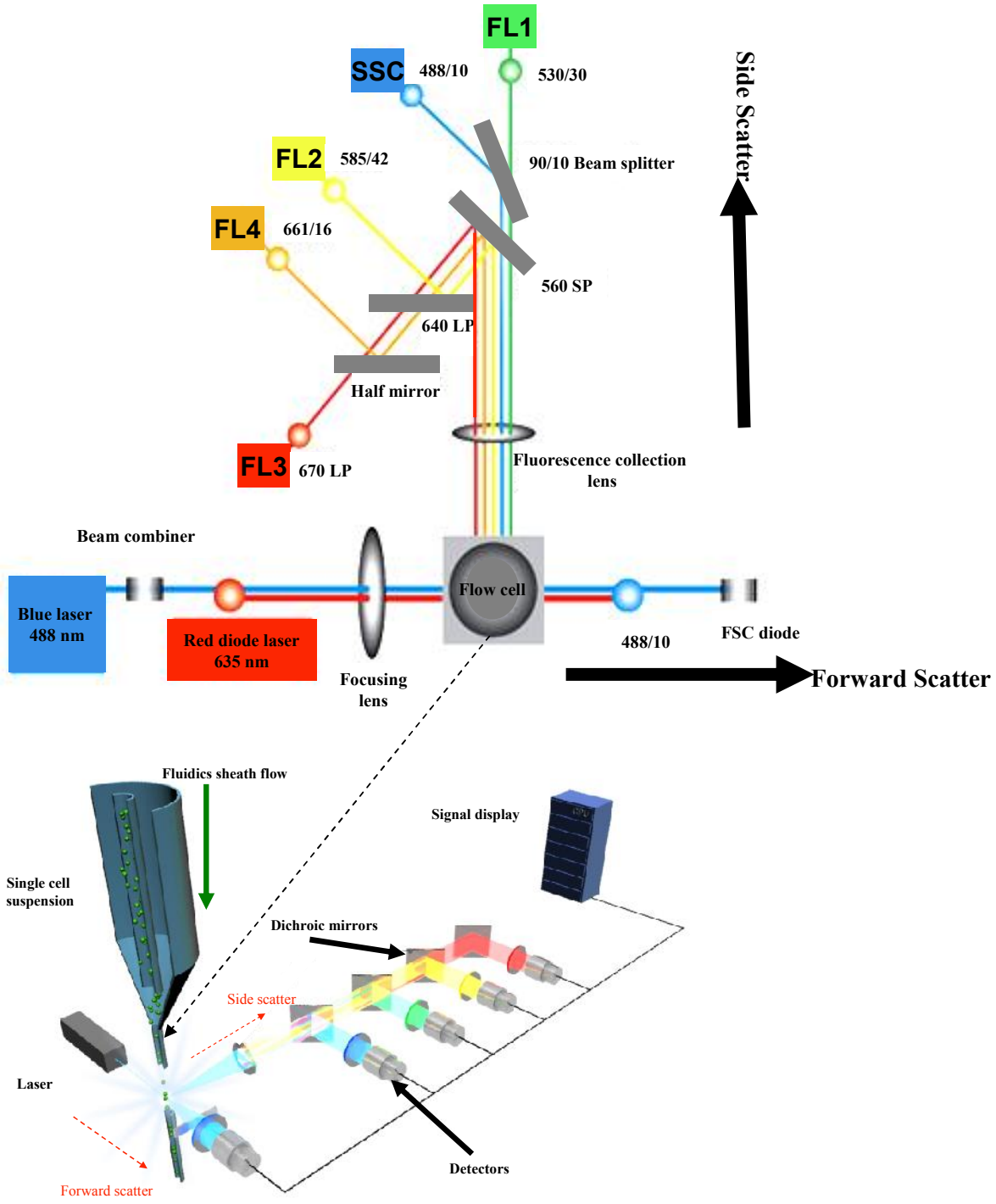


Figure 2.4: Schematic overview of typical flow cytometry setup (see text for details). Image was adapted from http://www.bdbiosciences.com/instruments/facscalibur/features/index.jsp.

3 - Identification of dysregulated pathways and genes in AML

3.1 INTRODUCTION

AML is genetically heterogeneous with a unifying feature of inhibition of terminal myeloid development arising from disordered gene expression and dysregulation of developmental pathways. As well as somatic mutations (Rocquain *et al*, 2010), chromosomal abnormalities and deletions (Ebert, 2010; Nagarajan, 2010), disordered gene expression can arise from dysregulated non-coding RNA (Calin *et al*, 2007; Isken *et al*, 2008) and epigenetic regulation (Jiang *et al*, 2009; Martin *et al*, 2010). For example constitutive activity within leukaemic cells can occur due to changes in expression of transcriptional regulators such as *STAT3* (Spiekermann *et al*, 2002), *GATA2* (Wieser *et al*, 2000), *DNMT3A* (Mizuno *et al*, 2001; Hayette *et al*, 2012) or *RUNX1* (Yan *et al*, 2009). Pathway activation could also be due to epigenetic inactivation of important gene regulators such as silencing of *SFRP1* and *PTEN* (Reins *et al*, 2010; Yoshimi *et al*, 2011; Li *et al*, 2012). In the past decade, GEP using microarray technologies and sequencing technologies have become popular tools for the discovery of dysregulated gene expression or identification of high frequency gene mutations involved in cancer pathogenesis (Mi *et al*, 2007; Bacher *et al*, 2009; Welch & Link, 2011; Borate *et al*, 2012). These abnormalities are likely to act in a concerted manner to dysregulate the activity of a number of cellular processes or pathways. Given that most high frequency mutations have already been detected, it is hoped that elucidation of changes in gene expression and pathway signalling may generate a path to identifying viable therapeutic targets (Ricciardi *et al*, 2012; Ungewickell & Medeiros, 2012).

This chapter uses GEP to identify commonly dysregulated genes in AML and to characterise these changes with respect to aberrant pathway activation which may be linked to the pathogenesis of this disease. Analysis of gene expression associated with clinical outcome data allowed the identification of gene targets with clinical relevance. To this end, this study describes a method used to compare microarray data from public repositories and

inter-related platforms using batch effects removal. Commonly dysregulated genes associated with clinical importance were prioritised for subsequent analysis.

3.2 AIMS

The overall aim of this chapter is to discover aberrant gene expression and signalling pathways that may contribute to the pathogenesis of AML. The specific objectives of this chapter were:-

- To identify aberrantly regulated pathways in AML. This will be achieved by comparing the GEP of AML patient blasts to normal HPCs derived from $CD34^+$ BM using comprehensive genome wide microarray analysis.

- To establish the most dysregulated genes in AML within the most significantly dysregulated pathway.

- To validate microarray expression data of significant genes using qRT-PCR.

- To determine the association of aberrant gene expression with AML patient clinical outcome.

3.3 MATERIALS AND METHODS

3.3.1 AML patient blast and normal CD34$^+$ BM sample GEP datasets

A subset of BM or PB samples from AML patients who had enrolled in MRC-UK NCRI AML clinical trial protocol were used in this study. Each AML patient sample was taken at *de novo* presentation of the disease. The samples taken for this study were approved by local and multicentre ethical committees in the UK. Each patient sample was taken at diagnosis before treatment in accordance with the 1964 Declaration of Helsinki. Biological characteristics of AML patients that passed quality control (QC) and were subsequently used for analysis (see 3.4.1) are shown in Table 3.1. The procedure for HG-U133 Plus2.0 array refers to the customized Roche AmpliChip Leukaemia Custom Microarray kit produced by Roche Molecular Systems, Pleasanton, USA to be used specifically for the MiLE study in conjunction with the HG-U133 Plus2.0 Affymetrix GeneChip®. The production of the Amplichip was developed independently for the MiLE study stage II conducted between November 2006 until October 2007 (Kohlmann *et al*, 2008b).

RNA sample processing for HG-U133A array was performed using the GeneChip® 3' IVT kit produced by Affymetrix® (High Wycombe, UK) according to manufacturer's recommendations:- http://media.affymetrix.com/support/downloads/manuals/expression_analysis_technical_manual.pdf. Both procedures were conducted previously by Amanda Gilkes from the Department of Haematology, Cardiff University.

GEP data from healthy CD34$^+$ BM samples (defined herein as Normal Controls) were downloaded from ArrayExpress website available at http://www.ebi.ac.uk/microarray-as/ae/at. A total of 10 raw .CEL files from HG-U133A were obtained from E-GEOD-13496 (n=8) and E-GEOD-12803 (n=2). A total of 13 raw .CEL files from HG-U133 Plus2.0 were obtained from the following experiments: E-GEOD-12662 (n=3), E-GEOD-4619 (n=10).

Characteristic at diagnosis	HG-U133 Plus2.0 (n=98)	HG-U133A (n=160)	DWD (merged GEP) (n= 246)
Trial group			
AML10	4	2	7
AML11	2	3	7
AML12	1	47	47
AML14	16	36	55
AML15	73	72	128
AML16	-	-	2
Not determined	2	-	-
Sample type			
Bone marrow	66	3	70
Peripheral blood	31	-	16
Not determined	1	157	160
Age			
15-39	26	50	65
40-69	64	95	158
>70	8	15	23
WBC count ($x10^9$ /L)			
<10	24	50	81
10-99	54	82	120
>100	18	28	44
Not determined	2	-	1
FAB subtype			
M0	3	9	12
M1	14	25	37
M2	29	39	57
M3	5	24	28
M4	18	31	46
M5	19	15	32
M6	1	-	1
M7	-	2	2
Not determined	9	15	31
Gender			
Male	49	88	131
Female	47	72	115
Not determined	2	-	-
Karyotype group			
t(15 ;17)	5	23	28
t(8 ;21)	6	14	17
inv(16)	7	16	23
Complex (>5 chromosomal abnormalities)	8	11	18
Normal karyotype	35	51	90
Other chromosomal abnormalities	11	31	47
Not determined	26	14	23
Cytogenetic prognostic classification			
Favourable	34	53	68
Intermediate	50	85	135
Adverse	11	17	28
Not determined	3	5	15

Table 3.1: Stratification of data of AML patients used in this study. Stratified AML patient parameters obtained at diagnosis according to whether the samples were analysed in the context of HG-U133 Plus2.0, HG-133A or DWD-merged GEP data. The DWD-merged GEPs data set was created to increase statistical power for clinical outcome analysis (see 3.3.6). The table shows the trial group (chemotherapy protocols for each group are detailed in Supplementary 3.1 and cytogenetic prognostic classification according to Grimwade *et al*, 1998 detailed in Supplementary 3.2 .

3.3.2 Data import and assessment of GeneChip® data quality

3.3.2.1 Visual inspection of GeneChip® raw data signal

Microarray 2-D raw images, .CEL and .CDF were initially imported into dChip MFC Application (version 1.0.0.1) for background visual inspection. The application tool for dChip was available from http://www.hsph.harvard.edu/cli/complab/dchip/ (Li, 2008). Pseudo images will show artefacts not visible in the raw data image. With the help of Dr Anna Evans from the Haematology Trials Unit, Cardiff University, pseudo GeneChip® images were created from probe level model (PLM) weights and PLM residuals generated using Bioconductor 2.11 packages written in R language for bioinformatics application (Gentleman *et al*, 2004). The standard Bioconductor suite software was installed from http://www.bioconductor.org/getBioC.R. Probe level investigation was carried out in RScript1 2.6.1 that was created by Ihaka and Gentleman (Ihaka & Gentleman, 1996) available for download at http://www.r-project.org.

3.3.2.2 Hybridisation and internal QC assessment

All .CEL files were imported into Partek® Genomics Suite version 6.5 (Partek Inc., St. Louis, USA). An overall QC analysis was performed for every GeneChip® to detect GEP with lower or poor quality. The QC analysis was performed using customized versions of the "affyQCReport packages" available from Partek® Genomics Suite. The QC pipeline analyses several internal control sequences as included in the array. The internal control probe sets are the hybridisation controls with internal probe calls for "AFFX-r2-Ec-bioB", bioC, bioD and cre which are *Escherichia coli* genes that must always be present, also the poly-A controls with the following probe sets "AFFX-r2-Bs-Dap", Thr, Phe and Lys which are the modied *Bacillus subtilis* genes and should be called present at a decreasing intensity. Other QC probes analysed by the package include a normalisation control set (of 100 probe sets) and housekeeping gene; Glyceraldehyde 3-phosphate dehydrogenase *(GAPDH).* GEP data are flagged if these probes have an unusually high variance, high RNA degradation levels, overall low signal levels (poor hybridisation or labelling), or if the raw image of array shows prominent spurious artefact.

3.3.2.3 Multiple array signal pre-processing and normalisation

The main goal of normalisation is to remove systematic bias in the data as completely as possible, while preserving the variation in gene expression that occurs due to biologically-relevant changes in transcription. In short, the raw optical microarray image files .DAT files and .CEL files were generated using default Affymetrix GeneChip® Operating System (GCOS) parameters. The imported .CEL files are in text format and contain intensity values for PM and MM features, and are converted from pixel-level .DAT files. Only raw .CEL files were used for downstream expression analysis using Partek®. For Partek® analyses, .CEL files were imported using Affymetrix annotation files (NetAffx, Version na31. hg19). Robust multi-array averaging (GC-RMA) background correction algorithms were used to normalise GEP data. GC-RMA is a method of converting .CEL files into expression data with the help of probe sequence and with GC-content background correction. The GC-RMA algorithm normalises and summarises probe level intensity of the PM probes by quantile normalisation, Log_2 transformation and median polishing for summarisation. This will result in a set of expression measures. The algorithm for the RMA model used is shown below (Bolstad *et al*, 2003; Irizarry *et al*, 2003b):

$$y_{ij} = m + \alpha_i + \beta_j + \varepsilon_{ij}$$

where $y_{ij} = \log_2 \mathrm{N}\left(\mathrm{B}\left(PM_{ij}\right)\right)$

α_i is a probe-effect i= 1,…,I

β_j is chip-effect ($m + \beta_j$ is log2 gene expression on array *j)* j=1,…,J

By using the method described as above, two GEP databases were created (i) HG-U133 Plus2.0 and (ii) HG-U133A; each based on the GeneChip® to which the RNA was hybridised (see 3.3.1). GEP data from each array type was analysed separately (3.3.3) or was combined into a single database (see 3.3.6.1) for further batch correction using DWD software to create a larger cohort of AML samples for clinical correlational study.

3.3.2.4 Exploratory analysis using Principal Components Analysis (PCA)

PCA is a statistical technique used to select key variables for differences seen in multidimentional data (eg: microarray gene expression) and helps to simplify visualisation of complex data into 3 dimensions, based on the published method for discriminant analysis (Krzanowski, 1980). PCA was applied to all GEPs hybridised to the HG-U133 Plus2.0 GeneChip® to provide insight into the underlying factors with most variability, to aid in QC (i.e: identifying outliers) and sample grouping. The information gathered from PCA was then used to find suitable factors to be included into the Analysis of Variance (ANOVA) model (as shown in Figure 3.8) towards finding significant sources of variation across all samples.

3.3.3 Calculating differentially expressed genes using Mixed-Model ANOVA

A total of 111 HG-U133 Plus2.0 .CEL files (representing 98 AML patients' blasts and 13 Normal Control GEPs) that had passed QC were used for this downstream analysis. The HG-U133 Plus2.0 dataset was chosen because this array provides more comprehensive coverage of probe sets compared to HG-U133A array. The specifications of both arrays are described in Section 1.5. Initially, the common underlying gene changes associated specifically with the more immature developmental subtypes of AML (M0 and M1) together with the changes seen in other subtypes (M2 - M6) were analysed. This was achieved by a two arm approach using either 1-way ANOVA or 4-way ANOVA (Figure 3.1).

The 1-way ANOVA was based on the following model: $Y_{ij} = \mu + FAB_i + \varepsilon_{ij}$, where Y_{ij} represents the j^{th} observation on the i^{th} FAB, μ is the common effect for the whole experiment and ε_{ij} represents the random error present in the j^{th} observation on the i^{th} FAB. The errors ε_{ij} are assumed to be normally and independently distributed with mean 0 and standard deviation δ for all measurements. The contrast method using Fisher's Least Significant Difference was performed to compare AML-M0 *vs* Normal Control and AML-M1 *vs* Normal Control.

The 4-way ANOVA is based on the following model: $Y_{ijklm} = \mu + Disease_i + Sample\ type(Disease)_{ij} + FAB(Disease)_{ik} + Scan\ Date(Disease)_{il} + \varepsilon_{ijklm}$, where Y_{ijklm} represents the m^{th} observation on the i^{th} Disease, j^{th} Sample type, k^{th} FAB, l^{th} Scan Date , μ is the common effect for the whole experiment and ε_{ijklm} represents the random error present in the m^{th} observation on the i^{th} Disease, j^{th} Sample type, k^{th} FAB, l^{th} Scan Date. The errors ε_{ijklm} are assumed to be normally and independently distributed with mean 0 and standard deviation δ for all measurements and the Scan Date is a random effect.

The principles for both ANOVA models performed by Partek® were adapted from Method of Moments (Eisennhart, 1947). Significant genes were then filtered to include only changes of > 1.5 or < -1.5 fold changes between AML subtype(s) *vs* Normal Control. Partek's ANOVA uses Least Squares (LS) mean (Tamhane & Dunlop, 2000) to calculate fold change. When the data is logged (such as by GC-RMA), the resulting LS means are anti-logged, producing geometric means. Statistically significant genes were set to pass the false discovery rate (FDR) threshold of P <0.05 based on the Benjamini–Hochberg test. This filtering produced three genelists A, B and C (Figure 3.9A) based on comparisons made as above.

Following pathway maps analysis in which WNT signalling was identified as a significantly dysregulated pathway (see Table 3.3), the ANOVA test was repeated using the above contrasts on a focused smaller subset of probe sets which were manually selected to include 651 of the WNT signalling related probe sets (low probe density) (for ANOVA see Figure 3.2). Following similar fold change and *P*–value filtering, an additional three genelists D, E and F were generated (for results see Figure 3.9B).

The details of the 651 WNT related probe sets (the 'WNT-ome') are listed in Supplementary 3.3 . These probe sets were compiled based on the curated genes obtained from the GeneGo Metacore™ WNT signalling pathway map database and also based on published WNT genes reviewed in other public pathway databases, primarily from the Stanford University WNT homepage :http://www.stanford.edu/group/nusselab/cgi-bin/wnt/main and Kyoto Encyclopedia of Genes and Genomes (KEGG) bioinformatics resources for WNT pathway: http://www.genome.jp/kegg/kegg3a.html. The accurate Affymetrix® probe set code for each gene chosen was then retrieved from NetAffx™ Analysis Centre, accessible from http://www.affymetrix.com/analysis/index.affx, in order to match the HG-U133 Plus2.0 annotation on the GeneChip® used. The information about alternative gene names, approved symbol abbreviation and aliases for each gene was

manually compiled and obtained from http://www.genecards.org (Rebhan *et al*, 1997; Safran *et al*, 2003). The approved gene symbol corresponding to each probe set was based on the HUGO Gene Nomenclature Committee (HGNC) database.

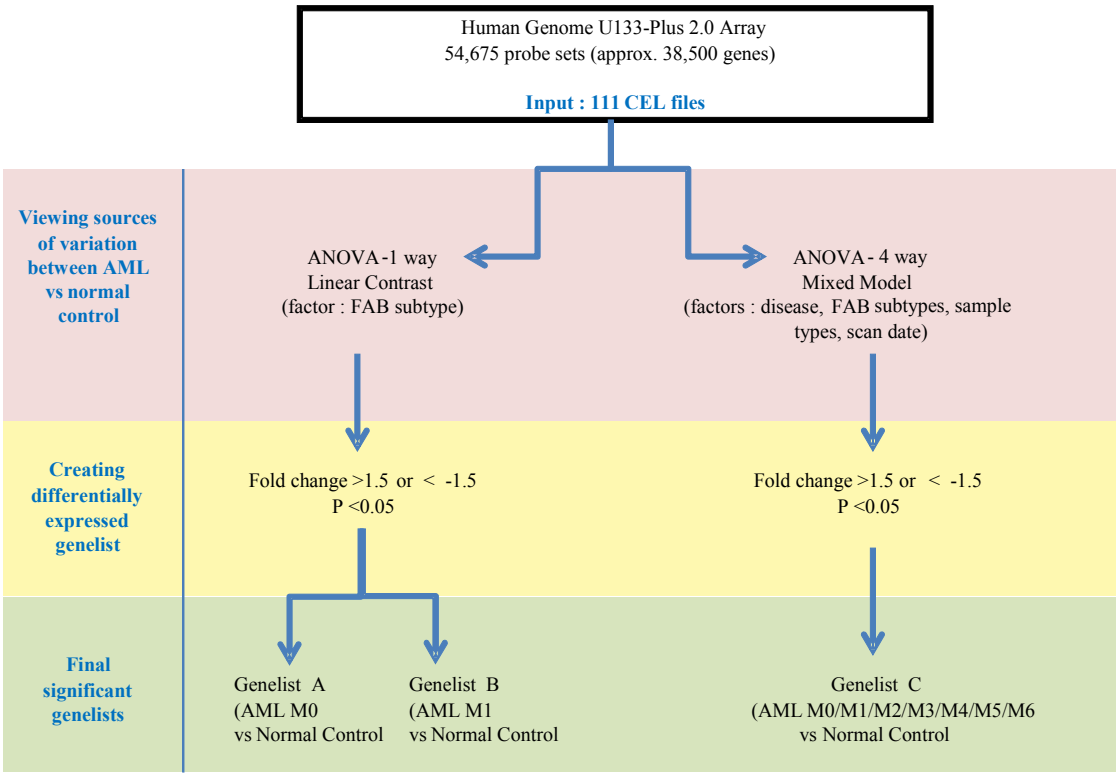


Figure 3.1 : Flow diagram illustrating the ANOVA workflow to generate genelists A, B and C. Two approaches (1-way and 4-way ANOVA) were used to search for differentially expressed genes in AML blasts *vs* Normal Control samples from whole genome probe sets. Sample types and FAB subtypes are as characterised in Table 3.1.

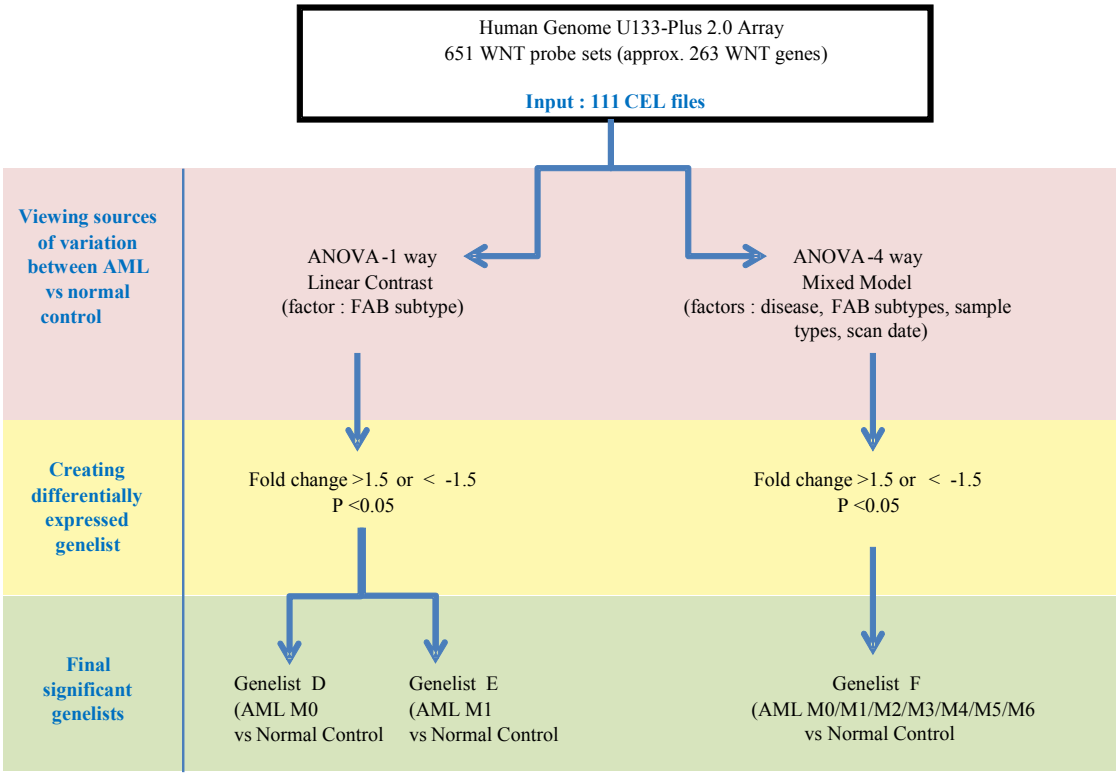


Figure 3.2 : Flow diagram illustrating the ANOVA workflow to generate genelists D, E and F containing differentially expressed genes in the 'WNT-ome'. Two approaches (1-way and 4-way ANOVA) were used to search for differentially expressed genes in AML blasts *vs* Normal Control samples from 'WNT-ome' probe sets. Sample types and FAB subtypes are as characterised in Table 3.1. 'WNT-ome'probe sets (n=651) are as characterised in Supplementary 3.3 .

3.3.4 Downstream analysis of significant genelists A-F

3.3.4.1 Pathway Maps analysis (genelists A-C)

Pathway analysis was generated through the use of specialised software provided by Thomson Reuter (Thomson Reuters, New York, USA) called MetaCore™ from GeneGo Inc. (GeneGo Inc., Michigan, USA) available at http://www.genego.com/faq.php. This is a web application that generates Pathway Maps combined with functional ontology enrichment analysis which can statistically rank the most relevant Pathways from the input dataset. All maps were created by a team of experts and scientists from Thomson Reuters by a high-quality manual curation process based on published peer-reviewed literature to ensure its reliability. There are over 70,000 pathways within Metabase™ which is the internal knowledge base behind the Metacore™ discovery tool.

For the input dataset, selected values from the three GEP genelists A, B and C were imported into MetaCore™ and the interconnectivity of each probe set was linked to its most biologically relevant pathways. These genelists were derived from the whole genome analysis thus allowing wider coverage of probe sets. In all three analyses, the following values were selected for the input dataset: gene identifiers (in probe set format) and their corresponding fold change characteristics and *P*-values obtained from the previous ANOVA (3.3.3). These values were uploaded as an Excel file into the MetaCore™ database. Statistically significant pathway map sorting was applied. The pathways were ranked in a format of table (and graphic) according to their newly calculated *P*-value of hypergeometric intersection (further details about *P*-value algorithm in 3.3.4.2). The *P*-value of hypergeometric intersection was set at <0.05 (no more than 5% false positives) and used to estimate the probability of a random intersection between the input probe set IDs with ontology gene IDs within the MetaCore™ database. The lower *P* -value means higher relevance of the gene ontology IDs to the dataset, which shows as a higher rating for the given probe sets of the corresponding pathway. The maps are arranged by hypergeometric *P*-value and the sorting must pass the 1.5 fold change expression background threshold.

3.3.4.2 Interactome and protein enrichment analysis (genelists D-F)

The Interactome tool provided by MetaCore™ allows estimation of interconnectedness for the input dataset based on published literature, (e.g. a gene may regulate another gene or they may bind with each other). The tool will estimate statistically significant interactions in the set, and enrichment of the dataset according to protein.

In order to identify statistically significant genes that regulate other genes (i.e by direct binding, activate phosphorylation, transcription regulation etc), interactome analysis using enrichment by protein function was applied. All probe sets within genelists D, E and F that represent only WNT signalling genes were imported and uploaded as a tab-delimited text file into the MetaCore™ database. All three genelists were analysed for relative enrichment with certain protein classes. The output data divided genes / proteins into seven different functions (protein classes): transcription factors (TFs), receptors, ligands, kinases, proteases, phosphatases and metabolic enzymes. The results were ranked by a *P*-value. The *P*-values were calculated using the basic formula for a hypergeometric distribution similar to the *P*-values for Pathway Maps, where r is the number of probe sets of a particular protein class from the set of interest (input probe sets); R is the number of probe sets in the set of interest; n is the number of probe sets of particular protein class in the whole GeneGo global network; N is the number of probe sets in the GeneGo global network.

$$P(r, n, R, N) = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}$$

$$p\mathrm{Val}(r, n, R, N) = \sum_{i=\max(r, R+n-N)}^{\min(n, R)} P(i, n, R, N)$$

All probe sets for given genes were associated with their proteins and subsequently, all protein lists were screened for the number of interactions with the GeneGo global interactome network (interconnections) and within the individual protein lists (intraconnections). The degree of over- and under-connectivity were evaluated by z-score. The z-score signifies the difference between the obtained number of proteins and the expected average number of proteins corresponding to genes expressed in units of standard dispersion, where r = number of proteins derived from the current protein list that have interactions with given protein; R = total number of proteins in the GeneGo global network

that have interactions with the given protein; n = total number of proteins in given protein list; and N = total number of proteins in the GeneGo global network.

$$Z - \text{score} = \frac{r - n\frac{R}{N}}{\sqrt{n\left(\frac{R}{N}\right)\left(1 - \frac{R}{N}\right)\left(1 - \frac{n-1}{N-1}\right)}}$$

All formulae used for scoring and prioritisation of genes and pathways according to their relevance to input data were published by MetaCore™ from GeneGo Inc. Significant data was considered to have $P < 0.05$ with FDR applied.

3.3.5 Validation of TCF7L2 microarray expression using qRT-PCR

Based on the degree of over-expression, *P*-value significance, interactome analysis and relevant literature reviews, the *TCF7L2* gene was chosen for subsequent analysis (results in 3.4.6). This study first sought to validate mRNA normalised GEP data for this gene by qRT-PCR. A small cohort (n=24) of randomly selected AML patients' RNA samples representing HG-U133A (n=15) and also HG-U133 Plus2.0 (n=9) were analysed for *TCF7L2* mRNA gene expression using qRT-PCR. The primer pair was designed to amplify the region between exon 7/8 and exon 9 that was partially covered by the *TCF7L2* Affymetrix probeset 212759_s_at (illustrated in Figure 3.3). The forward primer spans the boundary of exon 7 and 8 to increase specificity and to eliminate genomic amplification. The RNA template was reverse transcribed (2.5.3) and the qRT-PCR assay is described in 2.5.4. The source of total RNA used for qRT-PCR validation was similar to the microarray described in 3.3.1. For qRT-PCR data analysis, raw cycle threshold (Ct) values of *TCF7L2* and *ABL* were collected from LightCycler LCS4 4.0.0.23 (Roche Diagnostics). These raw values were then imported into GeneSpring GX-11.0 (Agilent Technologies, Berkshire, UK) and processed according to the baseline transformation option provided in GeneSpring GX RealTime PCR analysis tools. In short, *TCF7L2* averaged Ct values were normalised, in which signal differences between averaged Ct of *ABL* and the target gene *TCF7L2* within each sample was calculated. Subsequently, for each gene, the median of the Log_2 summarised values across all the samples was calculated in order to get the final `baseline transformed' normalised *TCF7L2* expression.

Statistical significance of the correlations was determined by comparing baseline transformed values from qRT-PCR with the median polished Log_2 values of the GC-RMA normalised data of *TCF7L2* after DWD adjustment (DWD data as described in 3.3.6.1). Pearson's correlation coefficient was used. Correlational statistical analyses utilized a 2-tailed significance level of $P < 0.01$ and were performed using SPSS for Windows version 11.5.0.

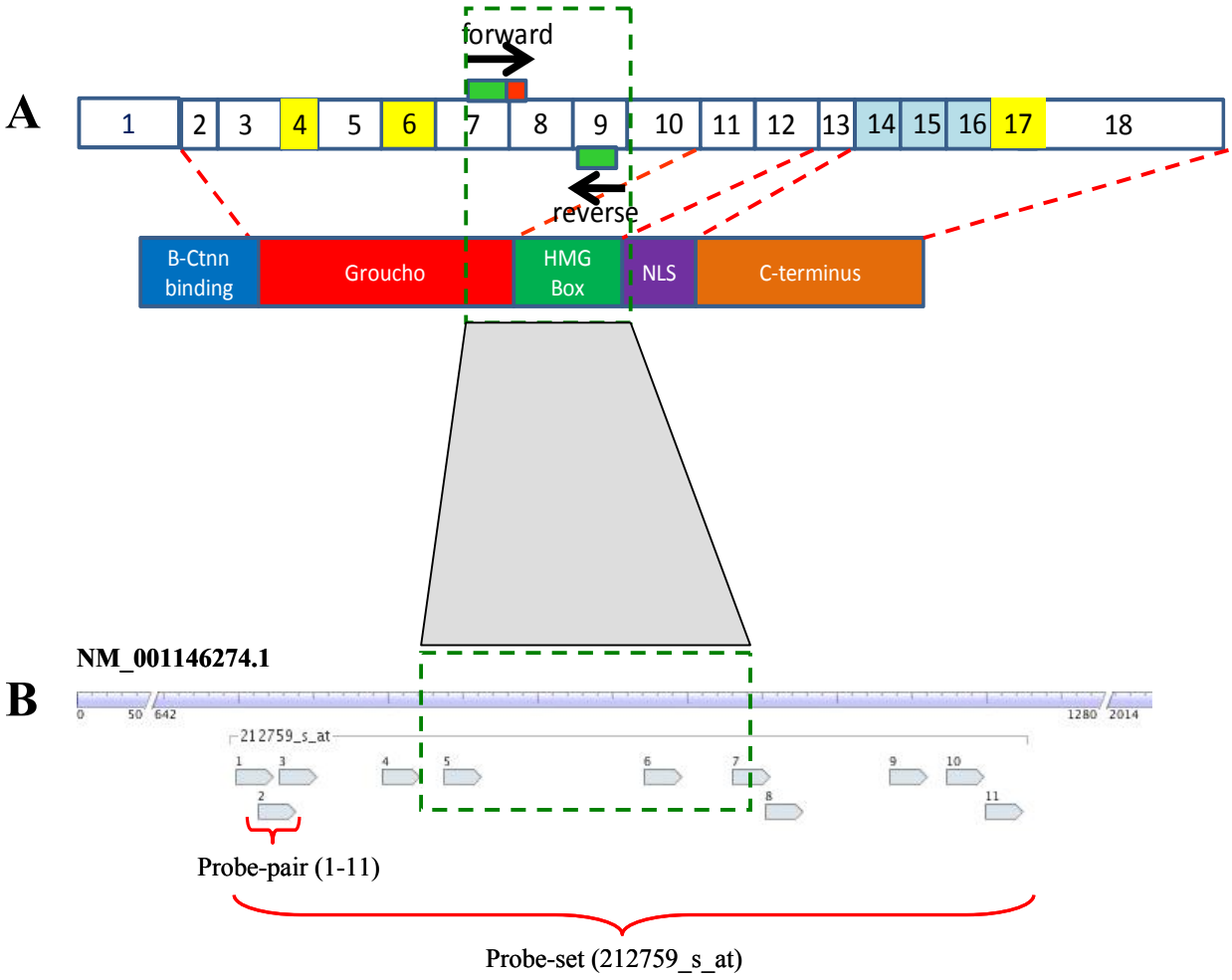


Figure 3.3: Location of primers and probe pairs within the *TCF7L2* probe set found to be statistically significant by microarray analysis. **(A)** Location of primer pairs used for *TCF7L2* validation by qRT-PCR. Schematic diagram of *TCF7L2* gene and protein domains encoded by the given exons, illustrating the position of the primer pair (forward and reverse primers) used for validation of microarray using qRT-PCR. The green and red forward primer denotes the exon-spanning primer. **(B)** Location of eleven probe pairs within the *TCF7L2* probe set. The probe set number for *TCF7L2* is 212759_s_at, and the sequence for each probe pair as in Table 3.2. NCBI RefSeq number for *TCF7L2* is NM_001146274.1.

Probe pair (no.)	Start position	End position	Exon no.	*(Perfect Match)* Probe sequence 5'-3'
212759_s_at (no.1)	969	993	5	GAAATGGCCACTGCTTGATGTCCAG
212759_s_at (no.2)	984	1008	5	TGATGTCCAGGCAGGGAGCCTCCAG
212759_s_at (no.3)	998	1022	5	GGAGCCTCCAGAGTAGACAAGCCCT
212759_s_at (no.4)	1067	1091	7	AAGTGCCAGTGGTGCAGCACCCTCA
212759_s_at (no.5)	1108	1132	7	ACGCCTCTTATCACGTACAGCAATG
212759_s_at (no.6)	1242	1266	8a/8b	ATCGCCTGGCACCGTAGGACAAATC
212759_s_at (no.7)	1301	1325	9	GTCAACCAGTGTACCCAATCACGAC
212759_s_at (no.8)	1323	1347	9	GACAGGAGGATTCAGACACCCCTAC
212759_s_at (no.9)	1406	1430	10a	CACATCATACGCTACACACGACGGG
212759_s_at (no.10)	1444	1468	10a	GCCATAGTCACACCAACAGTCAAAC
212759_s_at (no.11)	1470	1494	10a/b	GGAATCGTCCCAGAGTGATGTCGGC

Table 3.2: Probe pair sequences of the *TCF7L2* probe set 212759_s_at found to be significant in HG-U133 Plus2.0 analysis. Probe pair sequences were adapted from information published in the Annotation Database for Affymetrix Probesets and Transcripts (ADAPT) (Leong *et al*, 2005). NCBI RefSeq number for *TCF7L2* is NM_001146274.1

3.3.6 Measures to increase sample number for clinical outcome association studies in the AML cohort

3.3.6.1 Merging of HG-U133 Plus2.0 and HG-U133A array datasets

In order to increase statistical power for the purpose of clinical outcome analysis, it was necessary to merge the available GEP datasets from the two cohorts of AML array platforms that were available. Combining both HG-U133A and HG-U133 Plus2.0 microarray datasets resulted in better utilisation of the existing AML clinical outcome data available within Department of Haematology, Cardiff (n=246). The HG-U133 Plus2.0 contains 86 AML patients and the HG-U133A dataset provides an additional 160 AML patients GEP data. Non-intensive AML patients who previously received low dose Ara-C

treatment were excluded from analysis. The descriptions of the AML patients used for the merged data are shown in Table 3.1.

A list of common probe sets IDs between the two arrays was created. This resulted in a total of 22,277 probe-sets for downstream analysis. The .CEL files from the array with greater density (HG-U133 Plus2.0) were imported first into Partek® software and the probe-filtering option was applied to include only the common probe sets created earlier. GC-RMA background correction and quantile normalisation was applied. Subsequently, the second less dense probe GEP data was imported (HG-U133A) and similar settings were applied throughout. The presence of significant variation in GEP before and after merging was visualised by unsupervised hierarchical clustering. The GC-RMA normalised merged GEP was saved in a separate Partek® spreadsheet for further use.

3.3.6.2 Removal of cross-experimental batch effect using the DWD algorithm due to sample pooling

As described in 3.3.6.1, Partek® was able to merge and detect the presence of a residual batch effect. However, in order to compensate for systematic technical differences between different batches as detected, an additional approach (software) was required. This study used a Distance Weighted Discrimination (DWD) method. A requirement of DWD is that a large sample (batch) size is required for implementation due to the dependency on batch mean and variance estimates in sample sizes, preferably more than 40 samples for each batch. The software installation package for DWD ver1.0 was downloaded from https://cabig.nci.nih.gov/community/tools/DWD, developed by University of North Carolina Lineberger Comprehensive Cancer Center, USA. For the input dataset, the GC-RMA normalised values from the merged array were uploaded as a single tab-delimited text file into DWD software. The mean adjustment tool was applied to merge the two arrays simultaneously. This approach finds the best DWD hyper-plane and the data were adjusted by projecting different batches on the DWD plane (Benito *et al*, 2004; Hu *et al*, 2006). Multivariate projection views were used to illustrate the adjustment process using the DWD method. The final adjusted GEP generated by DWD was imported back into Partek® as a single tab-delimited text file for further analysis. The quality of this adjusted GEP was inspected using an exploratory tool (hierarchical clustering).

3.3.7 Correlation of TCF7L2 expression with clinical outcome in AML patients

The statistical analysis described in this section was kindly assisted by Dr Robert Hills from the Department of Haematology, Cardiff University. A total of 246 AML patient data sets (Table 3.1) were utilised for clinical and microarray correlational studies. The adjusted GEP generated by DWD for the expression of the *TCF7L2* gene (212759_s_at) obtained above (see 3.3.6) was selected as the input expression data. Patients who were not considered fit for intensive treatment and who had received non-intensive treatment, namely low dose Ara-C versus one of two novel treatments, which are low dose Clofarabine or Sapacitabine were removed from the dataset. This was done to remove noise and bias in the clinical outcome. Clinical outcome data have been collected over a period of 8 years for AML patients enrolled in the MRC AML studies.

Adjusted multivariate analysis of maximum likelihood estimates was performed to look for an association of *TCF7L2* as a prognostic independent factor. The following variables were included in the model: age, WBC count, cytogenetic prognostic groups, gender, WHO performance and *TCF7L2* gene expression. Proportional hazards Cox regression models on time-to-event endpoints was performed. Significance was set at $P < 0.05$ for all tests for multivariable analysis. All analyses were performed using SAS version 9.1.3 software (SAS Institute Inc.). WHO performance scores are detailed in Supplementary 3.4 (Oken *et al*, 1982).

3.4 RESULTS

3.4.1 Quality Control analysis identifies high quality GEP data

In order to analyse only high quality GEP data, a number of QC steps were performed. The initial RNA quality and integrity of AML blast samples were previously analysed by Amanda Gilkes, Department of Haematology, Cardiff. All samples hybridised to the GeneChips® were of high RNA quality as determined by Agilent Bioanalyser RNA Integrity number (RIN>8). The RNA integrity data included 5 µg of total RNA yielding at least 20 µg of biotinylated cRNA. Following fragmentation, the cRNA was electrophoresed on the Agilent Bioanalyser: fragment sizes were between 50 and 200 bp. According to MILE Working Group recommendations, samples that did not meet these criteria were discarded (Kohlmann *et al*, 2008a). As part of this study, a number of bioinformatics control steps were performed. Of the initial 155 GEP HG-U133 Plus2.0 GeneChip® data files that were available from Department of Haematology (not listed here), a total of 111 (72%) files passed all QC criteria as listed in 3.3.2.2. The excluded data files either had been deemed poor by visual inspection both by PLM weights and PLM residuals due to prominent artefact spots or uneven hybridisation (Figure 3.4A - Figure 3.4C).

As for the internal probe set QC, hybridisation controls showed increasing concentration as expected, with probe set 'AFFX-r2-Ec-bioB-avg' having the lowest concentration (Figure 3.5A). Additionally, the signal intensity ratio of the 3' probe set over the 5' probe set of the RNA-degradation control *GAPDH* transcript was less than 3 (indicative of good data) shown in Figure 3.5B, after removal of bad QC files. Most files showed poly-A signals with acceptable linearity; where the signal for probe set 'AFFX-r2-Bs-dap-avg' is present at a much higher concentration as expected than the other poly-A probe sets (Figure 3.5C). This spiking test of internal probe sets verified that there was no bias during the reverse-transcription between highly expressed genes and low expressed genes during sample preparation. In summary, QC analysis has identified a final dataset comprising of 98 GEPs of AML patients (Table 3.1) and 13 GEPs of Normal Control for RNA hybridised to the HG-U133 Plus2.0 GeneChip®. Similarly, the QC of HG-U133A had been previously validated in earlier studies using a similar AML cohort (Tonks *et al*, 2007a; Tonks *et al*, 2007b; Liddiard *et al*, 2010) with only 68% passing QC from the total HG-U133A samples available; the final 160 AML samples selected are also shown in Table 3.1.

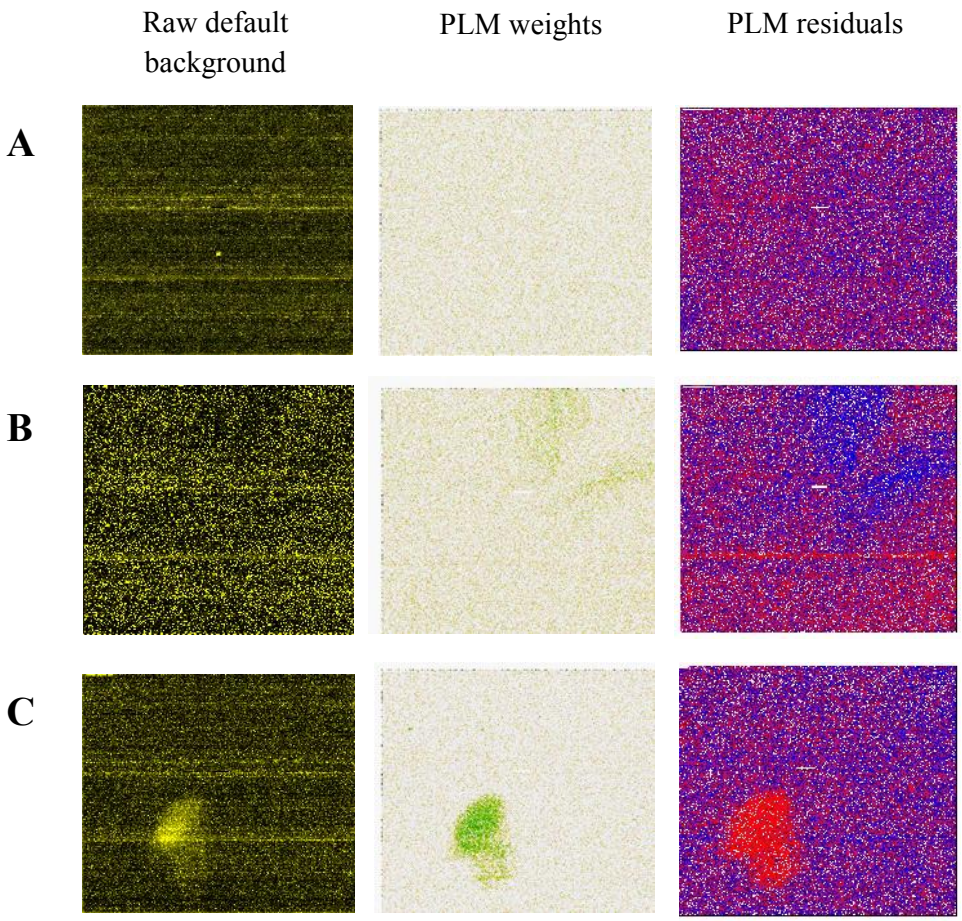


Figure 3.4 : Visual inspection of GEP data using 2-D pseudo-colour analysis. Each GEP data set was visually analysed for artefacts. **(A)** AML GEP with acceptable QC. **(B)** Poor AML GEP QC with uneven hybridisation and artefact 'line'. **(C)** Poor AML GEP QC with bright artefact spot. Note that the PLM residuals will allow a better view of the spurious areas.

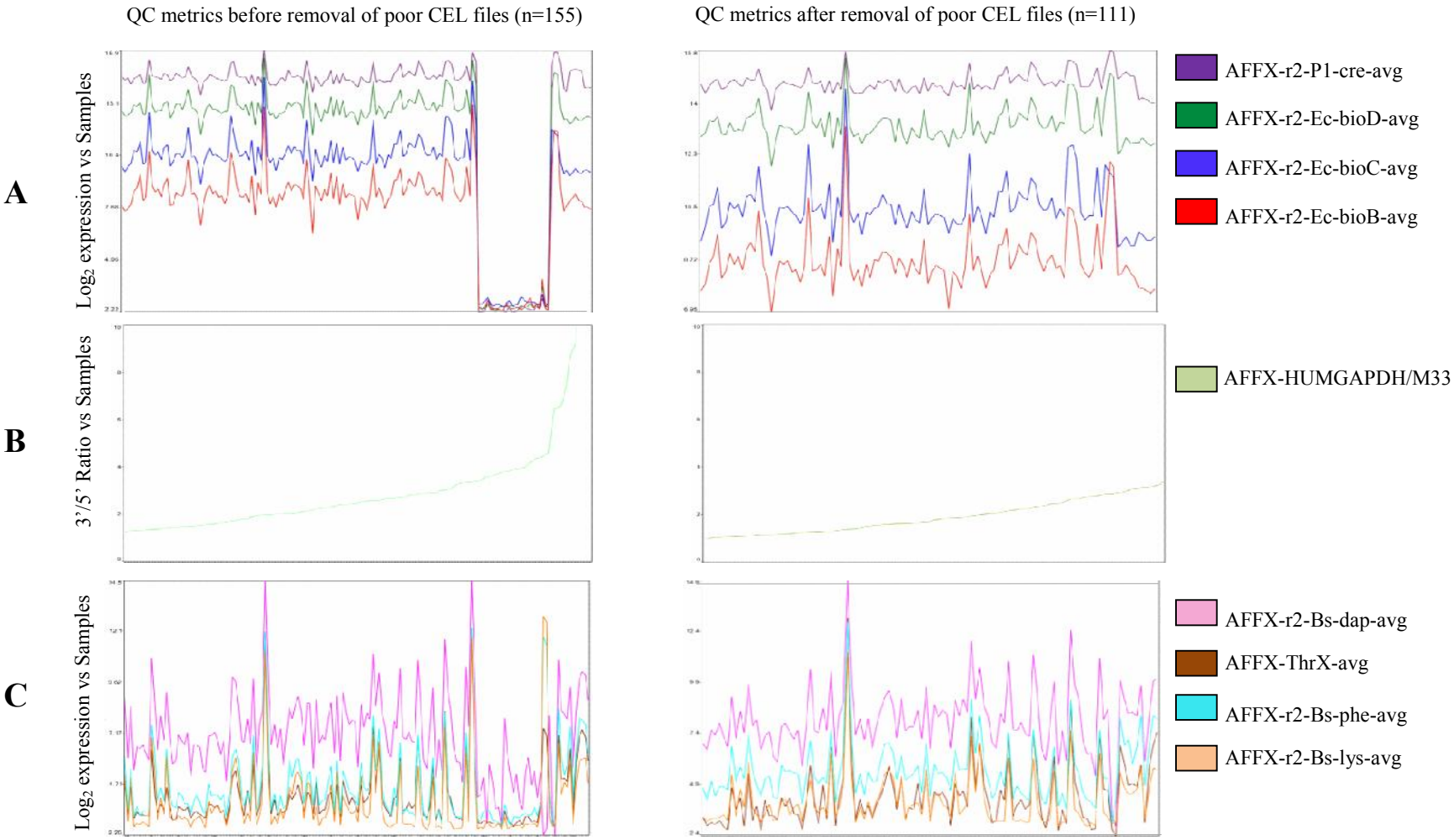


Figure 3.5 : Profiles of AFFX control probe sets before and after removal of poor quality GeneChips®. (A) Hybridisation controls showing increasing concentration for good QC. **(B)** RNA degradation signal showing 3'/5' ratio of *GAPDH* to be less than 3 for good QC. **(C)** Poly-A signals showing doubling of amplification and 'AFFX-r2-Bs-dap-avg' probe set having highest signal for good QC.

3.4.2 Exploratory data analysis discriminates AML from normal GEP

PCA mapping helps to reduce the dimensionality of the data set and to visualise underlying similarities capturing most of the variance in the original GEP. The analysis is based on all probe sets represented on the HG-U133 Plus2.0 microarray. As shown in Figure 3.6, PCA was able to detect several AML samples that have quite a distinct expression profile (dots outside the elipsoids), however this was a very small percentage of the overall AML cases analysed (6.1%). No obvious outlier was detected within the Normal Controls although they were derived from different donors taken from two different external experiments. The AML samples also clustered tightly regardless of the origin of sample type taken during diagnosis, either BM or PB. PCA exploratory output indicates that the disease factor (AML *vs* Normal Control) has a major effect on the data as expected and should be a fundamental factor to be included in the subsequent ANOVA model along with other factors described in the next section.

3.4.3 Finding differentially expressed gene changes between AML FAB subtypes and normal controls

In order to detect biological changes in the GEP data, normalisation was first applied to the raw intensity data followed by analysis of variance using ANOVA. Normalisation or log transformation of the GEP using GC-RMA (3.3.2.3) produced normally distributed and symmetrical GEP data as shown in Figure 3.7. In order to find differentially expressed genes, the normalised data was subjected to further variance analysis and this study employed several strategies of ANOVA to find differentially expressed genes as detailed in 3.3.3.

From the ANOVA model output (Figure 3.8) and consistent with the PCA, the AML *vs* Normal Control (or the "disease" factor) is still among the largest source of variation in these data with a signal to noise ratio of 1.23 followed by FAB having a higher variation signal of 1.82 above noise. This means that the "disease" (AML *vs* Normal Controls) and FAB subtypes are the factors that contributed to the variability in the gene changes explained by the ANOVA model. The scan date (signal to noise: 1.13) also contributed some degree of variation in the GEP whilst the sample type (signal to noise: 0.70) variability is not explained by the model used. By taking into account all the different factors, the heterogeneity in the GEP is highest between different FAB subtypes and contributed significant gene expression changes. Thus subsequent genelists were derived

based on FAB subtypes. As previously illustrated in Figure 3.1, the 4-way ANOVA can only compute one fold change value representing all FAB subtypes *vs* Normal Controls (genelist C), therefore 1-way ANOVA was added for additional fold changes to be computed separately for AML-M0 *vs* Normal Controls and AML-M1 *vs* Normal Controls (genelists A and B).

In summary, grouping all AML FAB subtypes into one group may result in obscuring gene changes due to the developmental status of the AML blasts. Therefore additional analyses were performed using FAB subtypes M0 and M1 since these are developmentally closest to the $CD34^+$ derived Normal Controls used. All three genelists A, B, C generated from whole genome ANOVA analysis were utilised for further enrichment analysis presented in 3.4.4.

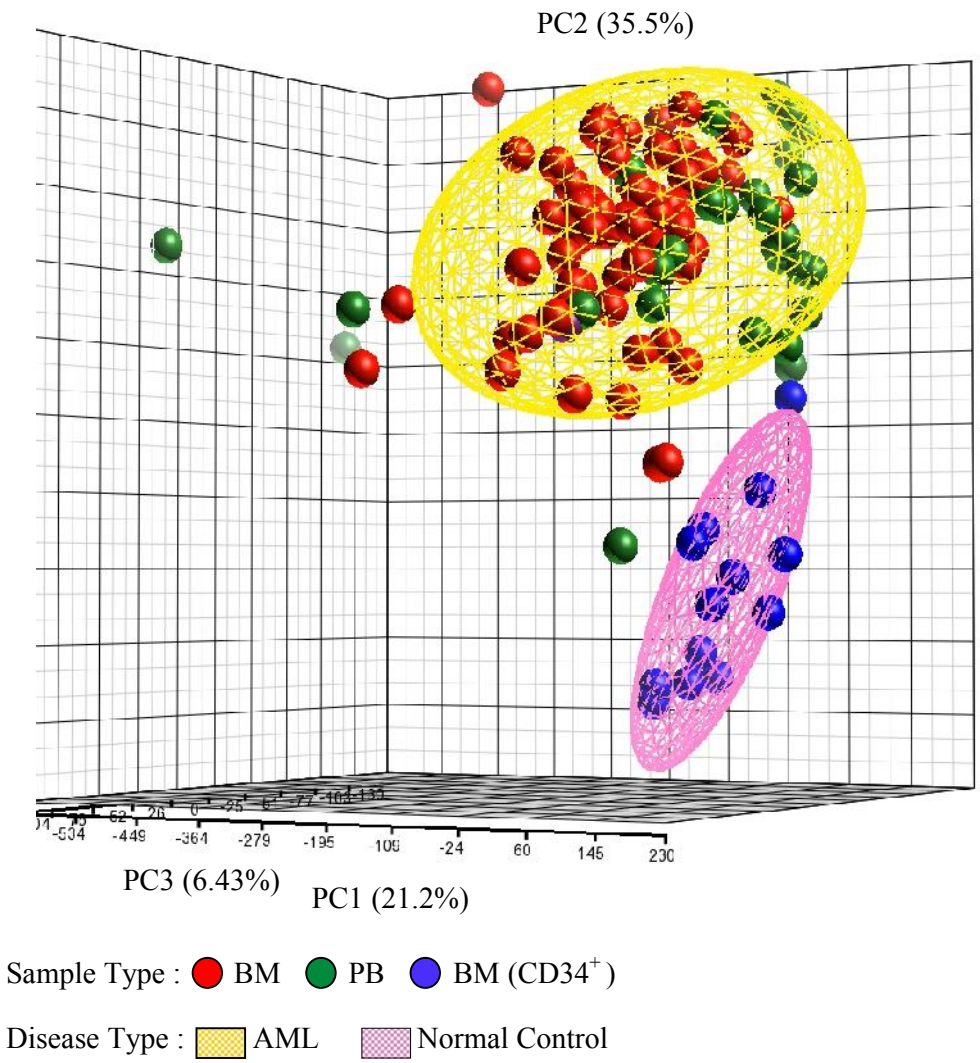


Figure 3.6 : Unsupervised PCA mapping of 98 AML samples and 13 Normal Controls derived from CD34$^+$ BM samples, showing 54,675 probe sets in a high-dimensional gene space. Individual GEP (dots) are coloured based on the sample type taken at diagnosis. Ellipsoids are drawn around each group and capture the underlying similarities of the AML samples and of the Normal Control samples.

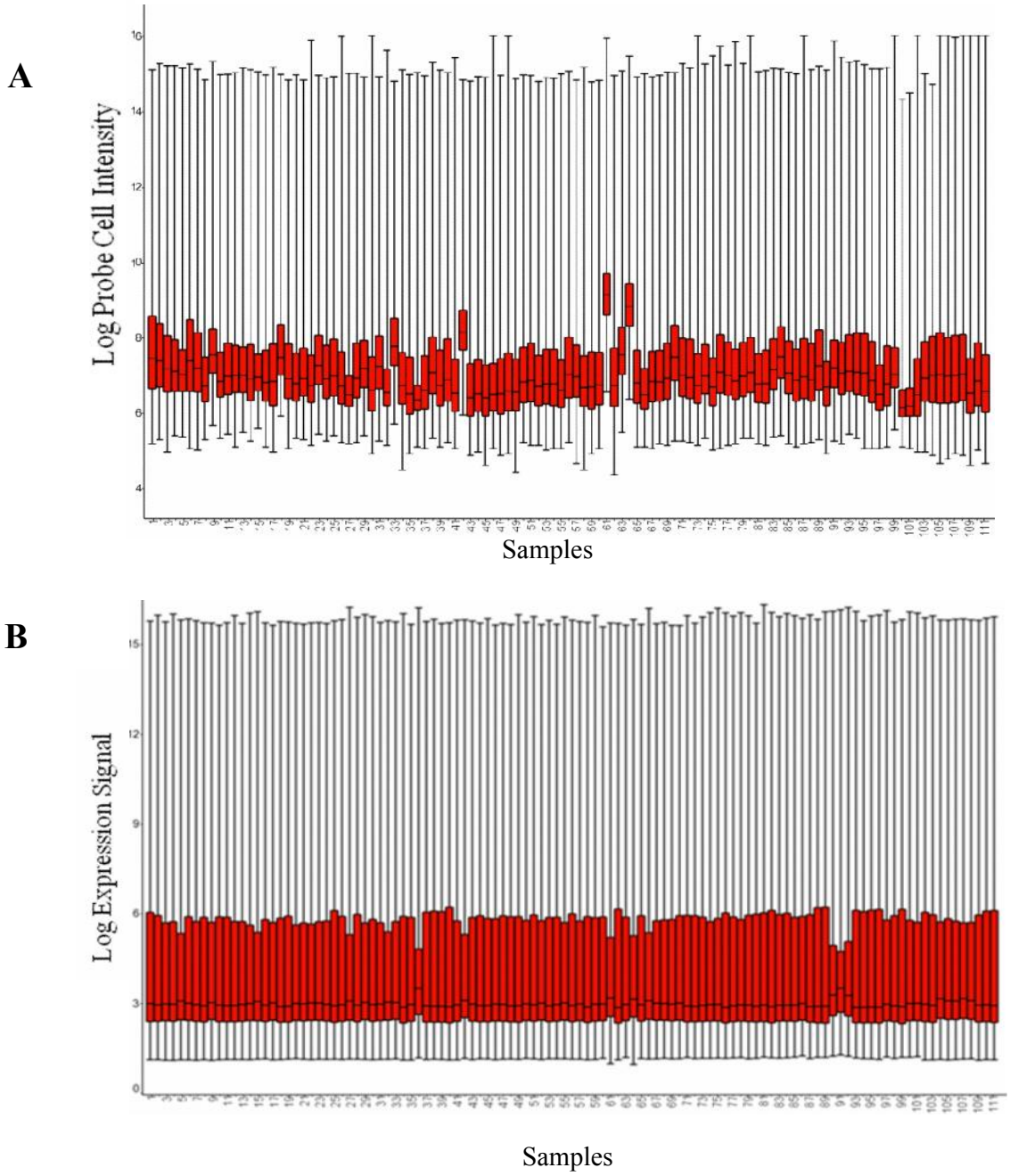


Figure 3.7 : Box plots showing pre and post normalisation performance of 111 GEP data. (A) Log Probe Cell intensity showing GEP distribution before normalisation **(B)** Log expression signal after GC-RMA normalisation showing a more stable median.

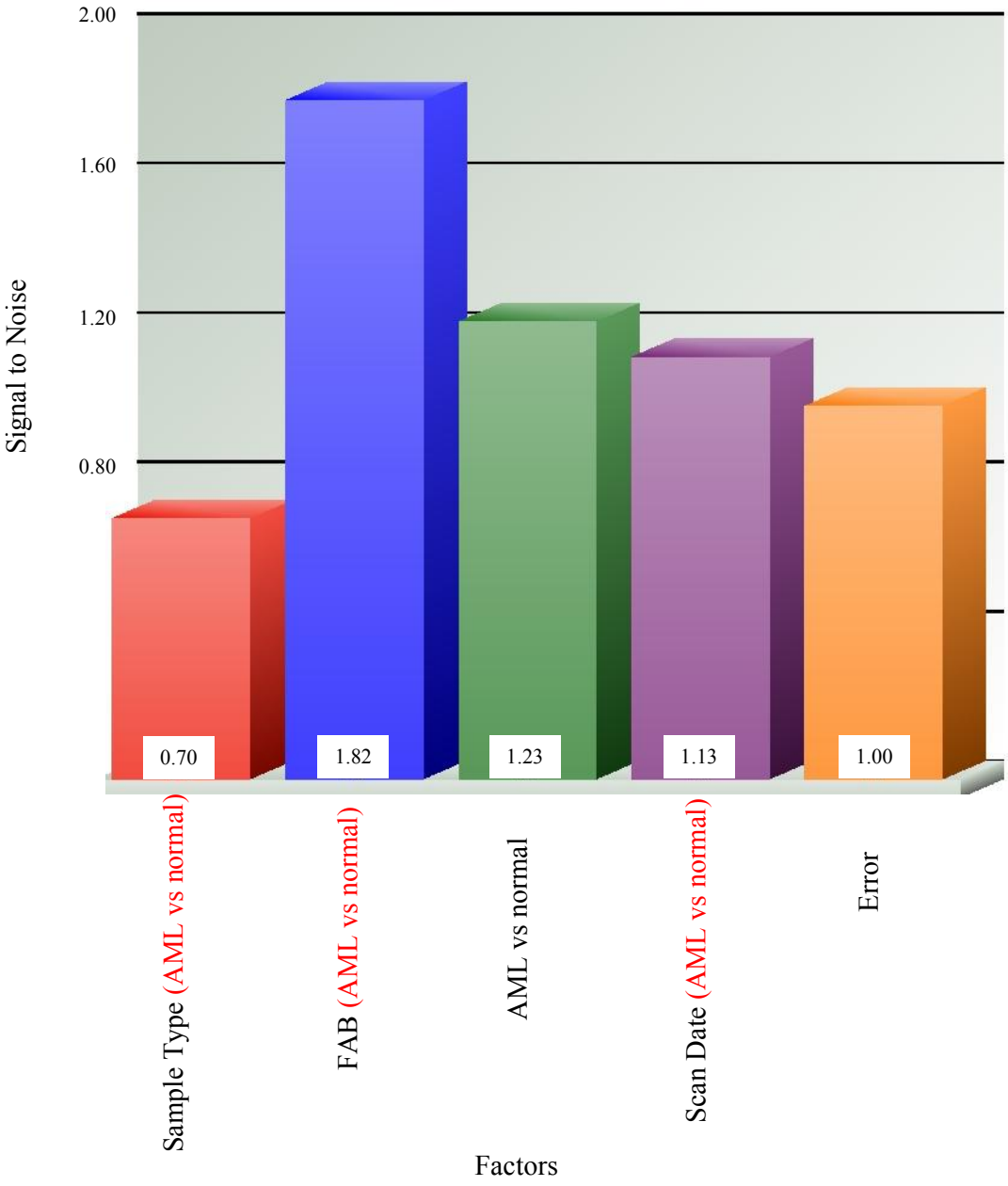


Figure 3.8 : Source of variation plot of the GC-RMA normalised HG-Plus2.0 dataset using mixed model 4-way ANOVA. The height of each bar indicates the average signal variance (mean square) for all 54,675 probe sets. The signal to noise ratio or mean F ratio to error noise (1.00) is shown for each factor in the model. The nested factor is shown in red. Signal to noise is variation contributed by each factor relative to error in the model. Error is the amount of variability not explained by mixed model ANOVA.

3.4.4 Enrichment analysis identifies WNT signalling as the most dysregulated pathway in AML

The mixed model ANOVA resulted in three genelists for subsequent analysis. These genelists contain GC-RMA expression value, fold change and *P*-value for each gene that passed the threshold (see 3.3.3). The three genelists and corresponding number of probe sets passing the criteria are as follow; genelist A (2769 probe sets), genelist B (8136 probe sets) and genelist C (4695 probe sets), illustrated in Figure 3.9A. Given the challenges of interpreting such large gene lists containing thousands of probe sets, it was necessary to incorporate bioinformatics tools in the analysis to systematically extract biological themes behind large gene lists produced by ANOVA. One particular interest is to apply significant gene list produced by ANOVA for pathway analysis to determine the relevance of gene expression within a signalling pathway. Genelists A, B and C were uploaded into MetaCore^{TM} analytical suite and analysed for significant deregulated pathways (see 3.3.4.1 for MetaCore^{TM} functional analysis; and results in Table 3.3). Although the different AML subtypes might differ from each other, understanding how they share common pathway characteristics should reveal fundamental information about AML pathogenesis. The top two significant pathways common to all three genelists were genes associated with "cytoskeletal remodelling" (a GeneGo term to describe cancer cells progression) and "WNT, TGF signalling". The pathway associated with "WNT, TGF signalling" was found to be highly significant in all comparisons and highest in AML-M0 *vs* Normal Control ($P = 8.572^{-16}$). The *P*-value of the intersection between a gene of an experimental dataset (such as from a genelist) and the same gene in a particular ontology is considered as a measure of the relevance of pathway analysis. In MetaCore™, *P*-value means the probability of a random intersection of genes in two different sets. The lower the *P*-value, the higher the "non-randomness" of finding such an intersection. In summary, global enrichment analysis by using Pathway Maps has identified WNT signalling to be perturbed in AML compared to Normal Control.

#	*AML-M0 vs Normal Control (Genelist A)*		*AML-M1 vs Normal Control (Genelist B)*		*Overall AML-M0:M6 vs Normal Control (Genelist C)*	
	Map title	***P-Value***	***Map title***	***P-Value***	***Map title***	***P-Value***
1	Cytoskeleton remodeling_ WNT,TGF	8.57^{-16}	Cytoskeleton remodeling	8.32^{-19}	Cytoskeleton remodeling	1.42^{-16}
2	Cytoskeleton remodeling	1.17^{-14}	Cytoskeleton remodeling_WNT,TGF	6.25^{-17}	Cytoskeleton remodeling_ WNT,TGF	2.94^{-15}
3	Cell cycle_Influence of Ras and Rho proteins on G1/S Transition	2.77^{-13}	Transcription_Sin3 and NuRD in transcription regulation	3.58^{-15}	Apoptosis and survival_p53-dependent apoptosis	2.68^{-12}
4	Immune response_IL-2 activation and signaling pathway	1.04^{-11}	Cell cycle_Influence of Ras and Rho proteins on G1/S Transition	3.85^{-13}	DNA damage_ATM/ATR regulation of G1/S checkpoint	5.59^{-12}
5	Cell cycle_Chromosome condensation in prometaphase	1.60^{-10}	Apoptosis and survival_p53-dependent apoptosis	5.32^{-10}	Cell cycle_Influence of Ras and Rho proteins on G1/S Transition	3.42^{-11}
6	Immune response_Fc epsilon RI pathway	1.17^{-8}	Transcription_P53 signaling pathway	2.04^{-11}	DNA damage_Role of Brca1 and Brca2 in DNA repair	1.07^{-10}
7	Development_VEGF signaling via VEGFR2 - generic cascades	8.60^{-8}	Cell cycle_Chromosome condensation in prometaphase	2.33^{-11}	Cell cycle_Chromosome condensation in prometaphase	1.47^{-10}
8	Transcription_P53 signaling pathway	1.72^{-7}	Development_VEGF signaling via VEGFR2 - generic cascades	8.25^{-11}	Cell cycle_The metaphase checkpoint	2.34^{-10}
9	Cell cycle_The metaphase checkpoint	2.21^{-7}	Cell cycle_The metaphase checkpoint	8.72^{-11}	Cell adhesion_Chemokines and adhesion	3.19^{-10}
10	DNA damage_ATM/ATR regulation of G1/S checkpoint	1.05^{-6}	Immune response_Fc epsilon RI pathway	9.09^{-11}	Development_GM-CSF signaling	1.39^{-09}

Table 3.3 : Statistically significant Pathway Maps generated from each genelist. Sorting was performed using FDR of < 0.05. The *P* -value calculation is the number of probe sets in the list that hit a given pathway map as compared to pure random chance.

3.4.5 Probe level filtering identifies dysregulated WNT signalling genes in AML

The data above (3.4.4) suggests WNT signalling is commonly dysregulated in AML. By mapping and compiling all genes known to be associated with WNT signalling (i.e. upstream and downstream regulatory genes and target genes) a more focused strategy was followed to facilitate the selection of the most dysregulated WNT associated gene(s) within AML. This was achieved by performing a cross validation analysis using an ANOVA approach as described in 3.3.3 and also illustrated in Figure 3.2; from here a further three genelists (D, E and F) were generated (Figure 3.9B). Two major observations were obtained from this approach. Firstly, all genelists (A – F) shared 74% similarity in terms of WNT genes that overlap, giving rise to the 14 common probe sets (Figure 3.9) and their corresponding fold changes are listed in Table 3.4 - Table 3.6. This shows that the output from the ANOVA model employed was reproducible even after reduction in the number of input probe sets. Secondly, the production of WNT genelists (D, E and F) has made it possible to search for significant interaction of different classes of genes (and proteins) within WNT signalling that might be important in AML, as described below.

The Interactome analysis of this set of genes (from genelist D, E and F) was conducted according to 3.3.4.2 using the "MetaCore™ enrichment by protein function" tool. This tool gives enrichment of gene expression changes based on different protein classes. From this, a series of protein rankings (similar to probe sets) and their level of interaction with other proteins present across all genelists were achieved. Table 3.7 lists 11 transcription factors, 2 receptors, 2 ligands, 8 kinases, 2 phosphatases, enzyme and other proteins which were identified to be significantly 'overconnected'. The protein (identified by probe set) with common agreement across all genelists and the most connected with the highest z-score was found to be a TF gene known as *TCF7L2* (z-score of 19.87). A representative image of a network showing direct interaction of *TCF7L2* with other genes in various cellular compartments is illustrated in Supplementary 3.5 (built using MetaCore™ Direct Interaction Map). Furthermore, *TCF7L2* was consistently found to be over-expressed in all other type of comparisons (Table 3.4 - Table 3.6), having the highest fold differences in AML-M0 *vs* Normal Control (10-fold). Following this collective observation and the potential role of a TF as a central nuclear regulator involved in transcriptional dysregulation of WNT signalling, *TCF7L2* was selected for further investigations.

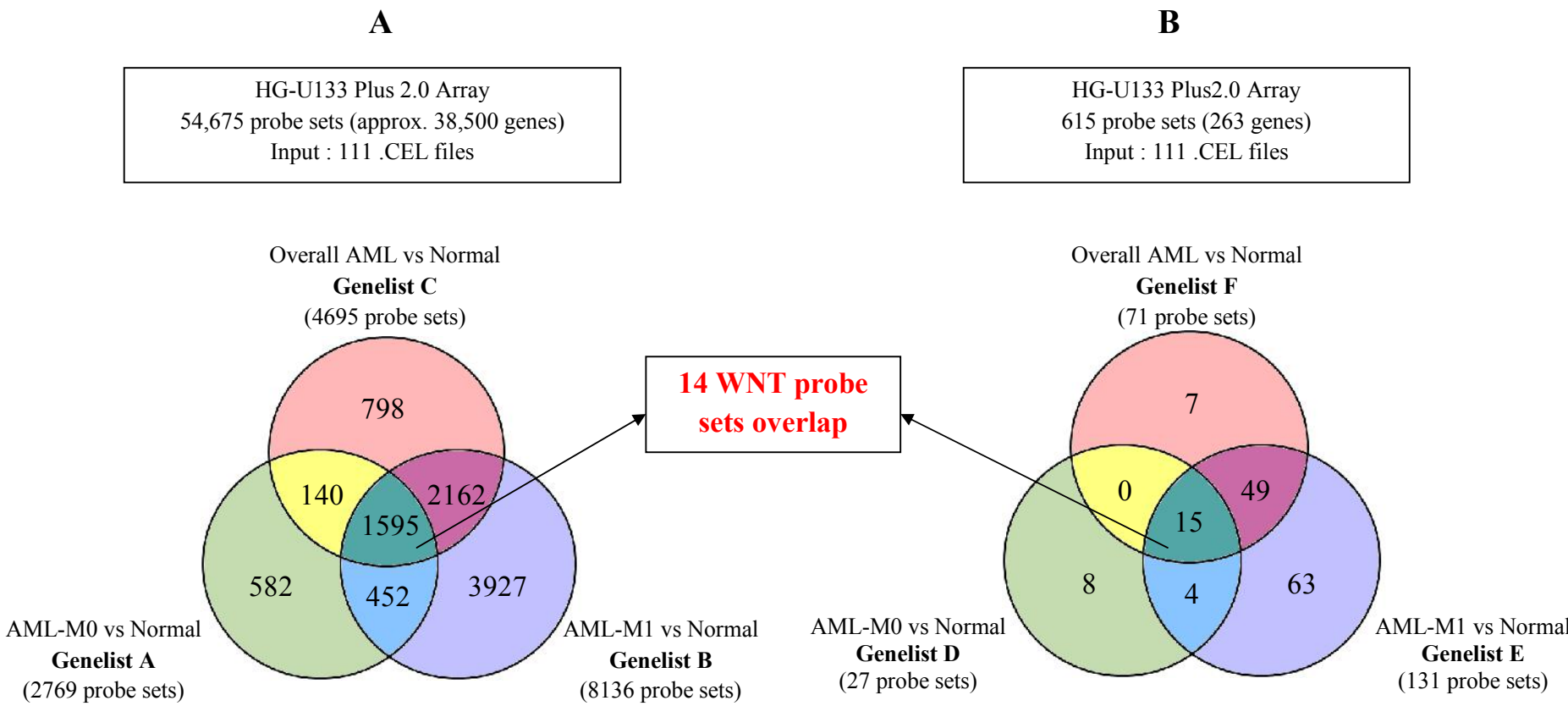


Figure 3.9 : Venn diagram showing significant genelists for both whole-genome and 'WNT-ome' comparisons (A) The intersection showing 1595 probe sets representing various genes from various pathways **(B)** The intersection showing 15 probe sets representing only WNT pathway genes. The number of probe sets found to be significant in each of the genelist is shown in parentheses. All genelists passed the 1.5 fold change threshold and *P*-value <0.05. Each genelist was generated by the Partek® List Manager tool.

Probe set ID	Gene symbol	*P* -value	Fold change	Description
203895_at	*PLCB4*	6.19^{-07}	-6.77	M0 down vs Normal Control
203896_s_at	*PLCB4*	2.08^{-07}	-4.35	M0 down vs Normal Control
212669_at	*CAMK2G*	1.12^{-05}	5.3	M0 up vs Normal Control
204489_s_at	*CD44*	5.46^{-06}	5.78	M0 up vs Normal Control
1567458_s_at	*RAC1*	5.52^{-05}	6.08	M0 up vs Normal Control
207993_s_at	*CHP*	0.0003891	7.24	M0 up vs Normal Control
214083_at	*PPP2R5C*	0.0005686	7.62	M0 up vs Normal Control
1557905_s_at	*CD44*	8.15^{-06}	9.93	M0 up vs Normal Control
204490_s_at	*CD44*	4.71^{-06}	9.93	M0 up vs Normal Control
209835_x_at	*CD44*	3.03^{-06}	9.96	M0 up vs Normal Control
212759_s_at	*TCF7L2*	1.03^{-05}	9.96	M0 up vs Normal Control
212014_x_at	*CD44*	3.51^{-06}	10.61	M0 up vs Normal Control
1555814_a_at	*RHOA*	5.84^{-05}	10.92	M0 up vs Normal Control
210916_s_at	*CD44*	2.26^{-07}	21.75	M0 up vs Normal Control

Table 3.4 : The common 14 dysregulated WNT probe sets identified in AML-M0 *vs* Normal Controls built from genelist D. Genelist D contains a total of 27 probe sets and shown here are the common 14 WNT probe sets overlapping between Genelist A-F. Each probe set was arranged based on the fold change from lowest to highest. Genes are significant at $P < 0.05$.

Probe set ID	Gene Symbol	*P* -value	Fold-Change	Description
203895_at	*PLCB4*	2.81^{-14}	-6.78	M1 down vs control
203896_s_at	*PLCB4*	9.50^{-14}	-3.93	M1 down vs control
212669_at	*CAMK2G*	3.01^{-09}	4.1	M1 up vs control
212759_s_at	*TCF7L2*	4.50^{-06}	4.23	M1 up vs control
204489_s_at	*CD44*	1.81^{-11}	5.27	M1 up vs control
1567458_s_at	*RAC1*	1.06^{-10}	6.4	M1 up vs control
1555814_a_at	*RHOA*	4.23^{-08}	7.62	M1 up vs control
1557905_s_at	*CD44*	9.54^{-11}	8.33	M1 up vs control
214083_at	*PPP2R5C*	1.15^{-08}	8.43	M1 up vs control
209835_x_at	*CD44*	1.04^{-11}	8.54	M1 up vs control
212014_x_at	*CD44*	9.67^{-12}	9.25	M1 up vs control
204490_s_at	*CD44*	4.14^{-12}	9.42	M1 up vs control
207993_s_at	*CHP*	3.51^{-10}	9.53	M1 up vs control
210916_s_at	*CD44*	4.04^{-13}	16.07	M1 up vs control

Table 3.5 : The common 14 dysregulated WNT probe sets identified in AML-M1 *vs* Normal Controls built from genelist E. Genelist E contains a total of 131 probe sets and shown here are the common 14 WNT probe sets overlapping between Genelist A-F**.** Each probe set was arranged based on the fold change from lowest to highest. Genes are significant at $P < 0.05$.

Probe set ID	Gene Symbol	*P* -value	Fold Change	Description
203895_at	*PLCB4*	3.94^{-06}	-4.29	AML down vs normal
203896_s_at	*PLCB4*	1.42^{-05}	-2.74	AML down vs normal
212669_at	*CAMK2G*	0.00059	2.75	AML up vs normal
212759_s_at	*TCF7L2*	0.00348	3.18	AML up vs normal
204489_s_at	*CD44*	1.43^{-05}	3.72	AML up vs normal
1557905_s_at	*CD44*	0.00029	4.48	AML up vs normal
1555814_a_at	*RHOA*	0.00179	4.54	AML up vs normal
209835_x_at	*CD44*	8.07^{-05}	4.61	AML up vs normal
1567458_s_at	*RAC1*	1.07^{-05}	4.81	AML up vs normal
212014_x_at	*CD44*	7.62^{-05}	4.95	AML up vs normal
204490_s_at	*CD44*	2.25^{-05}	5.5	AML up vs normal
207993_s_at	*CHP*	2.67^{-05}	6.19	AML up vs normal
214083_at	*PPP2R5C*	1.64^{-05}	6.84	AML up vs normal
210916_s_at	*CD44*	1.31^{-05}	7.8	AML up vs normal

Table 3.6 : The common 14 dysregulated WNT probe sets identified in overall AML (M0-M6) *vs* Normal Controls built from genelist F. Genelist F contains a total of 71 probe sets and shown here are the common 14 WNT probe sets overlapping between Genelist A-F. Each probe set was arranged based on the fold change from lowest to highest. Genes are significant at $P < 0.05$.

Gene type	Network gene name	*P*-value	z-score	Input probe set IDs	Fold change agreement (D/E/F)	Interactome agreement (D/E/F)
Transcription factors	*TCF7L2*	7.56^{-25}	19.87	212759_s_at	↑↑↑	Yes
Transcription factors	*JUN*	1.64^{-20}	13.97	201465_s_at	↑↑↑	
Transcription factors	*LEF1*	3.52^{-15}	14.55	221558_s_at	↓↓↓	
Transcription factors	*SMAD4*	1.12^{-13}	11.32	235725_at	↓↓↓	
Transcription factors	*SMAD2*	2.08^{-13}	11.89	226563_at	↓↓↓	
Transcription factors	*SNAIL1*	4.42^{-12}	13.74	219480_at	↑↑↑	
Transcription factors	*NFATC4*	1.39^{-08}	12.65	213345_at	↑↑↑	
Transcription factors	*TCF4/ITF2*	9.66^{-08}	12.23	212382_at	↓↓↓	
Transcription factors	*TCF12*	4.50^{-04}	6.13	208986_at	↓↓↓	
Transcription factors	*NFATC3*	9.80^{-04}	6.38	210555_s_at	↓↓↓	
Transcription factors	*NFATC1*	2.37^{-03}	4.27	211105_s_at	↓↓↓	
Receptors	*LRP6*	4.63^{-10}	14.42	225745_at	↑↑↓	
Receptors	*CD44*	2.53^{-08}	10.20	204490_s_at; 212014_x_at	↑↑↑	Yes
Ligands	*CYR61*	5.88^{-10}	14.20	210764_s_at	↓↑↑	
Ligands	*IL8*	1.35^{-09}	12.31	211506_s_at	↑↑↑	
Kinases	*GSK3B*	1.53^{-20}	17.71	242336_at; 226183_at; 226191_at	↑↓↓	Yes
Kinases	*NLK*	1.08^{-09}	15.29	222589_at	↓↓↓	
Kinases	*PRKCD*	2.67^{-08}	9.59	202545_at	↑↑↑	
Kinases	*MAPK9*	1.85^{-07}	10.32	225781_at	↓↓↓	
Kinases	*MAP3K7*	6.36^{-06}	7.61	206854_s_at	↓↓↓	
Kinases	*CAMK2G*	2.22^{-05}	9.46	212669_at	↑↑↑	Yes
Kinases	*PRKCE*	4.86^{-04}	5.07	226101_at	↓↓↓	
Kinases	*ROCK1*	2.28^{-02}	3.67	230239_at	↑↑↑	
Phosphatases	*PPP2R5C*	3.38^{-08}	11.81	214083_at	↑↑↑	Yes
Phosphatases	*PPP2R5B*	2.93^{-07}	16.22	204611_s_at	↑↑↑	
Enzymes	*EP300*	1.62^{-24}	16.18	202221_s_at	↑↑↑	
Other	*DVL-1*	1.21^{-22}	26.50	203230_at	↑↑↑	
Other	*CTBP2*	1.98^{-09}	12.02	201218_at	↑↑↑	
Other	*ABCB1*	2.09^{-08}	12.26	209994_s_at	↓↓↓	
Other	*RAC1*	3.14^{-07}	7.84	1567458_s_at	↑↑↑	Yes
Other	*PPP2R1B*	8.50^{-07}	10.05	202883_s_at	↓↓↓	
Other	*TBL1XR1*	1.47^{-06}	10.92	222634_s_at; 221428_s_at; 223013_at	↑↑↓	Yes
Other	*CUL1*	3.09^{-06}	8.13	207614_s_at	↓↓↓	
Other	*CDKN2A*	4.20^{-06}	7.90	209644_x_at	↑↑↑	
Other	*RHOA*	5.95^{-06}	7.19	1555814_a_at	↑↑↑	Yes
Other	*DAAM1*	7.62^{-06}	10.89	232552_at	↓↓↓	Yes
Other	*CCND2*	7.82^{-06}	8.10	200951_s_at	↓↓↓	
Other	*TLE3*	6.50^{-04}	6.90	206472_s_at	↑↑↑	
Other	*ABCB1*	3.21^{-03}	6.60	209993_at	↓↓↓	

Table 3.7 : Significant enrichment by protein functions in agreement with changes seen in all three genelists D, E and F. A gene (probe set) is considered overconnected if the corresponding z-score is more or equal to 0. *P*-value was calculated using the basic formula for a hypergeometric distribution. Fold change agreement shows upregulation or downregulation of a gene in each genelist dataset. Interactome agreement indicated by "Yes" means the gene is significantly overconnected across all 3 genelists datasets (interaction within < 3 datasets highlighted in black).

3.4.6 Correlation of TCF7L2 expression using Affymetrix microarray profiling and qRT-PCR

Microarray and Interactome analyses led to identification of *TCF7L2* as a potential key regulator in WNT signalling that may play important role in AML. Validation of *TCF7L2* mRNA expression obtained by microarray was necessary prior to any further functional validation. To achieve this, qRT-PCR was performed according to the method described in 3.3.5. This study focused on the mRNA region of *TCF7L2* represented on the array by probe set (212759_s_at) that was found to be significantly expressed in AML compared to Normal Controls. The expression of *TCF7L2* by microarray and qRT-PCR showed a significant degree of correlation (R=0.748, P <0.01). In conclusion, this study suggests that the observed changes in *TCF7L2* expression by microarray are reliable and successfully validated by an independent method.

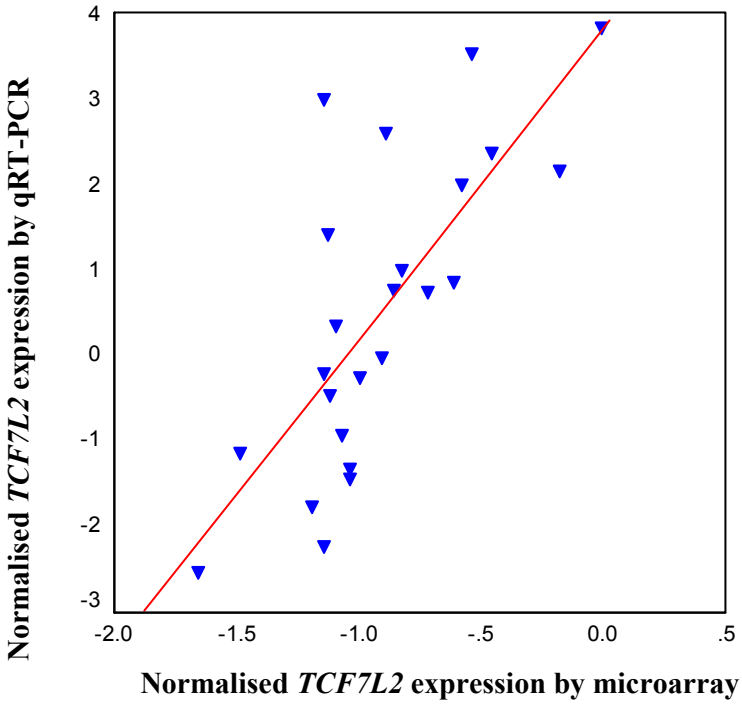


Figure 3.10 : Validation of *TCF7L2* mRNA expression using qRT-PCR. Correlation was shown between Log_2 array (GC-RMA normalised) and Log_2 qRT-PCR (deltaCt normalised) expression data. Correlation is significant at the 0.01 level (2-tailed) using Pearson's test.

3.4.7 Correlation of TCF7L2 expression and clinical outcome in a larger cohort of AML

3.4.7.1 Exploratory analysis showed a significant reduction in the source of batch effects after DWD adjustment in AML cohort

To increase the statistical power of the GEP dataset associated with clinical outcome, this study combined GeneChip® GEP raw data from HG-U133 Plus2.0 and HG-U133A into a single analysis. Due to the complexity in studying genes distributed in different experiments, background signal corrections were performed as described in the methods (3.3.6.1). Apart from signal normalisation, it is also possible to use other statistical methods such as DWD to remove unwanted batch effects (that can't be removed by normalisation); in this case the batch effect is GeneChip® platforms (i.e. HG-U133 Plus2.0 and HG-U133A). Prominent variation in GEP data was observed when the GeneChip® platform ('batch') was taken into account within the AML datasets. By finding the most fitted hyper-plane of the data, DWD was able to significantly reduce the variation between platform batches and the adjustment step was visualised by multivariate view as in Figure 3.11. This was also supported by unsupervised hierarchical clustering (Figure 3.12) which identified "batch' as the factor influencing the distribution of data suggesting that GEP data were initially clustered due to technical effects rather than biological effects even after normalisation.

The 'batch effect' was successfully removed following DWD adjustment supported by hierarchical clustering of DWD adjusted data. Figure 3.13 shows all AML samples from both GeneChip® platforms (red bar) are now clustered together. In summary, DWD adjustment helped to reduce substantial background noise and allowed creation of a single GEP dataset from the two different platforms. The expression value of *TCF7L2* gene derived from DWD adjusted merged GEP dataset (n=246) was then used for further clinical outcome analysis.

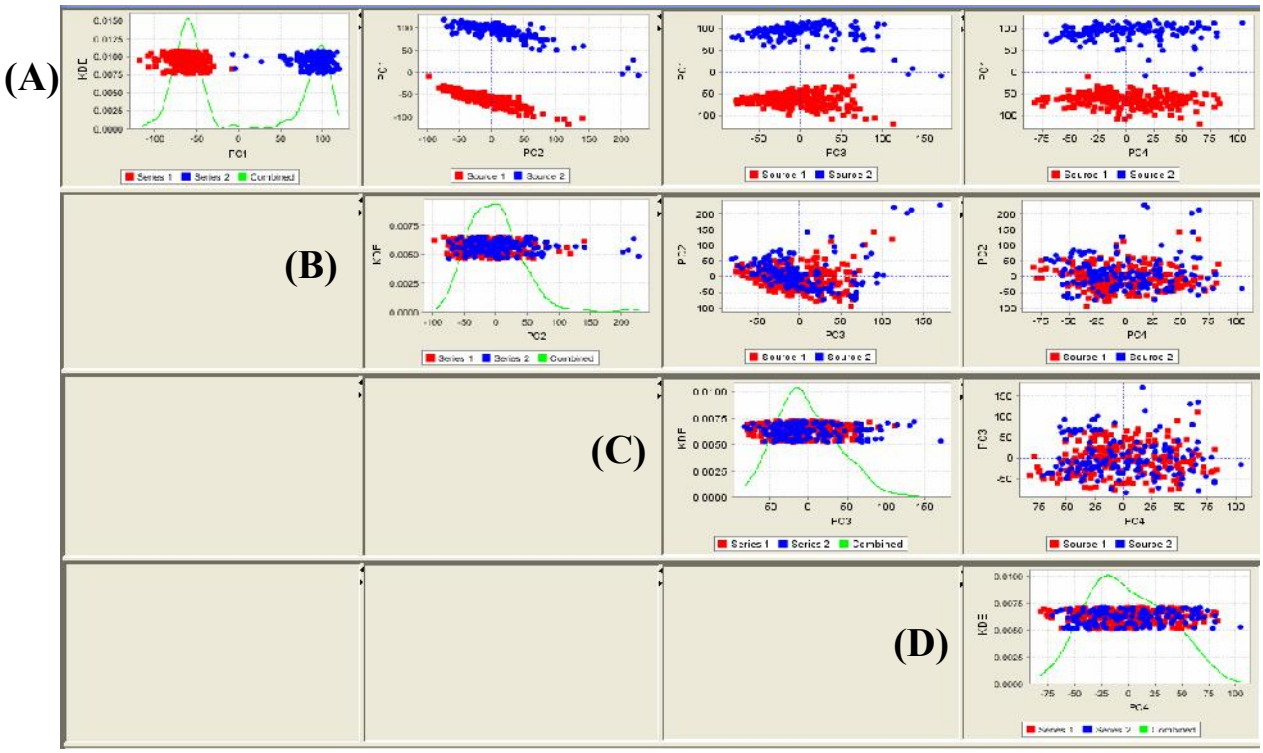


Figure 3.11 : Multivariate projection view of the merged GC-RMA data showing effective removal of platform bias using the DWD batch removal tool. The green line connects all samples. Red dots (Series 1 and Source 1) refer to GC-RMA normalised samples from HG-U133A and blue dots (Series 2 and Source 2) refer to HG-U133 Plus2.0. Kernel density estimations (KDE) show distribution of the values for the combined array expression. Multimodalities are indicated in principal component (PC) 1 for the HG-U133 Plus2.0 and HG-U133A and in PC2 - PC4 for the combined arrays. **(A – C)** Scatter plots with KDE are showing the distribution of the array expression means projected on the DWD direction. **(D)** KDE of the pooled DWD distribution of HG-U133 Plus2.0 and HG-U133A is shown in the scatter plot.

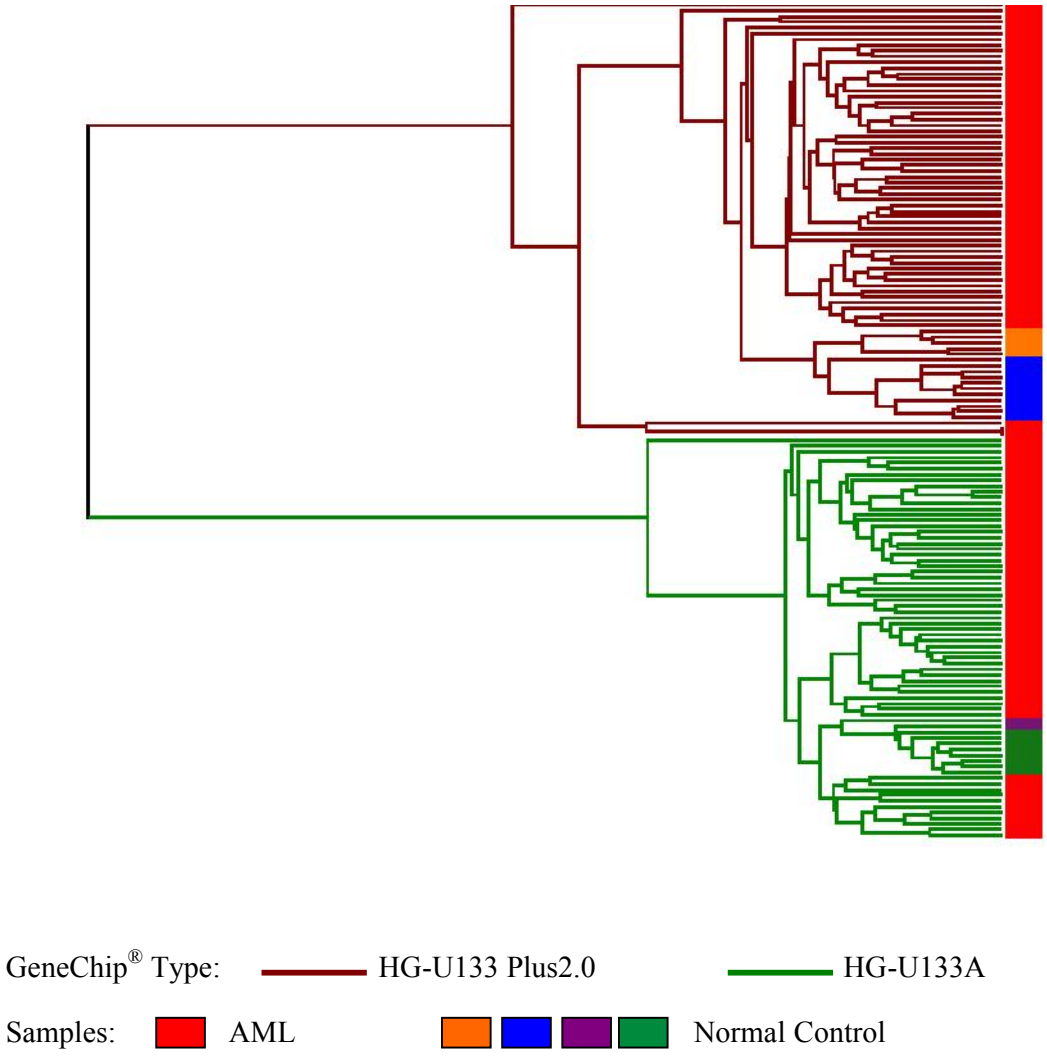


Figure 3.12 : Hierarchical clustering of the combined samples from HG-U133A (green dendrogram) and HG-U133 Plus2.0 (brown dendrogram) before batch effect adjustment.

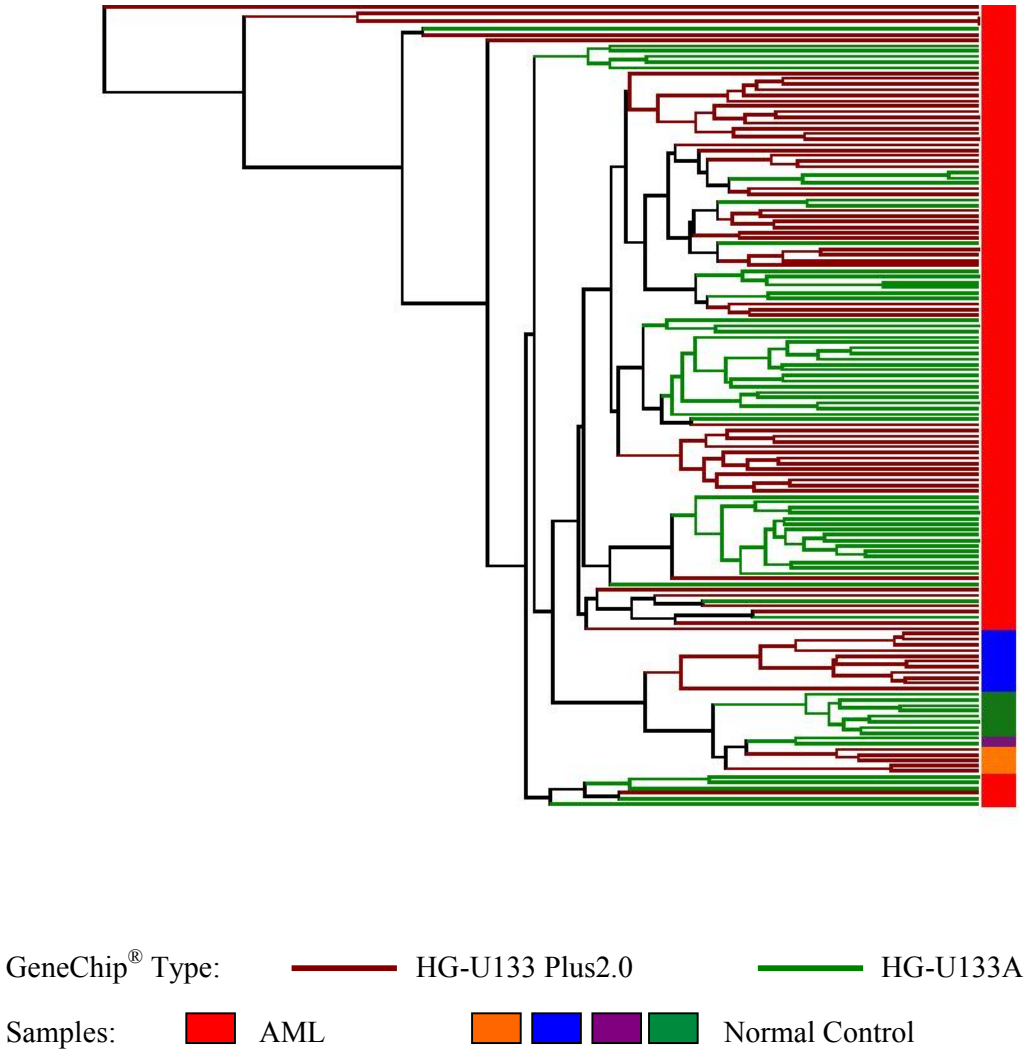


Figure 3.13 : Hierarchical clustering of the combined samples from HG-U133A (green dendrogram) and HG-U133 Plus2.0 (brown dendrogram) after batch effect adjustment.

3.4.7.2 High *TCF7L2* gene expression is associated with reduced complete remission rate in patients with adverse cytogenetic AML

To gain further insights into the prognostic impact of *TCF7L2* expression at diagnosis in AML, a cohort of AML patients (n=246) for which both microarray and clinical data were available were analysed. Among those, 55 patients didn't achieve CR. The prognostic impact of *TCF7L2* expression in AML was assessed using multivariate analysis described in 3.3.7. This analysis took into account patient co-variables including age, WBC count, cytogenetic prognostic group, gender and WHO performance. The analysis revealed that *TCF7L2* mRNA overexpression was found to be independently prognostic for reduced CR rate after the adjustment for co-variables with an odds ratio (OR) for CR: 5.19 (1.39-19.39) $P = 0.0144$. No statistically significant impacts were seen between *TCF7L2* gene overexpression and OS or relapse incidence. In summary, *TCF7L2* mRNA expression is an independent prognostic marker for CR status in AML patients.

3.5 DISCUSSION

3.5.1 Aspects of GEP processing and data integration

Analysis of microarray GEP data has increasingly become a common technique to provide insights into biological processes. Global inter-laboratory studies have recognised this tool for various clinical applications such as to predict patients' outcome, to discover cancer subclasses, and to identify new therapeutic targets (Bullinger & Valk, 2005; Haferlach *et al*, 2005). Although in the analysis of microarray data, many statistical methods exist for identification of relevant target genes, the choice of method used can greatly affect the set of genes identified (Allison *et al*, 2006; Jeffery *et al*, 2006; Dupuy & Simon, 2007; Ioannidis *et al*, 2009).

To identify important genes and pathways that were dysregulated in AML, the initial approach taken relied on identifying common underlying gene changes associated specifically with developmental subtypes of AML (M0 and M1) compared to Normal Control; the changes observed were then extended to all subtypes (M0 - M6). Unfortunately, $CD34^+$ Normal Control GEP data was not derived as part of the original experimental design at Cardiff. To circumvent this, this study downloaded pre-existing GEP normal $CD34^+$ data. Although these data were derived from a number of separate experiments, their expression profiles clustered tightly together suggesting their intrinsic gene expression was similar to each other but markedly different from the AML GEP (as shown by multidimensional visualisation Figure 3.6). In addition to the limited availability of Normal Control GEP data, there is some concern about what is the ideal control to compare an AML sample to. Although limited, several studies have compared AML patients' GEPs with normal cells derived from the $CD34^+$ cell population (Stirewalt *et al*, 2008; Majeti *et al*, 2009a). Heterogeneity of the $CD34^+$ cell subpopulation in leukaemia has been reported (Bonnet & Dick, 1997; Gentles *et al*, 2010; Rockova *et al*, 2011), as well as the heterogeneity of $CD34^+$ cells in normal samples. In contrast to AML, one study found there was no significant difference in gene expression signatures between $CD34^+$ cells derived from PB and $CD34^+$ data derived from BM in AML myeloblasts (Cheung *et al*, 2009).

Combining multiple microarray GEP data sets across different experiments as part of data pre-processing has always been challenging. Selecting a subset of discriminative genes related to AML pathogenesis requires robust statistical methods and normally multiple approaches would be used to derive a final genelist. This includes different normalisation and summarisation procedures. Because of this complexity in studying genes distributed in different experiments processed through sequential hybridisations, background signal corrections were performed based on the widely accepted algorithm GC-RMA that takes into account PM signals. The PM-only methods such as RMA, GC-RMA and Probe Logarithmic Intensity ERror (PLIER) prove particularly powerful in reducing false positives (Seo & Hoffman, 2006) and have previously been shown to perform better in detection of differential expression (Mieczkowski *et al*, 2010) as compared to the MAS5.0 algorithm. Even though RMA based normalisation is more precise, it may hide real changes (due to lower sensitivity), especially at low expression levels (Irizarry *et al*, 2003a; Millenaar *et al*, 2006).

To account for potential artefactual differences that may arise as a consequence of combining multiple GEP data e.g. AML trial GEP data hybridised on different Affymetrix GeneChips® (U133A and Plus2.0) a batch correction method was used. Systematic biases due to the two Affymetrix platforms were removed and GEPs were merged according to the DWD algorithm (Benito *et al*, 2004; Qiao *et al*, 2010). This approach has previously been reported to be useful for validating cancer intrinsic subtype gene sets that are predictive of survival outcome in breast cancer (Hu *et al*, 2006). DWD is flexible because it allows input data to be pre-processed using any form of normalisation of interest (as RMA used above). However with DWD, only two batches (i.e two sets of GEPs) can be applied at a time. Other methods such as Empirical Bayes (Johnson *et al*, 2007) and Singular-Value Decomposition (Alter *et al*, 2000) are among other options available that could remove biases caused by non-biological conditions when sample numbers per batch are much smaller (fewer than 25 arrays). Because the majority of microarray studies are conducted using much smaller sample sizes (eg. fewer than 30 samples), existing batch correction methods are not sufficient to remove the systematic biases in data, but this is not the case for this study. As a result of merging two cohorts of AML trial patients that give rise to a total of more than 200 samples, the adjusted GEP was then successfully utilised to assess *TCF7L2* expression as a prognostic marker in this study.

This study applied integration of GEP data of Normal Controls from external sources to be normalised with the existing AML GEPs. Through other meta analysis, researchers have shown reliability of output obtained from integrating microarray repository data (Choi *et al*, 2003; Rhodes *et al*, 2004; Warnat *et al*, 2005). This also promotes the discovery of small but consistent expression changes with increased sensitivity. To facilitate this process, a series of important checklists for conducting such a type of microarray analysis have also been extensively reviewed by Ramasamy *et al*; one may remove arrays that are of poor quality and check for batch effects before combining large data sets from multiple sources (Ramasamy *et al*, 2008).

3.5.2 Networks of aberrantly expressed genes in AML

Due to the growing computational discovery programmes available (eg: Cytoscape, GeneGo MetaCore™, Ingenuity Systems, Ariadne Pathway Studio etc) a much more coordinated and data rich framework exists in which regulatory networks and pathway analyses are among the advanced approaches applied to analysis of GEP data. For example, these topological tools have discovered important unsuspected cross talk between different haematopoietic pathways and have become the foundation for reconstructing a much more up-to-date stem cell transcriptional network due to increased number of known TF binding sites (Berger *et al*, 2006; Miranda-Saavedra & Gottgens, 2008; Philippakis *et al*, 2008). In order to understand more of the biological interactions behind aberrant gene expression in AML, this current study takes advantage of the Genego Metacore™ Interactome and pathway discovery tools. Using such approaches, this study has identified dysregulated expression of a number genes in AML that belong to the "Cytoskeleton remodeling, WNT, TGF" pathway. This pathway name is given by GeneGo MetaCore™ and consists mostly of WNT components that cross-talk to a small subset of genes involved in TGF signalling (eg: *SMADs, RHOA, RAC*). Cytoskeletal remodelling is a simplified term used to describe the role of WNT in cancer progression mechanisms such as microtubule stabilization, cell polarity (via non canonical WNT), motility and adhesion as reviewed by (Lai *et al*, 2009).

WNT signalling is well known to regulate many cellular processes including cell development and cell to cell interactions and is one of the most significantly affected pathways in CLL (Lu *et al*, 2004), AML (Wang *et al*, 2010) and is well described in solid tumours (Roman-Gomez *et al*, 2007; Liu *et al*, 2008). However, the role of many key proteins within WNT signalling is still not well described in the pathogenesis of AML in

comparison to solid tumours. Constitutive activation of canonical WNT / β-catenin expression has previously been shown in AML (Simon *et al*, 2005) including AML with normal karyotype or low risk groups (Mikesch *et al*, 2007). Deregulation of non-canonical WNT signalling (i.e independent of β-catenin function) has also been observed in haematologic malignancies. Hypermethylation of genes in the non-canonical WNT pathway (eg. *WNT5a*) has been associated with reduced survival and high relapse in AML patients (Roman-Gomez *et al*, 2007; Valencia *et al*, 2009; Martin *et al*, 2010), showing involvement of non-canonical WNT in AML progression.

To date, there are at least 200 upstream proteins known to modulate WNT activity including cell signalling ligands, receptors, phosphatases and transcription factors. In addition, around 150 or more WNT target genes are transcribed as a result of direct binding of TCF protein to relevant promoters and the list of WNT target genes continue to expand throughout the years. This current study represents an effort to simultaneously compile the growing number of WNT signalling target genes and other known WNT associated genes from a wide variety of resources using both discovery tools (GeneGo MetaCore^{TM}, DAVID, etc) and published literature. The resulting list of WNT genes and their corresponding probe sets has become a primary reference in the process of searching for the most dysregulated gene in WNT signalling. The analysis identified several common genes found to be overexpressed in AML, for example *TCF7L2, CD44 , RAC1, RHOA, PPP2R5C*; expression of these genes are highest in subtype M0 (Table 3.4).

TCF7L2 is of particular primary interest because it has been reported to be the central regulator of canonical WNT signalling that interacts with β-catenin acting in the nucleus and mediates transactivation of other genes. β-Catenin was also found to be significantly overexpressed by one genelist but not observed by Interactome analysis. Although not associated elsewhere with prognosis or any particular FAB subtype, *TCF7L2* has been described in AML, in which FLT3-ITD AML has shown to induce TCF transcriptional activity (Tickenbrock *et al*, 2005). Upregulation of *TCF7L2* expression was also documented during macrophage differentiation at the expense of granulocyte differentiation (Baek *et al*, 2009; Brown *et al*, 2012) showing an important role in normal haematopoiesis. Through DNA array-based genome-wide analysis of TCF7L2 chromatin occupancy, others have identified with high confidence TCF7L2-binding sites in target genes showing inappropriate activation of *TCF7L2* in colorectal cancer progression (Hatzis *et al*, 2008).

This study further suggests involvement of other genes that are not part of transcription factors in the dysregulation of AML. The upregulation of *RAC1* and *RHOA* seen in AML patients partly supports by the role of Rac GTPase signaling in myeloid-associated disease that is well documented to be up-regulated in murine AML (Rozenveld-Geugien *et al*, 2007). *RAC1* and *RHOA* are shown to enhance β-catenin accumulation in the nucleus leading to activation of canonical WNT through TCF and activation of *RAC1* has been reported to control nuclear localisation of β-catenin (Wu *et al*, 2008). Both *RAC1* and *RHOA* also play role in non-canonical WNT and this suggests involvement of non-canonical WNT dysregulation in AML. Identification of *CD44* overexpression in this study is in keeping with other studies showing overexpression in both AML and CML stem cells relative to normal HSC (Jin *et al*, 2006). Conversely, this current study also observed overexpression of negative regulators of WNT signalling such as *PPP2R5C*. The overexpression of *PPP2R5C* was reported to reduce the abundance of β-catenin and to inhibit the transcription of TCF / β-catenin target genes by interacting with Axin / APC (Seeling *et al*, 1999; Hsu *et al*, 1999). Among the most commonly downregulated genes seen in this study is *PLCB4*. *PLCB4* had been reported to be one of the gene related to a more primitive HSC stem cell and the expression is high in normal stem cells such as the multipotent mesenchymal stem cells (Tsai *et al*, 2007) and in the subpopulation of $CD133^{+}$ derived CB cells that play important role in haematopoiesis (Toren *et al*, 2005). Little is known about implication of *PLCB4* in the context of AML but *PLCB1* another key enzyme in nuclear signal transduction has been reported to play a role in the progression of MDS to AML (Cocco *et al*, 2005). However, inconsistencies were also observed where Majeti *et al* found downregulation of *RAC1* and did not find any significant changes of *TCF7L2, PLCB4, RHOA* or *PPP2R5C* when comparing normal to leukaemic stem cells (Majeti *et al*, 2009).

Subsequent analysis using Interactome as a gene interaction model revealed additional important genes in the dataset, which were missed by the direct fold change analysis. Moreover, this provides evidence that *TCF7L2* and other genes identified earlier are also inter-connected to each other. Following Interactome, the same set of genes *TCF7L2, CD44, RAC1, RHOA* and the *PPP2R* family once again appeared to be significant. These findings proposed that these genes are overconnected in the AML dataset along with several other genes from different protein classes. *JUN, LEF1* and the *SMAD* family are among important TF genes that were highly overconnected and also found to be

dysregulated (Table 3.7). The NFAT family (*NFATC1, NFATC3, NFATC4*) also appear to be among the overconnected TFs. NFAT is a key TF in non canonical WNT. Researchers have shown that NFAT protected CML cells and was responsible for Imatinib resistance in CML implying that NFAT plays important roles in leukaemia cell survival (Gregory *et al*, 2010). Transcription activation of NFAT is specifically enhanced by promyelocytic leukaemia (PML) protein, showing specific binding of NFAT to PML. PML also has a wide spectrum of target genes (Lo *et al*, 2008).

Besides *NFAT* family genes and *TCF7L2*, another TF that also has a high z-score (second most connected TF) is *JUN*. The upregulation of *JUN* has been documented in leukaemia patients previously (Casas *et al*, 2003; Rangatia *et al*, 2003; Pise-Masison *et al*, 2009). One particular interest is the complex between *TCF, JUN* and *DVL1*. *DVL1* is another WNT regulator that was found be overconnected as well as upregulated in this study. It is well known that the formation of the TCF / β-catenin transcriptional complex is subjected to many forms of regulation of other genes. Both *DVL1* and *JUN* stabilise the nuclear transcription complex whilst reduction of *DVL1* shown to suppress β-catenin / *TCF7L2* association (Gan *et al*, 2008). The upregulation of both *DVL1* and *JUN* could therefore lead to aberrant activation of WNT that disrupt the balance of this mechanism which is crucial for developmental processes. In the context of non-canonical WNT signalling such as planar cell polarity, *DVL1* interacts with the *DAAM1 / RHOA* axis, the *RAC1* axis and *GSK-3β* microtubule signalling (Habas *et al*, 2001) to regulate cytoskeleton remodelling through a transcriptional-independent mechanism, indicating the ability of *DVL1* to mediate complex and versatile signal transduction.

Of the eleven most overconnected TF genes identified, a number of important lineage-specific TF partners were also seen to be dysregulated. Among these are the *TCF7L2, SMAD* and *NFAT* families. These three groups of genes have been reported to be important for long term lineage decisions for T-regulatory cells and shared conserved binding sites for *FOXP3* locus controlling regulatory T cell development (Floess *et al*, 2007). Because TF genes are well known to play a significant role in cell lineage fate, it is important to understand their function and the consequences of TF dysregulation in this process. Normal haematopoiesis relies on lineage-specific transcription factors including *GATA2, RUNX1, TAL1, PU.1* that control cell differentiation; disruption of these factors in normal differentiation is a hallmark of AML blasts (Tenen, 2003; Pabst & Mueller, 2007). Complementary to that, *JUN* (Oh *et al*, 2000) and *TCF7L2* (Jaatinen *et al*, 2006) are among

the TF genes that are highly expressed in HSC. Another observation is aberrant expression and high connectivity of both *TCF7L2* and *SMAD* (*SMAD2, SMAD4*) indicating an altered activity of these ultimate effectors of BMP / WNT pathways. Both TCF7L2 / SMAD proteins interact during developmental processes and co-occupy genomic sites with GATA1 / GATA2 to regulate erythroid lineage (Trompouki *et al*, 2011). Disruption of either by silencing or aberrant expression of TCF / SMAD complex might affect their established roles in erythropoiesis.

Other genes such as *CAMK2, NLK* and *NFAT* that antagonise β-catenin-dependent canonical WNT signalling were also found to be dysregulated in this AML GEP study suggesting involvement of non-canonical pathway through either the CAMK2-TAK1-NLK pathway or NFAT-mediated transcriptional regulation. Both NLK and NFAT were previously reported to suppress β-catenin-dependent transcription (Ishitani *et al*, 1999; Saneyoshi *et al*, 2002). The WNT dysregulated genes presented so far indicate the possibility of involvement of non canonical WNT, in addition to the dysregulation in TCF / β-catenin activation in AML.

3.5.3 Aspects of gene selection

Affymetrix designed a certain number of probe sets to detect expression of each gene (ranging from 1 until 13 probe sets per gene). With regards to the number of probe sets found to be overexpressed or underexpressed in this study, there are several inconsistencies seen between different probe sets per gene. From Table 3.7, *CD44, GSK-3B* and *TBL1XRL* are examples of genes that appeared to have more than one probe set concomitantly significantly expressed, suggesting consistent signal detection of more than one probe set that exists on the Affymetrix GeneChip® belonging to these genes. However this is not the case for majority of other genes. Considering *TCF7L2*, there are actually seven probe sets engineered on the GeneChip® for hybridisation to occur. Nevertheless, only the probe set located within the centre of the coding gene was found to be consistently overexpressed (212759_s_at). For the other six probe sets that are located near to the 3' end of *TCF7L2* mRNA, the detection was hampered after statistical tests and fold change filtrations possibly due to weaker signals than the interior probe set. For *CD44*, there are 13 probe sets available to detect its expression on a single GeneChip® but only half are efficiently detected after statistical filtrations. These types of discrepancies could be explained partly by the location of probe sets that are close to the 3' ends of genes tending

to have a lower chance of coverage and to not be amplified as efficiently as the interior probesets, whilst probe sets located closer to the 5' end will have higher GC content with less specific binding to the genes (Bemmo *et al*, 2008). Therefore the interior probe sets could be the most reliable target for hybridisation. By taking *CD44* and *TCF7L2* as examples in this study, there is also a wealth of evidence showing multiple splicing within exonic regions of these two genes (Hsu *et al*, 1999; Shiina *et al*, 2003; Wang *et al*, 2009; Weise *et al*, 2010; da Cunha *et al*, 2010; Banky *et al*, 2012; Zeilstra *et al*, 2013). This explained the inconsistencies of the 3' IVT Affymetrix probe sets' ability to detect all signals with equal efficiency, particularly for highly spliced genes.

In summary, it is clear that AML is constrained by complex networks involving both positive and negative regulatory interactions among substantial numbers of WNT genes and that no single gene or TF is necessary and sufficient to cause AML progression. This study not only identified *TCF7L2* to be consistently upregulated across all datasets but is also the first study to show its prognostic value in AML. Constitutive activation of TCF / β-catenin remains the most widely accepted mechanism for dysregulation of WNT signalling. However, *TCF7L2* expression has more complexity involving its gene structure, variants and functions due to multiple splicing which require further investigation in AML. The functional specificities of the different *TCF7L2* isoforms remain incompletely understood in AML. The next chapter will attempt to distinguish the different variants of *TCF7L2* mRNA that might exist in AML (that cannot be distinguished by microarray) coupled with protein-mRNA correlational studies adding further insights into validating *TCF7L2* expression in AML.

4 - Validation of TCF7L2 as a dysregulated WNT signalling transcription factor in AML

4.1 INTRODUCTION

Data from gene expression analysis in the previous chapter established WNT signalling to be the most significantly dysregulated pathway in AML. Extended downstream analysis of differentially expressed WNT signalling genes using Interactome (MetaCore^{TM}) identified *TCF7L2* as the most overconnected TF to be aberrantly expressed in AML. As previously described (see 1.4.1), TCF7L2 is a member of the TCF family in the canonical WNT signalling pathway and affects a wide range of fundamental cellular processes such as embryonic development (Korinek *et al*, 1998b), cell lineage regulation (Trompouki *et al*, 2011), stem cell maintenance (Pinto & Clevers, 2005), tumour suppression (Angus-Hill *et al*, 2011) and oncogenesis (van de Wetering *et al*, 2002; van & Clevers, 2002). Misregulation of TCF / β-catenin activities and involvement of TCF7L2 in cancer formation have been reported in both myeloid and T-cell leukaemia (Chang *et al*, 2006; Pise-Masison *et al*, 2009) as well as solid tumours (van de Wetering *et al*, 2002; Shiina *et al*, 2003; Valenta *et al*, 2003).

Besides being dysregulated at transcript expression level as described in Chapter 3, previous studies in vertebrates and solid tumours have shown the presence of alternative splicing within the coding region of TCF7L2 particularly at the exterior of the β-catenin and HMG box domains (Arce *et al*, 2006; Weise *et al*, 2010). According to the NCBI database, *TCF7L2* is comprised of 18 exons, six of which are alternative exons (i.e. exon 4, 6 and 14 - 17), with additional complexities to exon 8 and exon 10 that have been reported to be alternatively spliced into two distinct sequences; i.e. type-a and type-b (Pukrop *et al*, 2001; Weise *et al*, 2010). To date, there are at least 13 types of TCF7L2 protein isoforms published in public databases (NCBI and UniProt), and the longest isoform of the full length TCF7L2 protein has been characterised to encode several functional domains that are important for its transcriptional regulation including the β-catenin binding domain, NLS and promoter-specific activation region (i.e. C-clamp) as illustrated in Figure 1.10. Through mRNA splicing, TCF7L2 protein isoforms with differing properties are generated. Full

length TCF7L2 cDNA with E-tail isoforms (hereafter known as C-tail E1 - E4) has originally been reported to express unique cysteine rich sequence-specific DNA binding motif known as C-clamp motifs (CRARF or CRALF) that recognise different groups of TCF binding elements which are located in the promoter regions of TCF target genes (Hecht & Stemmler, 2003; Wallmen *et al*, 2012). Several promoter genes such as Caudal Type Homeobox Transcription Factor-1 *(CDX1)* and *LEF1* specifically require a C-tail E2 TCF7L2 isoform for β-catenin activation (Atcha *et al*, 2003). As a nuclear mediator of WNT signalling, TCF7L2 also associates with several other corepressors such as Groucho, the MTG family and TLE to repress transcription in the absence of β-catenin (Moore *et al*, 2008). In addition to the C-clamp, TCF7L2 with the C-tail E2 isoform also encode another unique motif within exon 18 that has been reported to recognise a transcriptional corepressor known as CtBP that could prevent β-catenin from binding to TCF (Eastman & Grosschedl, 1999; Fang *et al*, 2006).

However, further details of TCF7L2 alternative splicing have not been characterised in AML. This chapter will therefore be focusing on detecting variations of *TCF7L2* mRNA transcript and protein expression, as well as in depth characterisation of each TCF7L2 protein isoform in both AML and normal cells. Findings from this chapter will allow functional studies to be performed on the predominantly overexpressed isoform found in myeloid cells.

4.2 AIMS

Affymetrix microarray data analysis has identified overexpression of *TFC7L2* in AML blasts compared to normal HPC. It is known from previous studies that this gene generates multiple protein isoforms, each of which might have a different functional significance. Therefore this chapter aims to:

- Characterise the *TCF7L2* mRNA variants in AML patients and leukaemic cell lines and to compare the expression to normal HPC and total BM cells. This will determine the exon composition of the different mRNA variants seen and their predicted protein molecular weight.
- Determine which *TCF7L2* isoforms are overexpressed at the protein level in AML. Characterise the TFC7L2 variants (again in comparison with normal HPC), their subcellular localisation and whether expression of mRNA correlates with protein TCF7L2 levels in AML.
- Compare how TCF7L2 isoform expression in AML differs from that seen in normal haematopoietic development.

4.3 MATERIALS AND METHODS

4.3.1 Source of primary AML blasts, myeloid cell lines, normal BM and normal $CD34^+$ HPC from CB.

Primary AML blasts (mononuclear fractions) were collected from the MRC / NCRI AML Trial bank. All samples had ethical permission and material transfer agreement (MTA) with Cardiff University. RNA was extracted using the protocol described in 2.5.1. The RNA samples were used for variant PCR analysis as detailed in Table 4.1. Thirty RNA samples from this cohort were already available and previously processed for the MILE study (Kohlmann *et al*, 2008a). Additional AML samples were obtained prospectively from MRC trials AML16 (n=5) and AML17 (n=5) that were not part of the MILE study but were processed similarly. From this cohort, a total of 19 AML samples (refer to Table 4.1) had

their proteins extracted using the method previously described in section 2.8. These were used for western blotting.

Normal BM cells (n=4) were derived from healthy donors purchased from AllCells® (Alameda, U.S.A.). Normal human HPC were derived from neonatal umbilical CB that was collected from the Maternity Unit, UHW, Cardiff following informed consent. Mononuclear cells were originally isolated from CB and the $CD34^+$ population was enriched according to a similar method to 2.4.2 – 2.4.3. RNA and protein samples from normal $CD34^+$ HPC (n=4) were provided courtesy of Dr Alex Tonks and Dr Richard Darley (Tonks *et al*, 2007b).

Myeloid cell lines were cultured as described in 2.3.1 and passaged at least twice before use in the experiments described below. A total of five myeloid cell lines were tested for alternative splicing using PCR namely K562, U937, THP-1, NB-4 and Kasumi-1. RNA from normal BM and myeloid cell lines was extracted using the method described in 2.5.1. Nuclear and cytosolic fractions were extracted as described in section 2.8.

Characteristic at diagnosis	Patients for variant analysis	Patients for protein analysis
Trial group		
AML14	9	4
AML15	21	5
AML16	5	5
AML17	5	5
Diagnostic sample type		
Bone marrow	4	-
Peripheral blood	13	11
Not determined / unknown	23	8
FAB subtype		
M0	5	2
M1	11	7
M2	7	1
M3	0	0
M4	8	2
M5	2	1
M6	0	0
RAEB	1	1
Not determined	6	5
Karyotype group		
t(15;17)	0	0
t(8;21)	2	0
inv(16)	4	0
Complex (>5 chromosomal abnormalities)	4	1
Normal karyotype	7	2
Other chromosomal abnormalities	6	2
Not determined / unknown	17	14
Total	40	19

Table 4.1: Details of AML patients included in the variant and protein analysis of TCF7L2.

4.3.2 Determination of TCF7L2 mRNA splice variants by PCR

The total RNA yield was determined for each sample using a Nanodrop spectrophotometer with a purity between 1.7 to 2.0 as determined by A_{260}/A_{280}. For the cDNA synthesis, 500 ng of total RNA was reversed transcribed (see 2.5.3). Two sets of primers were used to amplify *TCF7L2* mRNA. The primer sequences and their respective Tm are shown in Table 4.2. The PCR reactions were performed using 5 µl of the RT product, 0.5 µM each of the forward and reverse primer (Eurofins MWG Operon, Ebersberg, Germany), 200 µM of dNTP mix (Applied Biosystems®, Paisley, UK) , 1X PCR buffer II (Applied Biosystems®, Paisley, UK), 1 mM of $MgCl_2$ (Applied Biosystems®, Paisley, UK) for N-terminus primer pair (or 0.9 mM for C-terminus primer pair), 0.5 units of AmpliTaq® Gold DNA Polymerase (Applied Biosystems®, Paisley, UK) and made up to a final volume of 25 µl with sterile molecular grade water. The PCR reaction was carried out in an Applied Biosystem thermal cycler. The sample reaction mix was then subjected to Touch-down (TD) PCR cycling parameters as detailed in Table 4.3. PCR products or amplicons from the N-terminus primer pair amplification were separated on a 2% (w/v) agarose gel in 1X TBE buffer (Tris (89 mmol/L), boric acid (89 mmol/L) and EDTA (2 mmol/L), pH 8.3) at 120 V for 50 minutes. The PCR products amplified by the C-terminus primer pair were separated on 3% (w/v) agarose gel for better resolution due to the shorter amplicons generated. DNA ladders (New England Biolabs, Hertfordshire, UK) of either 50 bp or 100 bp sizes were included in every electrophoretic run. The agarose gel was visualised using ethidium bromide staining and ultraviolet (UV) light and digitised using a Fujifilm LAS3000 imaging system.

Type of primer pair	Exons coverage	Calculated size of full PCR product without exon skipping (bp)	Primer sequences (5' – 3')	Tm °C
N-terminus (5'-Primary)	exon 1 to exon 12	1213 bp product from linear template gi|226371763|ref| NM_001146274.1	Forward ex1: AAT TGC TGC TGG TGG GTG A	56.7
			Reverse ex12: TGC TCT TCT CTG GAC AGT GC	59.4
C-terminus (3'-Tail)	exon 13 to exon 18	151 bp product from linear template gi|226371763|ref| NM_001146274.1	Forward ex13: CAA GCA GCC GGG AGA GAC	60.5
			Reverse ex18: TGG GCT GAG GCA GCT GCC TT	63.5

Table 4.2: Primer combinations used for variant screening and sequencing. Calculated PCR size shown is based on *TCF7L2* gene exon sequences published for NM_001146274.1. Primer sequences were adapted from (Shiina *et al*, 2003).

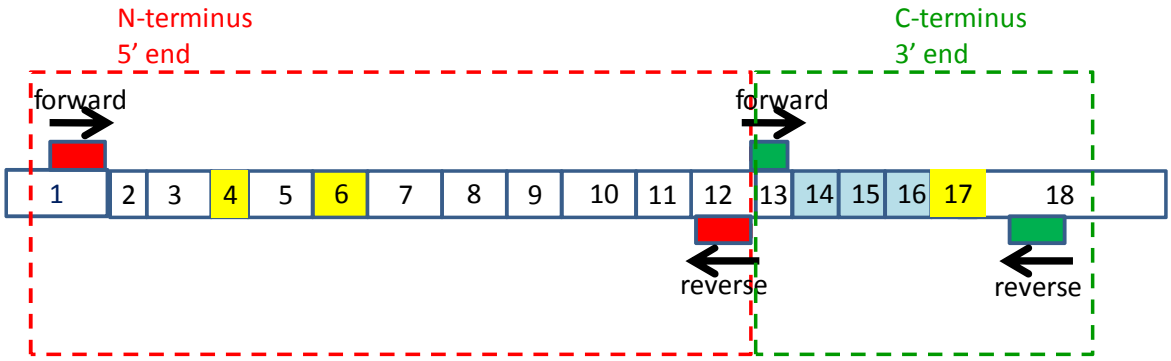


Figure 4.1: Schematic diagram illustrating the positions of both primer sets within the *TCF7L2* gene. The length shown for each exon in the diagram is approximate.

Cycle parameters	Temperature (°C)	Duration	Cycle
Pre-incubation:	94	10 min	(1 step x 1 cycle)
Touch-down PCR:- (***Part 1***)			(6 steps x 5 cycles)
Melt:	94	30 sec	
Anneal:	70 (reduce 2ºC every 2 cycles)	30 sec	
Extension:	72	1 min	
(***Part 2***: **additional cycles to stabilise the reaction in *Part 1***)			
Melt:	94	30 sec	
Anneal:	70 (reduce 2ºC every 2 cycles)	30 sec	
Extension:	72	1 min	
Melt :	94	30sec	(3 steps x 35 cycles)
Anneal :	60	30sec	
Extension :	72	1min	
Final extension:	72	5min	(1 step x 1cycle)
Hold:	4	-	-

Table 4.3: TD PCR parameters used for PCR screening. TD protocol was characterised by Part 1 and Part 2 that were added to the standard PCR cycling parameters. The annealing temperature of the TD reaction was decreased by 2°C every 2 cycles from 70°C to a final touchdown at 65°C after 5 cycles. This helps to reduce the formation of heterodimers and to increase the binding affinities of primers to the template DNA, particularly for long sequences. The TD cycles were followed by standard PCR cycling of 35 cycles.

4.3.3 Sequencing of PCR products

4.3.3.1 PCR products above 100 bp

Selected samples were sequenced using either the N-terminus primer pair (exon 1 and exon 12) or the C-terminus primer pair (exon 13 and exon 18) depending on the PCR product. In order to sequence a single band from a mixture of PCR products (indicated by multiple bands on the gel), each band was excised (see 2.6.2). After gel purification, the TD PCR step was repeated (see 4.3.2) using the gel purified DNA, to ensure specificity of each excised band. The resulting single PCR product was then purified again using the QIAquick PCR Purification Columns protocol (Qiagen®, West Sussex, UK) and sequenced. PCR products that contained a single band on a gel were directly purified using the QIAquick PCR Purification Columns protocol (Qiagen®, West Sussex, UK) described in 2.5.5 and sequenced according to steps described in 2.5.6.

4.3.3.2 PCR products below 100 bp

PCR products that contain a DNA amplicon of less than 100 bp were excised from the gel according to methods described in 2.6.2 and cloned into the appropriate plasmid vector (see 4.3.4). Plasmid DNA from at least two clones was selected for sequencing using universal M13 forward and reverse primers (see 2.5.6). The pDrive cloning vector map showing the position of the M13 primer pair site is available in Supplementary 4.1 material.

4.3.4 PCR cloning

PCR cloning allows small products (< 100 bp) to be sequenced more efficiently. PCR products of less than 100 bp were excised from the gel and cloned into pDrive vector using the Qiagen PCR Cloning Kit (Qiagen®, West Sussex, UK) in accordance with the manufacturer's instructions. In brief, after PCR reaction, a single band was purified from 3% (w/v) agarose gel according to 2.6.2. For ligation, 13 ng of gel purified PCR products were mixed with 50 ng pDrive plasmid vector, 5 µl ligation master mix (2X) and sterile water to a final volume of 10 µl. The ligation reaction mixture was gently mixed and then incubated at 4°C for 30 minutes before proceeding to transformation and plasmid purification (see 4.3.4.1 - 4.3.4.2).

4.3.4.1 Transformation using EZ Competent Cells

Following ligation incubation, 2 µl of the ligation reaction mixture was added into one vial of Qiagen® EZ Competent Cells, mixed gently and incubated on ice for 5 minutes. The cells were then heat shocked for 30 seconds in a 42°C water bath without shaking and returned on ice for 2 minutes. Following that, 250 µl of SOC medium was immediately added to the transformation mixtures and directly plated. Between 50 µl and 100 µl of transformation mixture was spread onto pre-warmed LB agar plates containing 100 µg/ml ampicillin, 80 µg/ml X-gal and 50 µM IPTG. The plates were incubated in a 37°C incubator for 16 hours. Plasmid incorporating the PCR products in the correct orientation produced white recombinant colonies instead of blue colonies.

4.3.4.2 Small scale plasmid purification

Colonies of interest were picked from LB agar plates and inoculated into 5 ml of LB broth containing 100 µg/ml ampicillin and incubated at 37°C overnight with a shaking speed of 220 rpm. After overnight incubation, the bacterial pellet was harvested from 3 ml culture by centrifugation at 10,000 x g for 10 minutes at RT. Plasmids were purified according to the QIAprep Spin Miniprep Kit (Qiagen®) protocol. In brief, the pellet was resuspended in 250 µl of chilled Buffer P1 and transferred to a fresh microcentrifuge tube. After cell clumps and pellets were completely resuspended by vortexing, 250 µl of Buffer P2 was immediately added and the tube was gently inverted for mixing. Subsequently, 350 µl of Buffer N3 was added, gently mixed and the tube was centrifuged for 10 minutes at 10,000 x g. A clear supernatant containing plasmid DNA was transferred on to the QIAprep spin column by pipetting and centrifuged at 10,000 x g for 1 minute. The QIAprep spin column was washed by adding 0.5 ml Buffer PB followed by 1 minute of centrifugation. Plasmid DNA was precipitated by the addition of 0.75 ml Buffer PE and centrifugation at 10,000 x g for 1 minute. The tube was spun twice to remove residual ethanol from Buffer PE. Plasmid DNA was eluted from the column by adding 50 µl Buffer EB (10 mM Tris-HCl, pH 8.5) to the centre of each QIAprep spin column followed by a final 1 minute centrifugation.

4.3.5 Analysis of nucleotide and protein sequences

The formal scientific name for *TCF7L2* transcript and protein was used according to HGNC guidelines and the National Centre for Biotechnology Information (NCBI) nucleotide database at http://www.ncbi.nlm.nih.gov/nuccore. The length of each exon, exon numbering, and exon sequences were obtained from NCBI and summarised in Supplementary 4.2. Base calling quality checks of DNA sequences obtained from sequencing were processed and assembled using the BioEdit Sequence Alignment Editor version 7.0.9.0. DNA sequence multiple alignment, similarity search, translation and identification of open reading frames (ORFs) were conducted using both pDRAW32 1.0 revision 1.1.114 by AcaClone software (www.acaclone.com) and the Clustal Omega Multiple Sequence Alignment programme available at http://www.ebi.ac.uk/Tools/msa/clustalo/ (Larkin *et al*, 2007). The search for regions of similarity using Protein BLAST was performed in http://www.uniprot.org/blast/uniprot/ and the nucleotide BLAST search was performed in http://blast.ncbi.nlm.nih.gov/. The protein sequences properties (eg: MW, residues, charge, isoelectric point) were retrieved from http://www.bi.up.ac.za/cgi-bin/emboss.pl?_action=input&_app=pepstats.

4.3.6 Determination of TCF7L2 protein expression

In order to assess cellular localisation and to detect different isoforms of TCF7L2 protein, nuclear and cytosolic proteins were extracted from each sample using the method described in 2.8 and employed for western blotting. Details of AML patients used for protein analysis are listed in Table 4.1. The sources of normal BM and normal $CD34^+$ HPC were previously described in 4.3.1. The different type of myeloid cell lines fractionated for nuclear and cytosolic proteins were previously cultured and are described in Table 2.1. The monoclonal antibodies used to detect TCF7L2 proteins and housekeeping genes Histone H1 and GAPDH by immuno-blotting are described in Table 2.3.

4.3.7 Total protein extraction using TEAB lysis buffer

Cell pellet derived from 20 x 10^6 cells were thawed and 2 µl of Benzonase (5 units per 10 x 10^6 cells) was added directly to each pellet and mixed thoroughly. Tubes were incubated on ice for 30 minutes and 100 µl of TEAB buffer (0.5 M TEAB, 0.05% (v/v) SDS, protease inhibitor cocktail, phosphatase inhibitor cocktail) was added. The mixture was incubated for another 30 minutes on ice and vortexed every 10 minutes. The cell lysates were then centrifuged at 4°C for 10 minutes at 10,000 x g. Potein supernatants were aspirated into fresh tube and stored at -80°C.

4.3.8 Quantitation of TCF7L2 protein expression in AML vs normal samples

The amount of TCF7L2 protein expressed in AML was compared to normal samples (BM and $CD34^+$ HPC). For this purpose, TCF7L2 protein expression in both cytosolic and nuclear fractions were quantified using western blotting and densitometry as described (see 2.8.2). All values for the raw protein signals were corrected from the background integral of the peak (Integral-Bkg signal unit), as determined by densitometry using the Advanced Image Data Analyzer software v4.5 (Raytek Scientific, Sheffield, UK). In order to compensate for variation between different transfers and different primary antibody detections, a control protein extract derived from Jurkat cells was selected as a standard for inter-blot calibration. The Jurkat unfractionated lysate produced a strong signal at 72 kDa and 58 kDa when probed with TCF7L2 antibody, both anti exon 11 and exon 1 TCF7L2. In each sample, densitometric signal values obtained from each protein band were first normalised to the Jurkat signal; this was performed separately for 72 kDa and 58 kDa bands. Normalised values were further standardised to the amount of total protein loaded (in µg) for each sample. Normalised total protein was defined by the sum of 72 kDa and 58 kDa isoforms detected in both subcellular fractions per sample.

4.3.9 Correlation of TCF7L2 mRNA and protein expression

A total of 19 AML patient samples (as in Table 4.1) were chosen for this correlative analysis. The protein samples were processed and quantified as described (4.3.8). The RNA samples from this cohort of AML patients were processed as described in 4.3.1. The qRT-PCR protocol and data analysis were performed similarly to the method described previously in 3.3.5. Due to the small number of samples, the normalised total protein data

didn't meet the assumptions required for parametric test and was positively skewed to the right (data not shown). Therefore, the normalised total protein values were log transformed and compared to the Log_2 normalised qRT-PCR data. The SPSS compute commands and guidelines were used when transforming data (Tabachnick & Fidell, 2007):-

Substantially positive skewness : use Logarithmic (Log 10)

SPSS commands: NEWX = LG10(X)

4.3.10 Statistical data analysis

The degree of correlation between paired variables (normalised total TCF7L2 protein *vs* normalised mRNA) was assessed using Spearman's coefficient (*R*). Test of significance were obtained through Fisher's Z transformation. The significance of differences when comparing two unpaired groups was calculated using the Mann-Whitney Test, and a value of $P < 0.05$ (two-sided) was considered statistically significant. Analyses were performed in SPSS for Windows version 11.5 (SPSS Inc., Chicago, USA).

4.4 RESULTS

4.4.1 Expression of TCF7L2 mRNA 5' variants in AML and normal haematopoietic cells

Exon skipping of TCF7L2 has previously been shown to generate multiple mRNA *TCF7L2* transcripts in normal murine neonates (Weise *et al*, 2010), in adult mice (Nagalski *et al*, 2012) and in several solid tumours (Shiina *et al*, 2003; Tsedensodnom *et al*, 2011). To investigate whether this is also the case in AML and normal HPC, this study examined 40 AML patients, five myeloid cell lines and compared with these normal HPC. TCF7L2 mRNA variants were examined by amplifying cDNA using two sets of primers (N-terminal and C-terminal primer sets), inherent exon complexity being too intractable to be resolved by just a single primer set. Exons located near to the 3' tail (or C-terminus) are mostly alternatively spliced with high GC content sequences that hamper perfect amplification of the whole gene using a single primer set. The N-terminus primer set amplified exon 1 through to exon 12 and thus was able to detect splicing within the 5'region sequence or the primary sequence (refer to Figure 4.1). In addition to complete loss of an exon through splicing, alternative types of exonic variants (namely type-a or -b) were also investigated for the presence of exon 8 and exon 10.

Results of the N-terminus amplification show that multiple *TCF7L2* transcripts were detected in mRNA isolated from both leukaemic and normal haematopoietic samples. Representatives of the PCR product banding patterns are shown in Figure 4.2. Nineteen (47.5%) AML patients produced PCR products of 1128 bp, 15 (37.5%) AML patients generated shorter products (ranging from 350bp – 600 bp), and six (15%) AML patients generated a combination of 1128 bp with the shorter products. Four (80%) out of five myeloid cell lines produced 1128 bp product and only one cell line generated a product of approximately 350 bp in length. For normal $CD34^+$ HPC, 50% of the samples produced product of 1128 bp whilst the remaining had a combination of 1128 bp with the shorter variants. Normal BM produced only the 1128 bp with the exception of one sample which had a slightly longer product of 1197 bp. To confirm the identity of these PCR products which covered only from exon 1 until exon 12 (i.e. primer set N-terminus) as mentioned earlier, DNA sequencing was carried out. PCR product that contained DNA sequences with complete open reading frame (ORF) were given a code (eg: P1), and PCR products

containing premature intervening stop codons were excluded from this analysis. Sequencing of the 1128 bp PCR product confirmed its homology with the published *TCF7L2* transcript variant 2, mRNA (NM_030756.4) from NCBI for the region covered by the primer pair; hereafter referred to as type P1 as shown in Table 4.4. Protein translation of this 1128 bp product sequence (or type P1) consistently generated a complete ORF (without intervening stop codons between exons). The one normal BM sample with the longest bp detected so far (1197 bp) was shown to retain a full exon 4 as detected by sequencing. The variable presence of exon 4 doesn't cause any changes in the downstream ORF apart from the local codon insertion. BLAST search showed that the region from exon 1 to exon 12 of this variant had 100% similarity to *TCF7L2* transcript variant 1 (NM_001146274.1) and variant 13 (NM_001198531.1), hereafter referred to as type P2. Sequencing of all samples confirmed the presence of exon 8 variant type-b which is longer than variant type-a. Similarly, all samples also harbour exon 10 variant type-b which is shorter than variant type-a, as defined by NCBI.

For the shorter PCR products (350 bp – 600 bp), sequencing data show that, irrespective of combination of exons, only one product actually had a complete ORF however this sequence has not been published in NCBI. This transcript (369 nucleotides) was referred to as type P3. The exonic identities of all these sequences are shown in Table 4.4.

To conclude, the most frequently detected transcript for the 5'-primary region of *TCF7L2* seen in AML blasts and in normal samples contains 1128 nucleotides. Sequencing of this transcript has confirmed the inclusion of exons 1 - 12 with a consistent loss of exons 4 and 6. All samples showed the presence of exon 8 variant type-b and exon 10 variant type-b. The next section will further characterise the splicing at the 3'-end of *TCF7L2* gene.

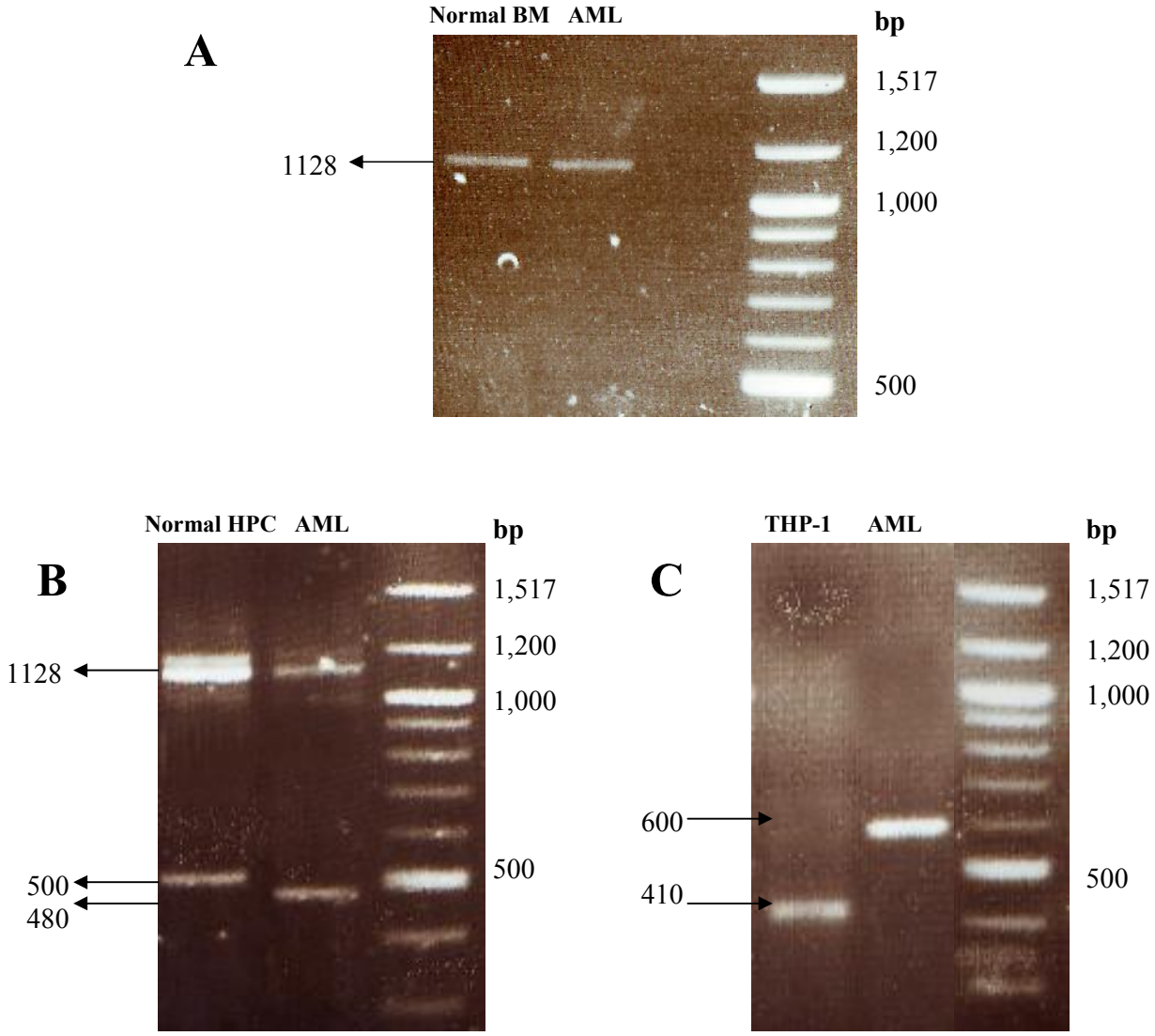


Figure 4.2: Representative PCR products showing variants of *TCF7L2* amplified by the N-terminus primer set. **(A)** Gel showing presence of the 1128 bp product. **(B)** Gel showing combination of the 1128 bp product and shorter products. **(C)** Gel showing AML and cell line sample (THP-1) with shorter PCR products. Ladders indicate product size in bp and all gels were electrophoresed using 2% (w/v) agarose.

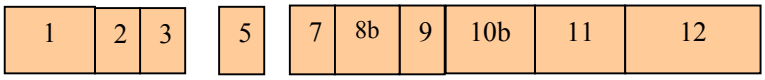
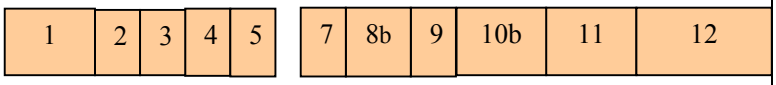
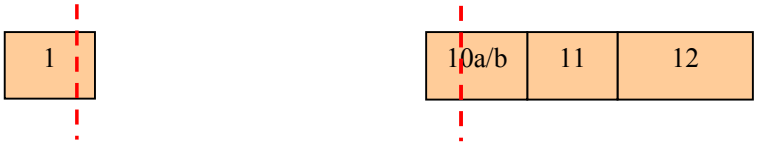
Exons combination of *TCF7L2* 5'- primary sequence confirmed by sequencing	Type of variant	PCR product length in bp	Present in AML?	Present in Normal?
	P1	1128	Yes	Yes
	P2	1197	No	Yes
 * exon 1 and exon 10 were truncated (shown above)	P3	369	Yes	No

Table 4.4: Comparison of 5'- primary sequence exon combinations that give rise to complete ORF. Exons shown were based on sequencing data. The predicted MW (kDa) was calculated from the start codon `ATG' in exon 1 until the last codon encoding exon 12. For type P3, the sequences within exon 1 and 10 were truncated, indicated by dashed line (red).

4.4.2 Complex exon combinations were detected at the C-terminus region of TCF7L2

Previous reports have shown that splicing of various exon combinations could occur between exons 14 and exon 18 of the C-terminus resulting in a shift within the ORF, therefore the proteins will utilize different stop codons within exon 18 (Shiina *et al*, 2003; Arce *et al*, 2006; Weise *et al*, 2010). To further verify the molecular diversity of variants expressed in the C-terminus region in AML blasts, PCR amplification and sequencing was performed and the identity of each sequence was compared to the published nomenclature of the C-tail TCF7L2 variants (Weise *et al*, 2010). The highly diverging C-termini of TCF7L2 had been reported to generate alternative splicing exons known as C-tail-E (E1-E4), C-tail-M (M1-M3) and C-tail-S (S1-S8).

Results from C-terminus PCR show that, in contrast to the multiple patterns of PCR banding produced at the N-terminus, all samples amplified using the C-terminus primer produced only one pattern of PCR banding (consisting of four bands) as detected by gel electrophoresis (Figure 4.3A). To independently characterise each PCR band, gel excision followed by a second round of TD PCR (as described in 4.3.2) was performed for each PCR band prior to sequencing. Gel electrophoresis following the second round of TD PCR reproduced similar PCR products, now as individual bands of the following sizes: 200, 150, 125 and 80 bp (Figure 4.3B). These individual bands were successfully purified and sequenced. Together, this suggests that the four PCR products were genuine variants. However, due to the identical sizes of exons 15 and 16 (i.e. 73 bp), the identity of variants carrying either one of these two exons could not be determined by size alone.

Following DNA sequencing, splicing patterns were compared between leukaemic and normal samples. Results as summarised in Table 4.5 illustrate the exon composition for all four bands found in leukaemic and normal samples. The 125 bp and 80 bp PCR products were similar between AML, cell lines and normal samples. However, sequencing data for the other bands (200 bp and 150 bp) showed that exon 15 and 16 are mutually exclusive and inclusion of exon 15 was observed only in the leukaemic cells. The presence of exon 17 was not detected in all samples analysed. In terms of the nomenclature, the 80 bp band was confirmed to be the correct tail for the variant carrying the C-tail-M1 as published previously, whilst the 125 bp was confirmed as a correct tail belonging to the variant carrying the C-tail-M2. The 150 bp band was confirmed to be either the C-tail-E1 or E3 and

finally, the 200 bp was found to be the correct sequence for either C-tail-E2 or E4. This was done by combining BLAST search with previous literature on nomenclature.

Data obtained from DNA sequencing were subjected to further analyses. The prediction of MW is made based on the DNA sequences. The protein coding region of the aforementioned six variants (C-tail E1-E4, M1 and M2) were compared as illustrated by multiple alignments in Figure 4.4. The C-tail-E1 and E3 induce in-frame deletion of the protein sequence after exon 13 and give rise to an exon 18 that harbours a stop codon at a position similar to type E2 and E4. The C-tail-M1 and M2 will induce out-of-frame protein sequence after exon 13 and exon 14 respectively, therefore creating a stop codon within exon 18 at earlier position.

The combined data obtained from both primer sets (section 4.4.1 and section 4.4.2) generate the full length *TCF7L2* sequences, as presented in Figure 4.5. Because the 5'-primary sequence type P1 was detected in both AML and normal samples, this transcript was further analysed to determine sequences of the full length TCF7L2. The relationship between 5'-primary sequence type P1 joined to different 3'- tails (or C-tails) is predicted to generate six different protein isoforms with MW varying from 56 kDa to 72 kDa.

In terms of functional roles, only C-tail-E1, E2, E3 and E4 preserve the cysteines within the C-clamp region identified by CRARF or CRALF motifs within exon 15 or exon 16 respectively. However only C-clamp with the CRARF motif (i.e. with exon 15) was reported to cooperate with β-catenin and the p300 complex for activation of downstream target genes (Hecht & Stemmler, 2003). No functional relevance has been reported for the CRALF motif so far. Among the group of *TCF7L2* C-tail variants detected, the inclusion of exon 15 that gives rise to E1 and E2 was not detected in normal BM or normal $CD34^+$ HPC. Even though the leukaemic samples produced all types of variants, the variants incorporating exon 16 were more commonly detected among the cohort of AML samples tested.

In summary, there are initially four bands detected by amplifying the 3'-tail region using the C-terminus primer set. These four bands were shown to represent the C-termini of six types of variants that, in turn, were predicted to generate six protein isoforms. The next section will determine whether these predicted transcripts are being translated.

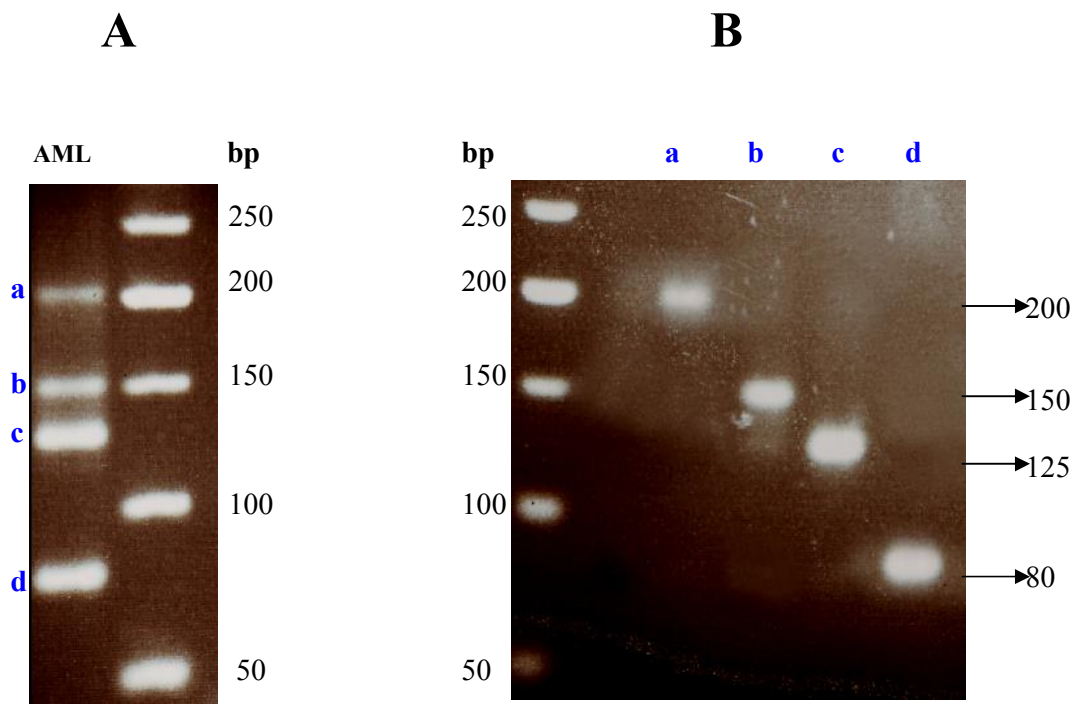


Figure 4.3: TD PCR of *TCF7L2* as amplified by the C-terminus primer set from a representative AML sample. (A) Gel electrophoresis using 3% (w/v) agarose gel showing presence of four PCR products **(a-d)** after initial TD PCR in the AML (and in all other samples) **(B)** Gel electrophoresis after a repeat TD PCR confirmed the identity of each PCR product **a-d** isolated from the agarose gel. Each product showed correct bp as seen by the first round TD PC (200, 150 , 125 and 80 bp). The ladder shown here is a 50 bp ladder.

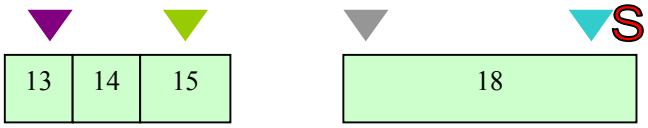
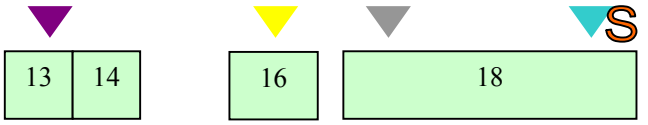
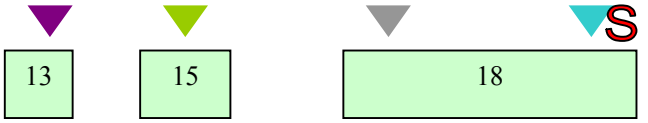
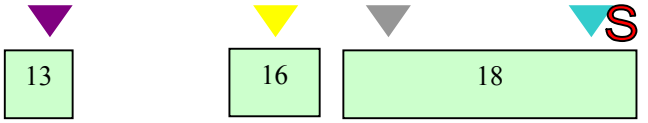
Exon combination of *TCF7L2* 3'tail sequence confirmed by sequencing	Type of variant	PCR product length in bp	Present in AML?	Present in normal ?
	E2	200	Yes	No
	E4	200	Yes	Yes
	M2	125	Yes	Yes
	E1	150	Yes	No
	E3	150	Yes	Yes
	M1	80	Yes	Yes

Table 4.5: Comparison of 3' tail (or C-tail) exon combinations seen in AML, leukaemic cell lines and normal samples (BM and $CD34^+$ HPC). Exons shown were based on sequencing data. The NLS motif is shown in purple, the cysteine rich C-clamp motifs are shown in green (CRARF) and yellow (CRALF) , the RKKKC motif at the start of exon 18 is shown in grey, Ctbp binding motif is shown in blue and stop codon is shown by the symbol "S". The nomenclature for E1-E4 and M1-M2 is from Weise *et al*, 2010

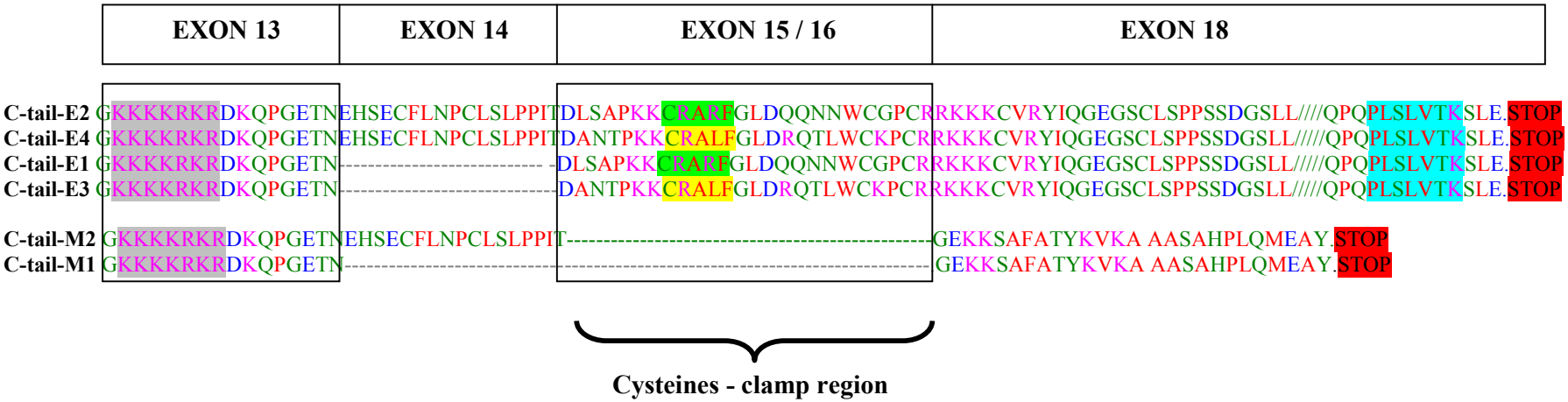


Figure 4.4: Multiple sequence alignments showing predicted amino acid sequence diversity starting from exon 13 until exon 18. Shown are the four types of 3'-tails (or C-tail-E) that encoded motifs for the C-clamp region and other target binding regions (C-tail E1 – E4). The cysteine motifs are highlighted in green and yellow, the NLS motif is highlighted in grey and the CtBP binding motif is highlighted in blue. The alternate exons are boxed. The amino acids encoding exon 18 of C-tail E1 - E4 were cropped to fit in the illustration. The position of each stop codon is highlighted in red.

4.4.3 Investigation at the protein level identifies 72 kDa, 58 kDa and 56 kDa TCF7L2 protein isoforms in AML patients

The data above suggest that six protein isoforms are predicted to be expressed in AML, illustrated in Figure 4.5B. To determine whether these were expressed at the protein level, western blotting was performed (see 4.3.6) and nuclear / cytosolic lysates were probed with a TCF7L2 monoclonal antibody raised to an epitope surrounding Leu330 in exon 11.

A predominant 58 kDa protein band was detected in 19 AML blast nuclear fractions but not in the cytosol following probing with exon 11 antibody. A 72 kDa protein was also expressed but only found in the cytosolic fraction (Figure 4.6A). Doublets of protein bands at approximately 56 kDa and 58 kDa were also seen in several AML samples, suggesting co-expression of different isoforms (see below). Several AML samples also expressed both 72 kDa and 58 kDa proteins in the nucleus, depicted by AML samples in lanes 6 – 7 in Figure 4.6B. There was no consistent banding detected at lower than 56 kDa, suggesting that any shorter isoform of less than 56 kDa (if any) is not being translated efficiently (see Figure 4.6B).

The data above demonstrate that all bands detected by SDS-PAGE were approximately 7 kDa larger in size than the predicted MW. The heaviest TCF7L2 protein observed was 72 kDa and could possibly match the predicted 65.3 kDa (or 65.4 kDa) isoform. Detection of similar isoforms by other studies has also shown the presence of 72 – 75 kDa (instead of 65 kDa) in nuclear but not cytosolic fractions (Weise *et al*, 2010). Likewise, there is only 72 kDa in whole protein lysates (Struewing *et al*, 2010; Tsedensodnom *et al*, 2011). The MW seen in the western blot is also 7 kDa higher than the size indicated in UniProt (eg: TCF7L2 Isoform 8, Q9NQB0-8). Possibly, this protein migrated slower than predicted when electrophoresed under SDS-PAGE, due its highly charged nature. As for 56 – 58 kDa isoforms, other studies (Vacik *et al*, 2011; Le *et al*, 2011) had indicated presence of TCF7L2 bands at around this same region when probed with TCF7L2 antibodies.

In summary, the distribution of TCF7L2 protein isoforms found in AML encompassing exons 1 – 18 of the full length gene follow the PCR variant exon composition data where it generates the 72 kDa, 58 kDa and 56 kDa isoforms seen in western blot.

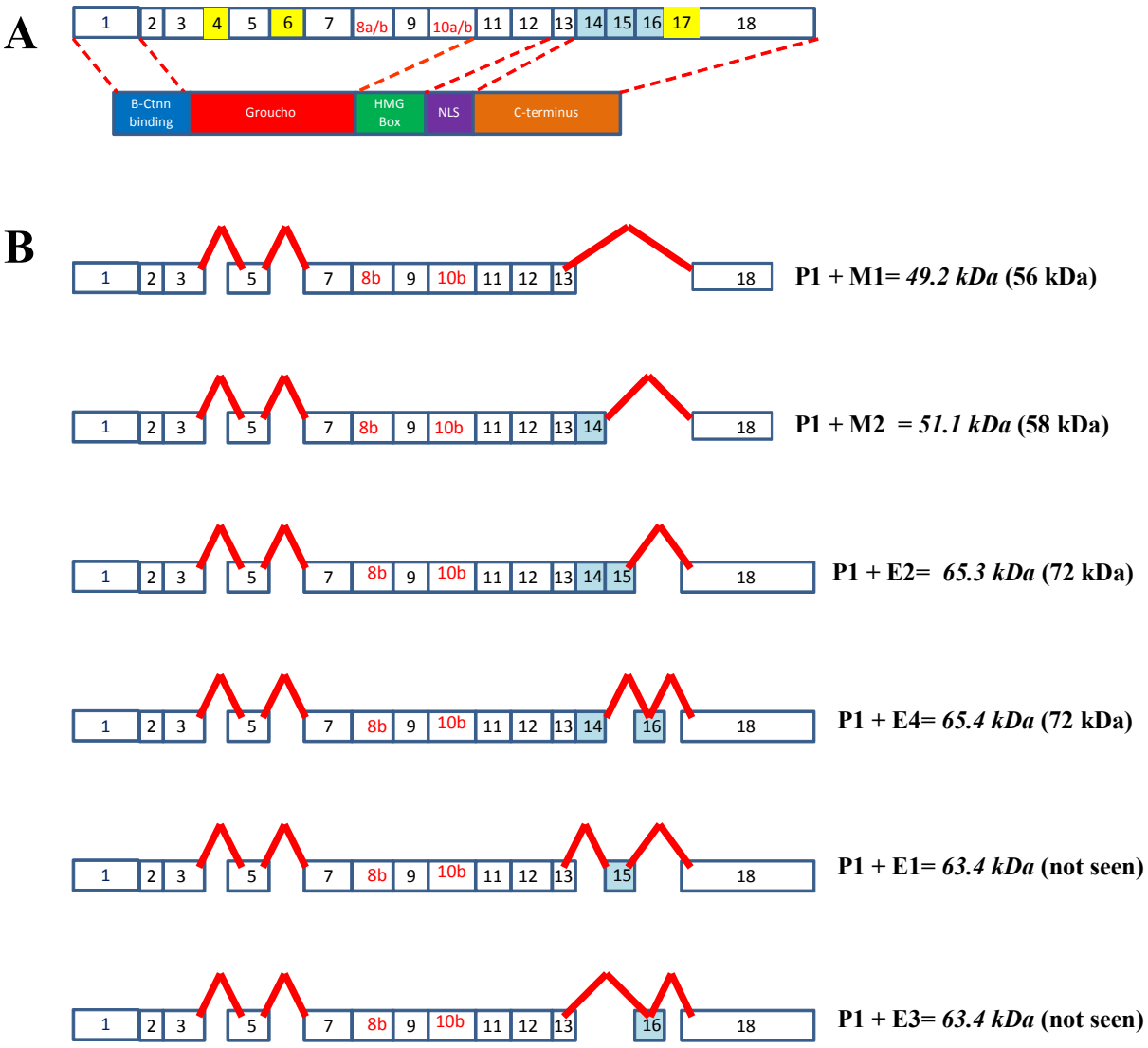


Figure 4.5: Schematic diagram showing a summary of TCF7L2 expression. (A) Established exons of *TCF7L2* are represented as white rectangles, commonly absent exons are shown in yellow and alternative exons as blue shaded rectangles. Protein domains are indicated below the corresponding exons. Exon numbers in red font indicates the presence of dual variants. **(B)** Exon combinations used to predict the ORF for each TCF7L2 isoform with corresponding protein MW. Predicted MW (italic) was calculated from the first "ATG" in exon 1 until the stop codon in exon 18. The apparent MW (in parentheses) at which these protein run on SDS PAGE (and western blotting probed with exon 11 antibody). Only full combinations of 5'-primary sequence (or type P1) joined to different type of 3'- tails (or C-tail) detected by sequencing are shown.

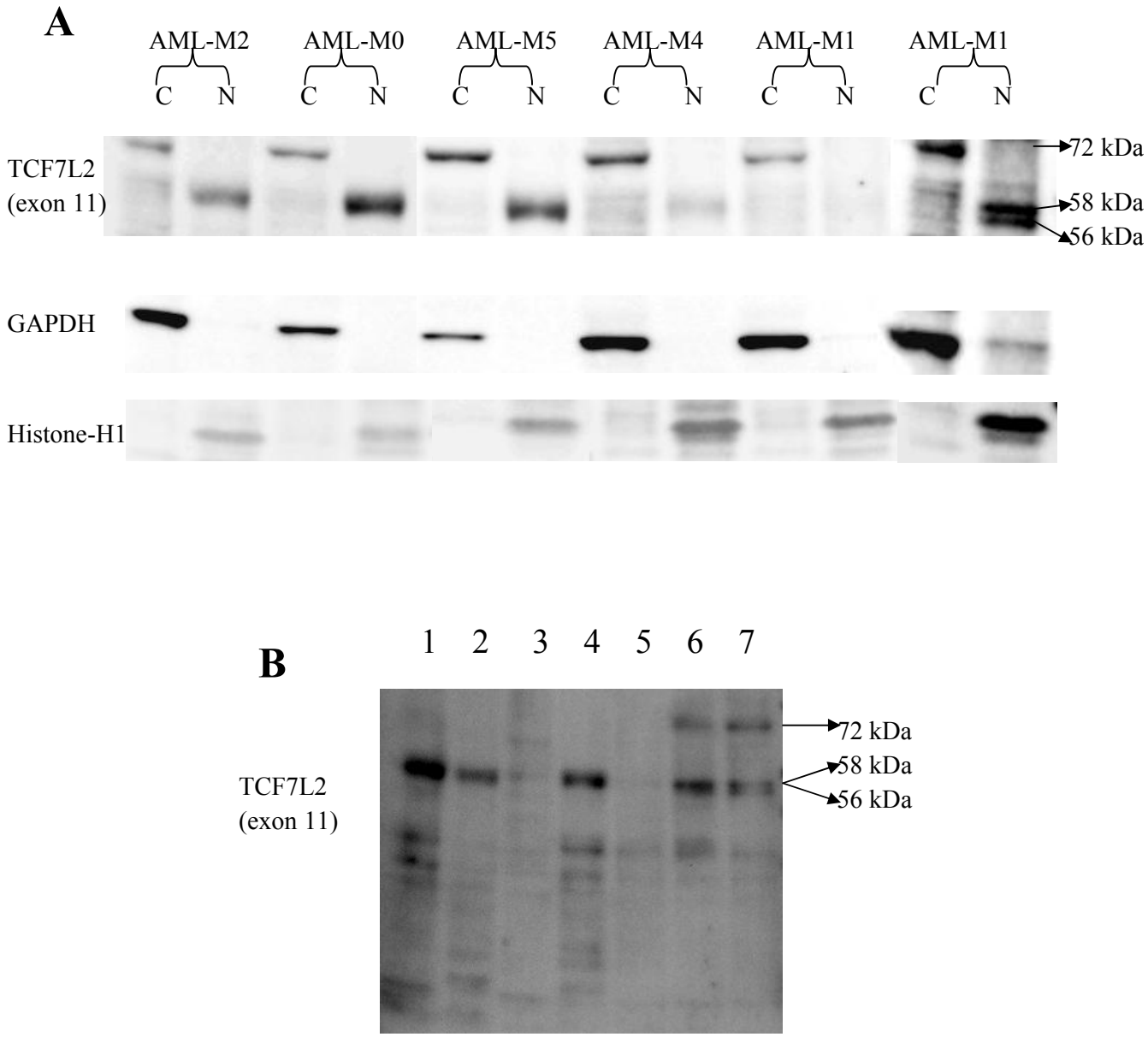


Figure 4.6: Distribution of nuclear and cytosolic TCF7L2 protein isoforms detected by western blotting in AML. Protein extracts were prepared from indicated patients and the blots were probed with a monoclonal TCF7L2 antibody (exon11 epitope). Loading controls used to show the purity of each fraction were probed with GAPDH and Histone H1. **(A)** Cytosolic and nuclear protein extracts were prepared from primary AML samples with different FAB subtypes (M0-M5). **(B)** Seven nuclear extracts of AML samples of FAB M1 (not listed in Table 4.1) and protein lysates were extracted using the TEAB buffer method (see 4.3.7), 20 μg protein was loaded and probed with TCF7L2 antibody (exon11 epitope).

4.4.4 Expression of the 72 kDa and 58 kDa TCF7L2 isoforms correlates with TCF7L2 mRNA expression.

The previous chapter correlated *TCF7L2* microarray expression (spanning exon 5 – 10) with qRT-PCR data derived from a similar region. Here, mRNA expression was correlated with expression at the protein level (as described in 4.3.9). These data (Figure 4.7) show a positive correlation between normalised mRNA (analysed by qRT-PCR) and protein expression (sum of 58 kDa and 72 kDa) in AML patient blasts (R=0.498, *P*= 0.042 using Spearman correlation coefficient). These findings suggest that expression of TCF7L2 protein is regulated at the transcriptional level.

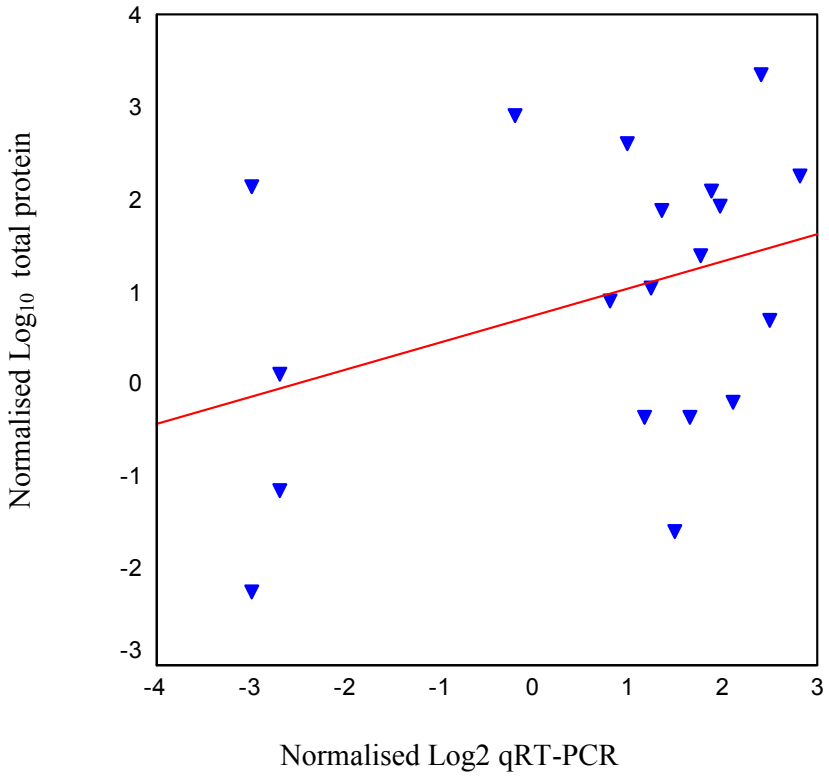


Figure 4.7: Correlation between TCF7L2 protein and mRNA expression. Correlation was done using normalised Log_{10} total protein expression of TCF7L2 72 kDa and 58 kDa (detected by exon 11 / Leu330 epitope) and normalised Log_2 mRNA expression using matched AML patients (n=19). Spearman correlation, R=0.498; $P < 0.05$.

4.4.5 Expression of TCF7L2 protein isoforms in normal haematopoietic cells

In order to determine whether TCF7L2 protein isoforms are aberrantly expressed in AML, it was necessary to analyse expression in normal haematopoietic cells. A total of four normal BM and four $CD34^+$ HPC were included in the analysis. Due to the different localisation of protein, expression of isoforms was measured separately for nuclear and cytoplasmic protein fractions.

As seen in Figure 4.8, normal cells express the similar predominant 56 or 58 kDa nuclear protein as seen in AML, thus being consistently detected when probed with TCF7L2 exon 11 antibody. The normal cells also expressed the 72 kDa protein in their cytosolic fractions.

Each band detected in the cytoplasmic and nuclear fraction was quantified (see 4.3.8) and analysed separately. When comparing the relative level of TCF7L2 protein expression between AML and normal samples, the mean protein expression in AML is lower than in normals (both total BM and $CD34^+$ HPC). However, AML subtype M0 was found to be expressing higher levels of TCF7L2 protein than normals (denoted by outlier values), whereas other subtypes of AML generally produced lower mean protein expression for both isoforms (Figure 4.9 - Figure 4.10). The mean protein expression of normal $CD34^+$ HPC was apparently high as compared with AML and total BM. Work performed since these experiments were completed show this unexpected conclusion is likely to be erroneous (see Discussion section 4.5); additionally this could result from the use of cultured normal controls for western blotting. There were insufficient cells and samples to perform nuclear extraction from fresh normal $CD34^+$ HPC derived from CB. Similarly, total BM is not a perfect match for blasts AML as it contains only a small proportion of blasts.

In summary, protein isoforms detected in normal cells are similar to those in AML. Due to sampling limitations, the relative expression of these isoforms between AML and normal is inconclusive at the protein level (see Discussion).

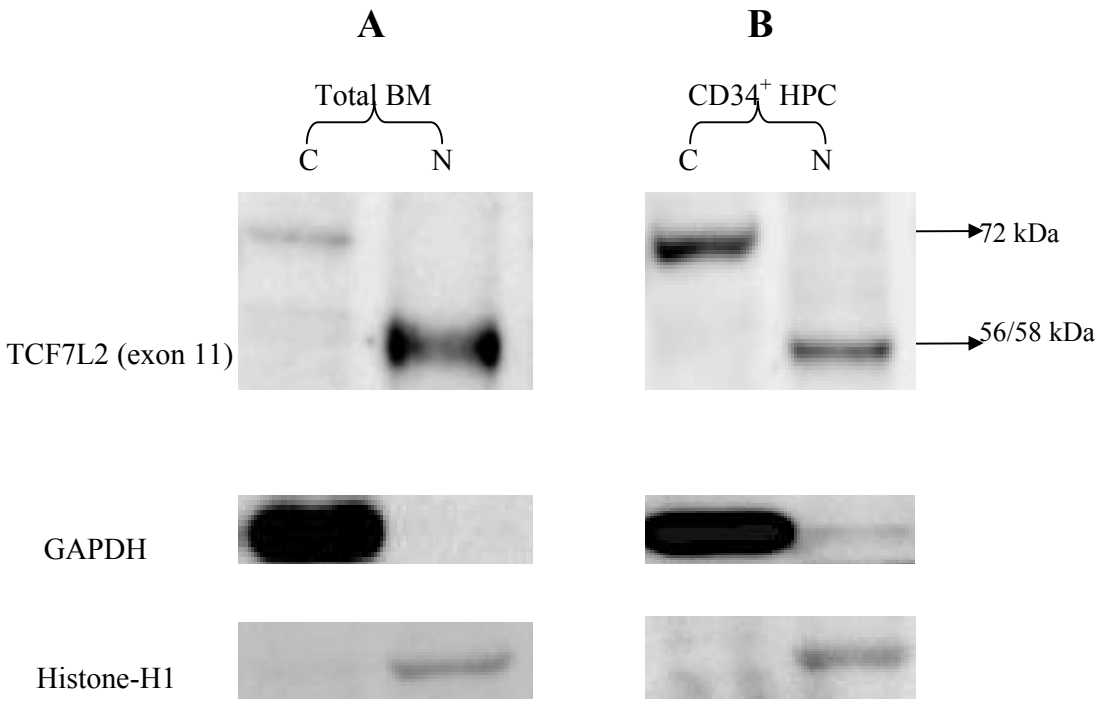


Figure 4.8: Distribution of nuclear and cytosolic TCF7L2 protein isoforms detected by western blotting in normal BM and CD34$^+$ HPC. (A) Normal BM sample with cytosolic and nuclear protein extracts, probed with exon 11 TCF7L2 monoclonal antibody. **(B)** Normal CD34$^+$ HPC day 5 culture sample with cytosolic and nuclear protein extracts, probed with exon 11 TCF7L2 monoclonal antibody. Loading controls used to show the purity of each fraction were probed with GAPDH and Histone H1.

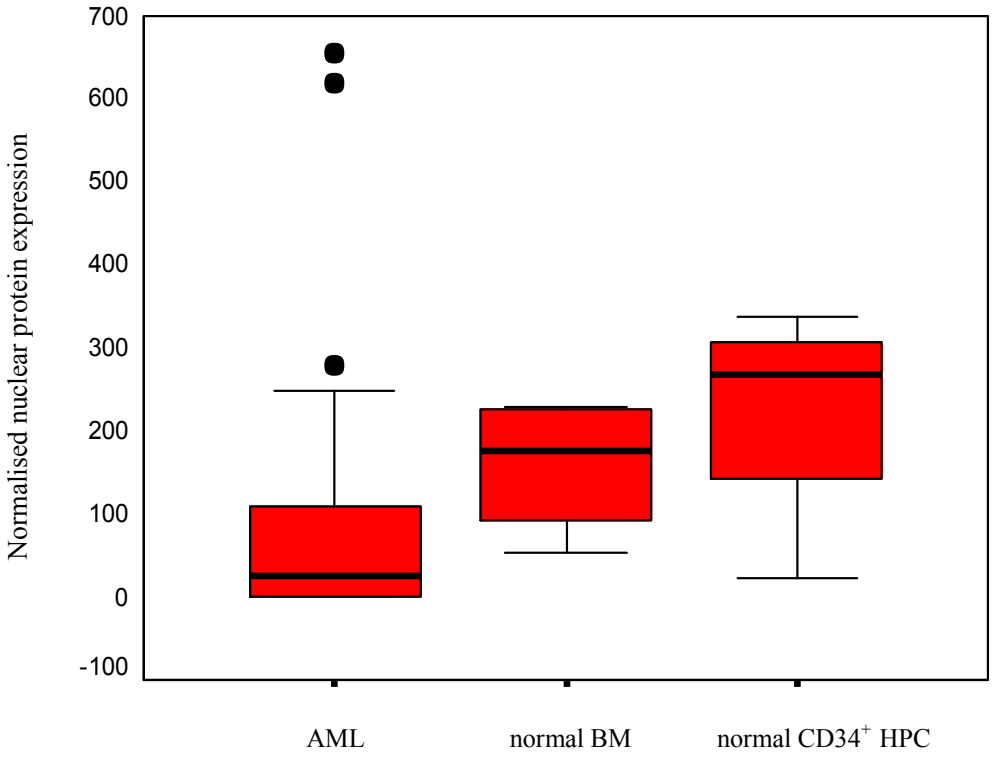


Figure 4.9: Relative expression of 56/58 kDa nuclear TCF7L2 probed with exon 11 antibody. Statistical significance was calculated using the Mann-Whitney Test compared to AML. • denotes data outlier. There is no significant difference in levels.

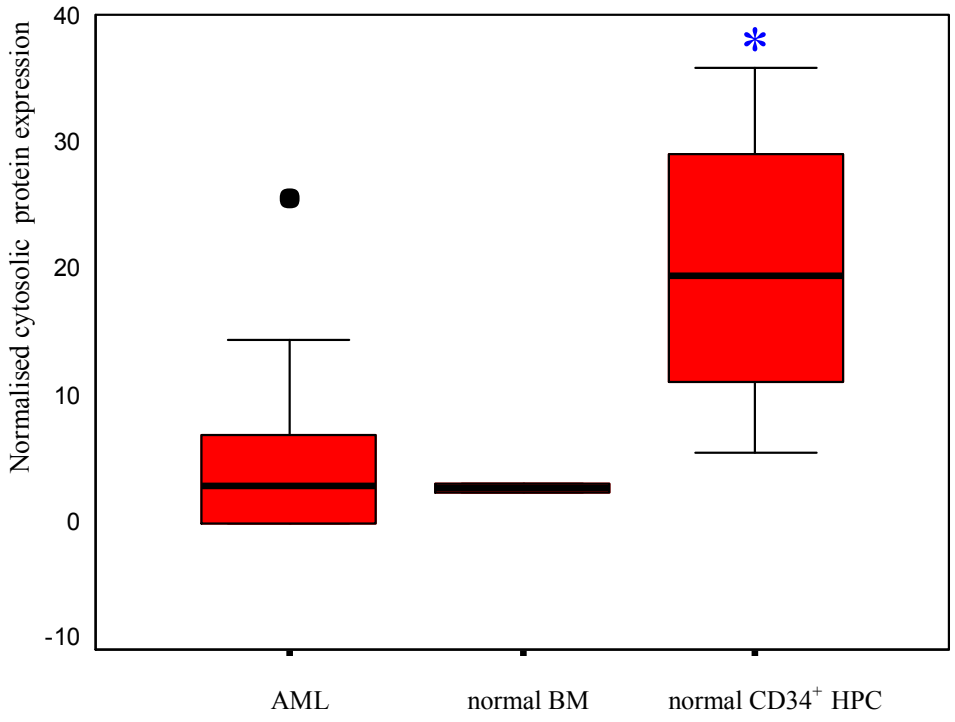


Figure 4.10: Relative expression of 72 kDa cytosolic TCF7L2 probed with exon 11 antibody. Statistical significance was calculated using the Mann-Whitney Test compared to AML, with *P <0.05. • denotes extreme value.

4.4.6 Identification of 56 kDa and 58 kDa isoforms as the dominant proteins expressed in myeloid lines

The data presented so far show that endogenous TCF7L2 proteins are readily detected in primary samples with different TCF7L2 isoforms being translated in both AML and normal samples. In the nucleus, the two dominant TCF7L2 proteins (56 / 58 kDa) are detected; however the functional consequences at the cellular level of these proteins are still poorly understood.

Eight different myeloid cell lines were assessed by western blotting, in which nuclear and cytosolic fractionation was performed similarly to 4.3.6. Either the 56 or 58 kDa isoform was found to be the most dominant isoform seen in the nuclear fraction of all cell lines (see Figure 4.11). In this figure, a doublet band of 56 and 58 kDa was also visible in NB-4 cells. Towards the end of this study, a number of improvements were made to the nuclear and cytosolic protein extraction. The improved extraction procedure was able to resolve the doublet much better than the old method and has helped to elucidate the banding of 56 kDa and 58 kDa and also the absence of any TCF7L2 protein at lower molecular weights, illustrated in the K562 cell line (Figure 4.12).

In summary, two TCF7L2 isoforms were expressed in the nuclear fraction and TCF7L2 was absent from the cytosolic fraction in all myeloid cell lines tested. This finding has given a clear picture of the overall expression of TCF7L2 in myeloid cell lines which will assist further TCF7L2 functional studies in the next chapter.

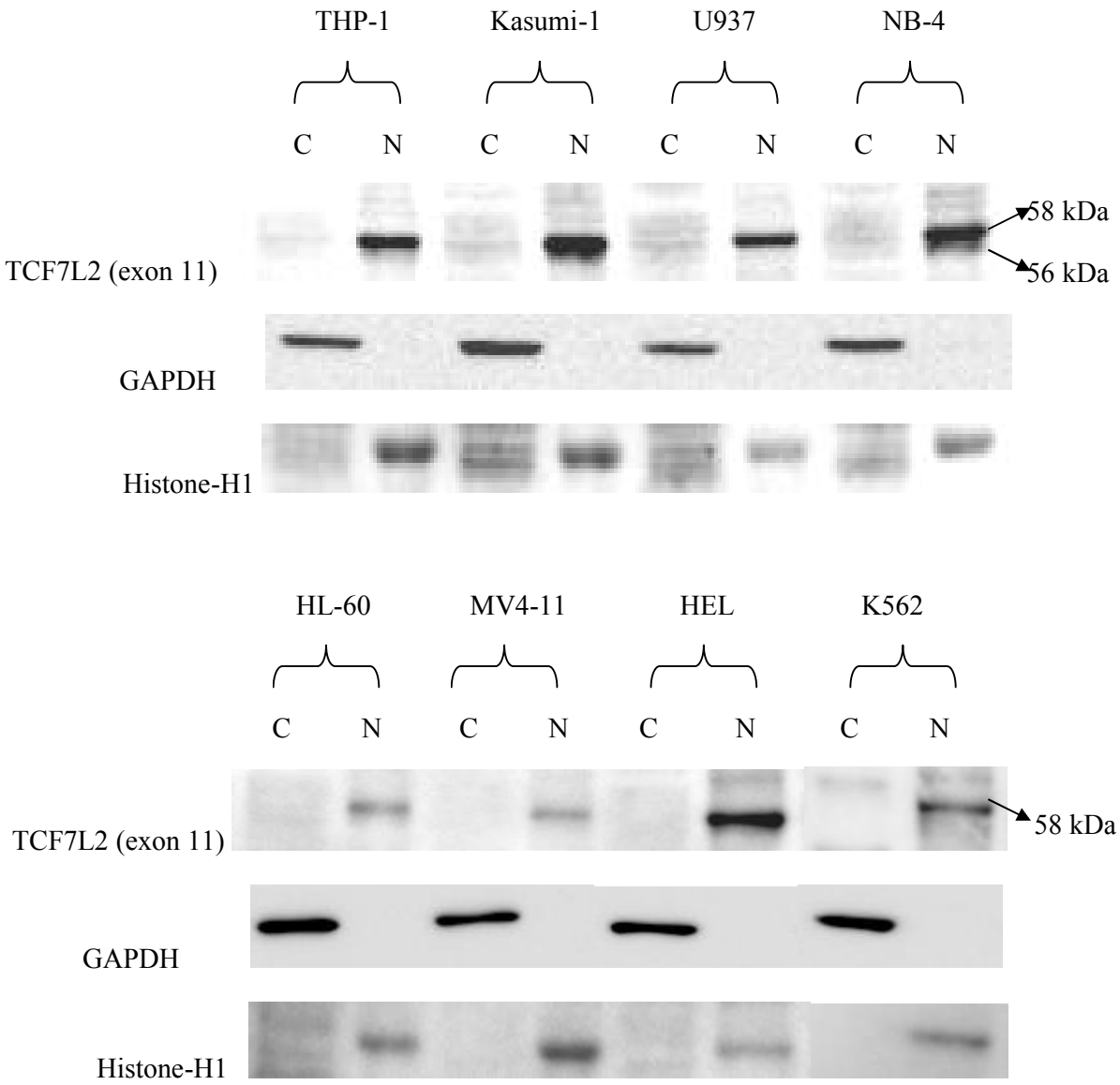


Figure 4.11: Distribution of nuclear and cytosolic TCF7L2 protein isoforms in myeloid cell lines. Protein extracts were prepared from indicated cells and western blot analysis was performed. The blots were probed with TCF7L2 antibody (exon11 / Leu330 epitope). Loading controls used were anti-GAPDH and anti-Histone H1.

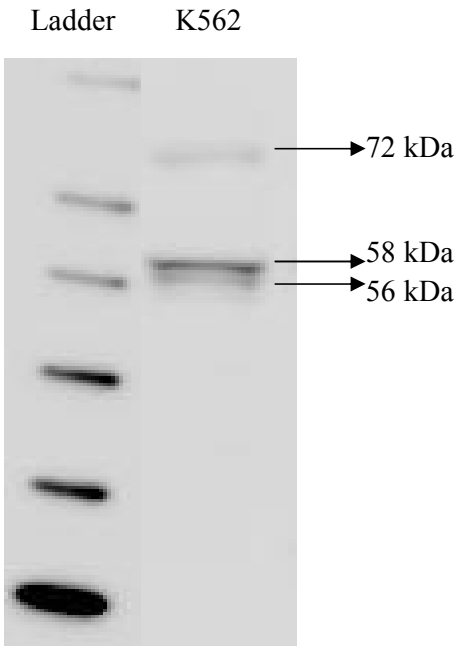


Figure 4.12: Improved detection of TCF7L2 protein isoforms in the K562 myeloid cell line. The blot was probed with TCF7L2 antibody (exon11/Leu330 epitope), and extracted using improved TEAB buffer protocol (see 4.3.7). Image is courtesy of Dr Richard Darley.

4.5 DISCUSSION

4.5.1 Comparison of TCF7L2 alternative splicing in AML and normal samples.

The study here revealed that within normal HPC and AML samples, a number of TCF7L2 variants are generated as a result of splicing at both the N-terminus (or 5'-primary) and the C-terminus (or 3'-tail). At the region near to the N-terminus, the complete length of cDNA amplified by the method used is 1128 bp (5'-primary type P1 as illustrated in Table 4.4) representing exon 1 – exon 12 with consistent absence of exons 4 and 6 confirmed by sequencing in both normal HPC and AML. Using the same primer pair, Shiina *et al* reported there was no splicing variant across exons 1 – exon 12 in other renal cancer cell lines; A498, Caki-1, and Caki-2 (Shiina *et al*, 2003), showing differential variant expression in different cancer types.

Comparative studies of haematopoietic cells have not been published. In this study, both exon 4 and exon 6 are found to be commonly absent in normal samples and also in AML blasts. With the exception of one normal BM sample where exon 4 was detected; however, the protein sequence encoded by exon 4 does not comprise any known functional domain as revealed by a BLAST search. Other types of tissue exhibited only the loss of exon 4 but not exon 6, such as in hepatocellular carcinoma and in zebrafish (Young *et al*, 2002). The presence of exon 4 and absence of exon 6 has been reported in murine intestinal eplithelia (Nazwar *et al*, 2009) and in human pancreatic islets with pro-apoptotic phenotype (Le *et al*, 2011). In addition, differential expression of exon 4 and exon 6 has been implicated during developmental stages in which exon 4 is lowest in foetal endothelial cells and highest in adult endothelial cells (Struewing *et al*, 2010).

Apart from the identification of 5'-primary type P1 and P2 (see Table 4.4), electrophoreses data also revealed an abundance of shorter DNA bands. Sequencing had confirmed only one PCR product of 369 bp to be informative, given the name 'type P3' (i.e. with a correct ORF without intervening stop codons). This transcript showed loss of exon 2 through to exon 9 and was only detected in leukaemic cells not in the normal samples tested. The functional significance of this variant is unknown, although co-repressor proteins Groucho and TLE have been reported to interact with TCF within this missing region and in the HMG domain within exon 12 (Daniels & Weis, 2005); however, one

functional study conducted in epidermal cells suggested that the absence of the Groucho domain did not affect transcriptional activity (Nguyen *et al*, 2009).

Depending upon the exon combinations near to the C-terminus, different ORFs are used for translation of exons 15, 16 and 17. Previous reports have used this distinction to classify TCF7L2 transcripts into three groups giving rise to TCF7L2 with C-tail-E, TCF7L2 with C-tail-M and TCF7L2 with C-tail-S protein isoforms (Tang *et al*, 2008; Weise *et al*, 2010). Normal HPC and AML also differ by their C-terminus (or 3'-tail) exon combinations. Sequencing data from the C-terminus primer revealed that there are two variants with complete C-clamp found in both AML and cell lines but not in normal samples. These are the C-tail-E2 and C-tail-E1 that harbour a CRARF motif. The other variants, namely C-tail-E3 and –E4 that display a CRALF motif were detected in AML, cell lines and normal samples. CRALF harbours an incomplete C-clamp and its function has not been extensively studied.

TCF7L2 with C-tail-M variant which lack both CRARF or CRALF motifs was also detected by PCR and sequencing in both AML and normal samples. This short C-terminus isoform lacks a functional C-clamp and was not able to bind to the coactivators, p300, and has also been reported to produce decreased activation of WNT/β-catenin target gene promoters (Hecht & Stemmler, 2003). Taken together, the existence of function-specific variants of TCF7L2 could produce distinct effects affecting promoter-specific transactivation and cell phenotype.

4.5.2 Predicted and observed TCF7L2 protein isoforms.

Western blot analysis was performed to confirm TCF7L2 protein expression in AML and normal samples. As presented in Figure 4.5, the combination of C-tail-E2 or C-tail-E4 with the N-primary type-P1 (the most prevalent N-terminus variant confirmed by sequencing) was predicted to produce a protein of 65 kDa. The heaviest protein detected by western blotting migrated at 72 kDa. This isoform was found to be expressed in all normal HPC and BM, but only in 61% of AML samples analysed. The significance of the 72 kDa protein in AML is still unknown but CtBP has been described to interact with the domain in the C-tail-E of this isoform to repress expression of certain target genes (e.g. *NKD*) in the absence of WNT stimulation in parallel with Groucho (Fang *et al*, 2006). In addition to

repression, the nuclear coactivator also binds to the C-tail-E to promote *CDX1* promoter activation showing coupling effects of this protein.

In relation to the published variants, there are at least 13 transcripts of TCF7L2 published in NCBI, and the UniProt entry describes 12 predicted protein isoforms produced by TCF7L2 alternative splicing. The closest match to the longest variant carrying C-tail-E2 (seen as 72 kDa by western blotting) shows 100% similarity to the transcript variant 2, mRNA (NM_030756.4), as supported by BLAST search. A protein sequence similarity search in Uniprot also showed 100% homology to TCF7L2 Q9NQB0-8 isoform 8, a fully reviewed protein, suggesting that the 72 kDa isoform described in this study is the full length transcript, and C-tail-E2 is the most described E-tail for full length TCF7L2. At transcript level, sequencing revealed that C-tail-E2 (containing exon 15) was only found in AML and cell lines but not in normal samples. This suggests the presence of a domain which may be of regulatory significance in leukaemia.

Very limited studies were performed using C-tail-E4, although this isoform could also be equally expressed. The 72 kDa C-tail-E4 protein (with exon 16) had only 99% similarity to transcript variant 2, mRNA (NM_030756.4) and no perfect match was found in NCBI database. The closest match is 99% similarity to TCF7L2 Q9NQB0-8 isoform 8, with variation of amino acid between position 435 until 455. Whilst this C-tail-E4 isoform has not been fully reviewed, it is impossible to distinguish the 72 kDa isoform derived from C-tail-E2 or C-tail-E4 by migration alone.

The combination of alternative exons in the C-terminus (or 3'-tail) produced different ORFs which modified the reading frame in exon 18 and lead to another shorter group of proteins identified as C-tail-M. For AML, 67% samples expressed this protein at nuclear level identified as 56 kDa or 58 kDa isoforms by western blotting. In some AML cases (~15%), doublets of protein bands were seen to migrate between 56 and 58kDa suggesting the expression of C-tail-M1 and C-tail-M2 respectively in combination with the common N-primary type-P1. The detection of this doublet is consistent with previous reports that have shown to be near identical MW to the variants within the TCF7L2 with C-tail-M group, and which was resolvable by SDS-PAGE (Tsedensodnom *et al*, 2011). This protein was found to be strongly expressed in both cell lines and normal cells. In line with other studies, C-tail-M was expressed in U937 cells (Baek *et al*, 2009) and detected only in nuclear fractions (Golan *et al*, 2004; Shitashige *et al*, 2007).

Similarity search showed that the putative 56 kDa transcript is 100% similar to variant 6 (NM_001146286.1), with 100% homology to T-cell factor-4 variant B (B4DRJ8_HUMAN) which is an unreviewed protein, published in UniProtKB/TrEMBL, whereas the 58 kD transcript has a 100% similarity to transcript variant 11, mRNA (NM_001198529), and the closest match in Uniprot is TCF7L2 Q9NQB0-10 isoform 10, a fully reviewed protein but only with 99% homology. The amino acid variation starts from position 233 until 236. In the context of AML, it is still unclear which isoforms might play the critical role in predisposition to TCF activation instead of repression.

4.5.3 Differential expression of TCF7L2 protein

TCF7L2 has been reported to be differentially expressed in colon and neural cells and different splicing events were detected during development. In normal HPC, high expression levels of both 72 kDa and 56 / 58 kDa TCF7L2 were detected in early $CD34^+$ culture (day 5), owing to the growth factor enriched culture condition, where the cells were actually harvested for western blotting. Day 13 cultures (and later) derived from $CD34^+$ cells are devoid of TCF7L2 expression that may equate to more differentiated status (Supplementary 4.3). Differential TCF7L2 expression with development has been observed in other tissue types. The intestinal epithelium is an example of tissue showing a highly restricted expression pattern related to the developmental stage. Strong TCF7L2 expression was present within the crypts of early human foetal small intestine (Barker *et al*, 1999; Barker *et al*, 2000), with the villi showing barely detectable TCF7L2 protein levels. Conversely, TCF7L2 expression levels increased in the more developed villi suggesting presence of differential TCF7L2 expression.

Electrophoretic mobility on SDS-PAGE indicated a MW 7 kDa higher than the predicted value. For example, the longest isoform was predicted to be 65 kDa in accordance to the value published for TCF7L2 Q9NQB0-8 idsoform 8 but migrated at 72 kDa in SDS-PAGE. This behaviour on SDS-PAGE appeared to be due to the large positive charge of this protein counteracting the negative swamping charge of the SDS and hence slowing migration. Studies have also suggested low electrophoretic mobility of some proteins due to disordered N- and C-terminal domains (Iakoucheva *et al*, 2001). Incomplete unfolding of proteins in SDS and poor binding of the detergent rather than to an unusual

shape of the SDS-protein complex can also be the cause of the abnormal protein migration (Matagne *et al*, 1991).

Subcellular localisation of TCF7L2 was not specifically examined in vertebrate studies but TCF is abundant in nuclear extracts from NLK activated cells (Ishitani *et al*, 2003). Consistent with findings from other studies, the shorter isoforms that migrated at 56 -58 kDa were also found to be expressed in the nucleus, however there is no literature to support localisation of TCF7L2 protein in the cytosol. The presence of the 72 kDa band in the cytosol is therefore contentious and subsequent work following completion of these experiments has shown that this cytosolic banding is a likely artefact. Hence the relative expression analysis of TCF7L2 in AML and normals is probably incorrect given the much higher levels of the "72 kDa isoform" detected in the cytosol of normal samples. This would suggest that the levels of TCF7L2 are in fact higher in AML than in normal samples which would be consistent with the transcript analysis.

In summary, this chapter identified combinations of splicing events in the N- and C-terminus region of TCF7L2 producing several variants and proteins which have been reported to display differential DNA binding specificity and regulatory potentials. The potential functional differences of these isoforms will be examined in the next chapter.

5 - Functional Roles of TCF7L2 Isoforms in Haematopoietic Cells

5.1 INTRODUCTION

Chapters 3 and 4 characterised the mRNA and protein expression of TCF7L2 in AML and normal haematopoietic cells. The identification of TCF7L2 isoform sequences responsible for the observed protein expression in AML (Section 4.4.3) has led to the investigation of its function as presented in this chapter.

Several reports in the literature support the role of TCF7L2 protein as a downstream nuclear effector of WNT signalling in association with β-catenin. TCF reporters, containing multiple TCF binding sites have shown activity specifically in response to β-catenin stabilisation (Korinek *et al*, 1997; DasGupta & Fuchs, 1999). This study employs a lentiviral β-catenin activated reporter (BAR) which stably integrates multiple TCF binding sites coupled to a GFP reporter (Biechele & Moon, 2008; Biechele *et al*, 2009).

The rationale for studying TCF reporter activity is supported by the constitutive TCF or LEF transactivation seen in carcinogenesis and in AML (Simon *et al*, 2005). WNT signalling was reported to be activated by AML fusion proteins (RUNX1-ETO, PML-RARα or PLZF-RARα) (Zheng *et al*, 2004) and also during blast crisis in BCR-ABL. Whether activation of WNT is important for all types of AML or only feature in AML with balanced translocations has not been established (Muller-Tidow *et al*, 2004). In addition, different isoforms of TCF7L2 have been shown to have different effects on TCF-dependent transcription depending on the differential promoter binding properties of the C-terminal sequence with different sets of target genes (Eastman & Grosschedl, 1999; Atcha *et al*, 2003). The functional relevance of the different TCF7L2 isoforms identified in AML will therefore be verified in this chapter.

There is compelling evidence that the TCF7L2 / β-catenin complex is an important factor that determines the balance between cell proliferation and differentiation of normal haematopoiesis (Kirstetter *et al*, 2006; Luis *et al*, 2012); however it is not currently known

how the expression of TCF7L2 isoforms will influence the development of normal human $CD34^+$ haematopoietic cells. These studies will promote the understanding of the role of TCF7L2 overexpression in AML.

5.2 AIMS

The isoforms of TCF7L2 identified in Chapter 4 were assessed for their functional effects in leukaemic and normal haematopoietic cells. To achieve this objective, this chapter aims:-

- To measure the effect of TCF7L2 knockdown in myeloid cell lines on TCF-dependent transcription using a TCF-reporter system.
- To examine effect of TCF7L2 knockdown on proliferative and survival capacity of myeloid cell lines.
- To establish the effect of TCF7L2 isoform overexpression (72 kDa, 58 kDa and 56 kDa) on TCF-dependent transcription in myeloid cell lines.
- To investigate the consequences of TCF7L2 72 kDa and 56 kDa isoform overexpression on the differentiation of normal human haematopoietic cells.

5.3 MATERIALS AND METHODS

5.3.1 Generation of TCF-reporter stable cell lines

To investigate the activity of TCF7L2 upon activation of WNT signalling, stable reporter lines using the BAR platform were generated in myeloid and epithelial cell lines. These reporter lines were created and supplied by Dr. Richard Darley.

In brief, K562, U937 and HeLa cells were engineered to express a TCF-venus GFP reporter which enabled analysis of TCF-dependent transcription. Cells were transduced with BAR lentiviral packaging plasmids (plasmid no. 200 or 186) (Table 5.1) supplied by Randy Moon (Biechele & Moon, 2008). The BAR reporter contains a concatemer of 12 TCF response elements to improve TCF binding that were inserted between the long terminal repeats (LTR) of a lentivirus plasmid (Figure 5.1).

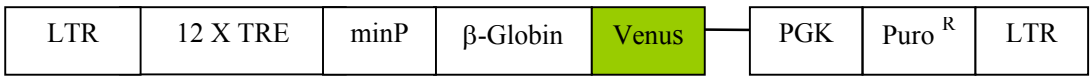
For the generation of stable reporter lines, the pBAR-VubiR (plasmid no. 200) or pBAR-VS (plasmid no. 186) lentiviral constructs were initially transfected into producer cells and the lentiviral particles were harvested (courtesy of Dr Richard Darley). Subsequently, parental K562 and U937 cells were infected with pBAR-VubiR lentivirus and HeLa cells were infected with pBAR-VS lentivirus in order to create stable reporter cell lines. Both BAR reporters (plasmid no. 200 and 186) drive the transcription of venus GFP to report the WNT signalling activation upon TCF binding. A second promoter known as Phosphoglycerate Kinase (PGK) drives the puromycin resistance gene (of vector pBAR-VS) to allow selection of the transduced HeLa cells and in the case of pBAR-VubiR, the ubiquitin promoter (instead of PGK) drives DsRed expression for selection of transduced K562 or U937 cells. The schematic diagrams of these promoters are shown in Figure 5.1.

Transduced reporter cell lines were selected based on DsRed positivity (K562 and U937) and puromycin resistance for HeLa cells. Following selection, a reporter clone of K562 and U937 cells was obtained by a limiting dilution method in a 96 well plate. Dr. Richard Darley kindly created and supplied these stable reporter BAR clones (refer Table 5.2).

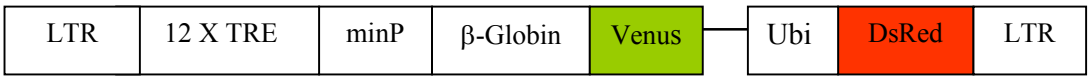
Name of plasmid vector	Plasmid Code No.	Coding sequence	Type of plasmid	Antibiotic resistant	Selectable marker	Source
pBAR-VubiR	200	Venus GFP	Lentiviral Reporter	Ampicillin	DsRed	(Biechele & Moon, 2008)
pBAR-VS	186	Venus GFP	Lentiviral Reporter	Ampicillin	Puromycin	(Biechele & Moon, 2008)
pHR	179	TCF7L2-72kDa with MYC tag	Mammalian expression vector	Kanamycin	Neomycin	Addgene, Cambridge, USA
pBSII-SK+	311	TCF7L2-56kDa	Bacterial expression vector	Ampicillin	*lacZ*	Eurofins MWG Operon, Ebersberg, Germany
pEX-A	209	TCF7L2-24kDa	Bacterial expression vector	Ampicillin	*lac*	Eurofins MWG Operon, Ebersberg, Germany
pLKO.1-puro	214	Non mammalian -shRNA	Lentiviral vector expressing shRNA	Ampicillin	Puromycin	Sigma-Aldrich®, Poole, UK
pLKO.1-puro	235 – 239	TCF7L2 -shRNAs	Lentiviral vector expressing shRNA	Ampicillin	Puromycin	Sigma-Aldrich®, Poole, UK
PINCO-DsRed	206 262 216	TCF7L2-72kDa TCF7L2-58kDa TCF7L2-56kDa	Retroviral vector expressing TCF7L2 insert cDNA	Ampicillin	DsRed	(Grignani *et al*, 1998)

Table 5.1: Details of plasmids and expression vectors used in this study. Plasmid numbers are those used in the laboratory plasmid catalogue.

(A)



(B)



Legend:-

LTR: Lentiviral long terminal repeat

12 X TRE: 12 TCF response elements

minP: Promega minimal promoter

β-Globin: β-Globin gene intronic sequences

Venus: GFP from eYFP variant gene

Ubi: Ubiquitin promoter

PGK: Phosphoglycerate Kinase promoter

DsRed: DsRed gene

Puro^{R}: Puromycin resistance gene

Figure 5.1: Schematic illustration showing components of the BAR reporter system used to create stable reporter lines. (A) pBAR-VS expression plasmid vector **(B)** pBAR-VubiR expression plasmid vector. Images were adapted from (Biechele & Moon, 2008).

Reporter cell line	Type	Source	Culture conditions
K562-pBAR-VUbiR	Subline of K562-myeloid leukaemia	Created by lentiviral transduction of K562 parental cells with BAR reporter plasmid no. 200. The DsRed positive clones were subsequently screened for BIO-induced activity. The clone with the highest fold induction was selected and expanded to produce this stable cell	RPMI-1640, 10% FBS, 20µg/ml Gentamicin, 2mM L-glutamine
U937-pBAR-VUbiR	Subline of U937-myeloid leukaemia	Created by lentiviral transduction of U937 parental cells with BAR reporter plasmid no. 200. The DsRed positive clones were subsequently screened for BIO-induced activity. The clone with the highest fold induction was selected and expanded to produce this stable cell	RPMI-1640, 10% FBS, 20µg/ml Gentamicin, 2mM L-glutamine
HeLa-pBAR-VS	Subline of HeLa-cervical carcinoma	Created by lentiviral transduction of HeLa parental cells with BAR reporter plasmid no. 186. Cells were selected in 0.5 µg/ml puromycin for 9 days. The puromycin resistant clones were subsequently screened for BIO-induced activity and expanded to produce this stable cell	DMEM, 10% FBS, 20µg/ml Gentamicin, 2mM L-glutamine

Table 5.2: List of stable TCF-reporter cell lines used in this study. Conditions for induction with the WNT agonist BIO are described in 5.3.4.

5.3.2 The effects of TCF7L2 knockdown in cell lines using lentiviral shRNA

5.3.2.1 Design of TCF7L2 shRNA plasmids

In order to specifically knock down the expression of the *TCF7L2* gene; MISSION® shRNA Lentiviral plasmid based RNA interference was used. This lentiviral system permits efficient and stable expression of shRNA using the pLKO.1-puro vector (Zufferey *et al*, 1998). The shRNA contains 21 sense bases that are identical to the TCF7L2 target gene, a loop and 21 antisense bases (see Figure 5.2).

A total of six shRNA clones were obtained from MISSION® shRNA, supplied in glycerol stocks. The pLKO.1-puro non-mammalian shRNA control (plasmid no. 214) supplied by the manufacturer contains a sequence that should not target any known human or mouse gene and serves as a non-targeted control for interpretation of knockdown results. The target sequences to be silenced within the *TCF7L2* gene were selected based on predicted sequences of shRNA clones accessible from The RNAi Consortium (TRC) database; http://www.broadinstitute.org/rnai/public/trans/candidates?transId=TRCT0000020432. The list of shRNA targets are shown in Table 5.3.

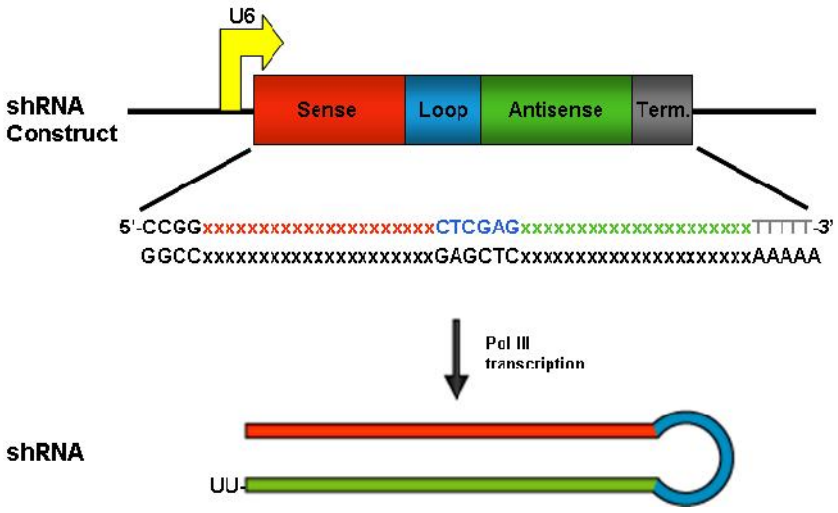


Figure 5.2: Schematic diagram of the shRNA insert cloned into the pLKO.1 puro vector. Hairpin sequences are comprised of a 21 base stem and a 6 base loop. The human U6 promoter (a pol III promoter) is used to drive expression of the shRNA hairpin. Expression using pol III promoter is optimal for producing shRNAs due to precise initiation and termination of transcription by the polyT termination sequence. This image was obtained from http://www.sigmaaldrich.com/life-science/functional-genomics-and-rnai/shrna/library-information/vector-map.html

Type of shRNA	Exon target	shRNA target sequence (21 bp)
TCF7L2-exon1 MISSION® shRNA	1	GCGCCAACGACGAACTGATTT
TCF7L2-exon2 MISSION® shRNA	2	CCGAAAGTTTCCGAGACAAAT
TCF7L2-exon11 MISSION® shRNA	11	TAGCTGAGTGCACGTTGAAAG
TCF7L2-exon13 MISSION® shRNA	13	CCTTTCACTTCCTCCGATTAC
TCF7L2-exon18 MISSION® shRNA	18	CGTCACCAAGTCTTTAGAATA
Non-target control MISSION® shRNA	Control	CAACAAGATGAAGAGCACCAA

Table 5.3: List of TCF7L2 shRNAs used in this study.

5.3.2.2 Production of TCF7L2 shRNA lentiviruses

The MISSION® shRNA bacterial glycerol stock (25-50µl) was streaked onto LB agar plates and plasmid DNA was purified from bacterial culture as described in sections 2.6.4. and 2.6.5.

Lentiviral particles of each TCF7L2 shRNA including the control were produced in packaging cells HEK293T by co-transfection with compatible packaging plasmids and harvested as described in sections 2.7.2 and 2.7.3.

5.3.2.3 TCF7L2 shRNA lentiviral transduction

The shRNA lentiviruses were used to transduce the stable reporter cell line K562-pBAR-VubiR according to section 2.7.4.

5.3.2.4 Puromycin selection of shRNA transduced cells

Following lentiviral transduction, cells expressing stable gene knockdown were selected using puromycin as follows. Cells (2×10^5) in fresh medium were transferred to a new TC treated 24-well plate and treated with 1 µg/ml puromycin and cultured at 37°C and 5% CO_2. Selection was carried out in the presence of puromycin until all uninfected control cells were killed (~1 week). Growth medium containing puromycin were replaced every 3 days during the period of selection.

5.3.2.5 Validation of shRNA knockdown by Western Blotting and qRT-PCR

Cell pellets were harvested from shRNA transduced K562-pBAR-VubiR cells at day 14 after transduction. Nuclear protein samples were extracted from 2×10^6 cells as described in 2.8.1. To complement protein validation, gene knockdown was also validated at the mRNA level. For this purpose, total RNA was prepared from 2×10^6 cells derived from shRNA transduced K562-pBAR-VubiR cells harvested at day 14 using Trizol as described in 2.5.1 – 2.5.2. Reverse transcription was carried out using 500 ng of RNA as described in 2.5.3. Expression of *TCF7L2* was assessed by amplifying two sets of primer pairs using the SYBR-Green protocol. The first primer pair set amplified the region between exons 7/8 and exon 9. The second set of primers amplified the exon 18 region of the *TCF7L2* gene. The primer design, primer sequences for *TCF7L2* and *ABL*, cycling parameters and protocol used for the qRT-PCR assay are described in 2.5.4. Raw data were imported into DataAssist™ v2.0 software (Applied Biosystems). The mRNA expression of *TCF7L2* level in the target shRNA sample relative to the pLKO.1-puro control sample was calculated by the comparative cycle threshold (Ct) method. Once the control sample was selected as a reference, the results were calculated by the DataAssist™ v2.0 software algorithm as below:-

$$\text{Relative Fold change} = 2^{(-\Delta Ct\ \text{target})} / 2^{(-\Delta Ct\ \text{control})}$$

Average Ct = mean of the technical replicates

ΔCt= Average Ct – Normalisation Factor

5.3.2.6 Cell growth and viability assessment

To determine whether TCF7L2 knockdown affects cell growth, cells were counted using a haemocytometer under a microscope (see 2.3.3).

To assay cell viability and apoptosis, 7-Amino-actinomycin D (7-AAD) was used. The 7-AAD dye intercalates cytosine and guanine bases of DNA; apoptotic and non-viable cells can be characterised and identified (Philpott *et al*, 1996). Viable cells retain their membranous integrity and are impermeable to 7-AAD (Schmid *et al*, 1992). Briefly, 1 x 10^5 cells were washed with 1 ml of 1 X PBS by centrifugation at 180 x g for 5 minutes. Cells were incubated with 1 μg/ml of 7-AAD for 20 min protected from direct light and analysed by flow cytometry as described in 2.9.1.

5.3.3 The effects of TCF7L2 overexpression in cell lines using a retroviral system

5.3.3.1 Creation of TCF7L2 clones and expression plasmids

Standard cloning methods including restriction enzymes used to construct TCF7L2 coding PINCO-DsRed plasmids (no. 206, no. 262 and no. 216 (see Table 5.1) are described in 2.6.1.

The TCF7L2-58kDa and TCF7L2-56kDa constructs contain the 5' Kozak sequence (GCCACC) at the start of the ORF. The sequence encoding TCF7L2-56kDa isoform was generated by creating a *Bam*HI/*Eco*R1 fragment to 'drop' the 56kDa insert from the source pBSII-SK+ plasmid (no. 311) (Eurofins MWG Operon, Ebersberg, Germany). The PINCO-DsRed plasmid vector was also digested using a similar RE for directional cloning. The TCF7L2-56kDa insert (1.7 kb) was purified from agarose gel (2.6.2) and sub cloned into the *Bam*H1/*Eco*R1 sites of PINCO-DsRed yielding the expression plasmid no. 216.

The sequence encoding the TCF7L2-58kDa isoform was generated by introducing different alternative C-terminal "tail sequences" (containing exon 14) into the existing TCF7L2-56kDa isoform expression plasmid vector no. 216 created as above. The tail sequence was derived from expression plasmid no. 209 (Eurofins MWG Operon, Ebersberg, Germany) containing exon 13, 14 and 18. This tail region was excised using *Xho*1/*Eco*R1 digestion and the predicted 200 bp product was gel purified and was ligated into plasmid TCF7L2-56kDa isoform (plasmid no. 216). This modified plasmid then became the TCF7L2-58kDa isoform (plasmid no. 262).

The sequence encoding TCF7L2-72kDa was generated by creating a *Bgl*II/*Xba*1 insert fragment of 72kDa sequence from the pHR-TCF7L2 plasmid no. 179 (Addgene, Cambridge, USA). Klenow was used to blunt the *Xba*1 end site. The resulting 2.4 kb insert sequence contains a 5' MYC tag within the ORF. The insert fragment was purified and subcloned into the *Bam*H1/ blunt end sites of the PINCO-DsRed expression plasmid vector to create expression plasmid no. 206. DNA sequence verification of each construct DNA was checked by direct sequencing using PINCO vector primer pairs around the cloning sites.

5.3.3.2 TCF7L2 retroviral transduction and western blotting

The amphotropic packaging cell line Phoenix was transfected with 45 μg of each TCF7L2 plasmid DNA (as created in 5.3.3.1) using calcium phosphate as described in section 2.7.1. Retroviral constructs were packaged and used to infect myeloid cell lines and HeLa cells as previously described (2.7.4 - 2.7.3).

The HeLa-pBAR-VS reporter line was seeded at approximately 30% confluency prior to infection. In order to infect the adherent HeLa cells, 1ml of retrovirus (PINCO-DsRed empty vector (EV), 72kDa and 56kDa TCF7L2 virus) was added directly to each culture in a 5mL flask and incubated overnight at 37°C with 5% CO_2. On the following day, virus containing medium was replaced with fresh medium.

In order to verify the success of retroviral transduction at the protein level, western blotting was performed to assay TCF7L2 isoform expression using protocols described in section 2.8. To isolate discrete subpopulation of transduced cells from the K562-pBAR-VubiR reporter line, DsRed expressing cells were enriched by fluorescent activated cell sorting. Specifically, 5 x 10^6 cells were aliquoted and centrifuged at 300 x g for 5 minutes. The growth medium was discarded and the cell pellet was resuspended in 1X PBS buffer containing 1% BSA and passed through a 40 μM cell strainer. Cells were then flow sorted with the help of Dr Chris Pepper, Cardiff University, using FACSDiva™ version 6.1.3 (BD Biosciences, Oxford, UK). DsRed positive cells (2 x 10^5) were recovered post cell sorting; each condition with more than 70% DsRed positivity. Sorted cells were expanded and harvested for TCF7L2 protein expression by western blotting.

5.3.4 Assay for TCF reporter activity

To observe the effect of TCF7L2 on WNT signalling, cell lines (K562-pBAR-VubiR and HeLa-pBAR-VS) overexpressing TCF7L2 isoforms and shRNA transduced cell line (K562-pBAR-VubiR) were assayed for TCF reporter activity. The assay was performed on days 5 and 7 post transduction. Cell lines expressing the BAR reporter system that co-expressed the PINCO-DsRed EV construct or non-targeted shRNA served as controls for TCF reporter activity. All cells were treated for 24 hours with 2.5 μM of the GSK3β inhibitor BIO (2'Z,3'E)-6-Bromoindirubin-3'-oxime) (Tocris, Bristol, UK) prior to analysis. BIO is a potent ATP-competitive GSK3β inhibitor that stabilizes β-catenin levels by preventing its degradation. At the same time, a separate set of cells were treated with 1 μg/ml recombinant WNT3A (R&D Systems, Abingdon, UK) using similar conditions. WNT3A is a physiological agonist of WNT ligands that activates β-catenin via WNT signalling. In each case, cells were seeded at a density of 2 x 10^5 cells/ml in a 48 well plate and cultured in appropriate growth medium (see Table 5.2) and incubated for 24 hours at 37°C with 5% CO_2. The following day, 250 μl of culture was washed with 1X PBS buffer containing 1% BSA, and centrifuged at 300 x g for 5 minutes. Cell pellets were resuspended in 100 μl staining buffer and analysed for reporter expression using the Accuri® C6 flow cytometer (Accuri Cytometers, Ann Arbor, USA). Histogram data analysis was performed as described in 2.9.2.1.

5.3.5 The effects of TCF7L2 overexpression in normal haematopoietic cell development using a retroviral system

5.3.5.1 Transduction of normal $CD34^+$ HPC with TCF7L2 expression plasmid

$CD34^+$ cells were purified from mononuclear cells from neonatal CB using MiniMACS (Miltenyi Biotec, Camberley, UK) and infected with retrovirus encoding for TCF7L2 (5.3.3.1) as described in 2.4.3 and 2.7.4. Transduced $CD34^+$ HPC were seeded at 0.5 x 10^5 /ml in supplemented IMDM containing the following growth factors; 50 ng/ml IL-3, SCF, FLT-3 and 25 ng/ml IL-6, G-CSF and GM-CSF at 37°C with 5% CO_2. After three days of culture, transduced $CD34^+$ HPC in logarithmic growth phase were seeded at 2 x 10^5 /ml and sub-cultured in supplemented IMDM containing 5 ng/ml IL-3, G-CSF, GM-CSF, and 20 ng/ml SCF. The optimal concentration of growth factors used was adapted from previous experiments (Darley *et al*, 1997; Darley *et al*, 2002).

5.3.5.2 Immunophenotypic analysis of transduced cells

To assess cell differentiation, immunostaining was carried out using lineage discriminator antibodies (CD13 and CD36) as well as specific differentiation antibodies (CD14, CD15, CD34) as listed in Table 5.4. Immunophenotyping was performed in a 96-V well plate using a minimum of 5 x 10^4 cells taken from each culture. Appropriate cell numbers were aliquoted into each well and centrifuged at 300 x g for 3 minutes and growth medium was discarded. Cells were then washed with 150 µl flow cytometry staining buffer. Following centrifugation as above, the supernatant was discarded and cell pellets were resuspended in 15 µl of staining buffer. Between 2.5 µl – 5 µl of primary antibodies were added (according to manufacturer's recommendations). The plate was securely covered with a plate sealer and briefly vortexed prior to incubation at 4°C for 30 minutes. Following incubation, cells were washed with 150 µl staining buffer as above. Supernatant was aspirated and replaced with 15 µl of staining buffer. The diluted secondary antibody streptavidin (SA) - peridinin chlorophyll protein (PerCP) Cy5.5 (diluted 1:20 in flow cytometry staining buffer) (BD Pharmingen™, Oxford, UK) was added to each well (5µl), vortexed and incubated at 4°C for 30 minutes. Finally, cells were washed with 150 µl staining buffer, centrifuged as above for 3 minutes and resuspended with 100 µl staining buffer for analysis using an Accuri® C6 flow cytometer. Data analysis was performed using FCS Express as described in 2.9.2.2.

Antibody reagent	Fluorescent conjugate	Ig type	Manufacturer	Target cell receptor
CD13	APC	IgG	BioLegend, London, UK	Myeloid lineage
CD36	Biotin	IgM	Ancell, Bayport, UK	Myeloid/erythroid lineage
CD34	FITC	IgG	BD Pharmingen™, Oxford, UK	Progenitor cells
CD14	FITC	IgG	DAKO, Ely, UK	Monocytes
CD15	FITC	IgG	BioLegend, London, UK	Granulocytes
IgG	FITC	1gG1	DAKO, Ely, UK	Isotype control (non-specific binding)

Table 5.4: List of antibodies used for immunophenotyping of transduced haematopoietic cells. Biotinylated CD36 was subsequently labelled with SA-PerCp Cy5.5.

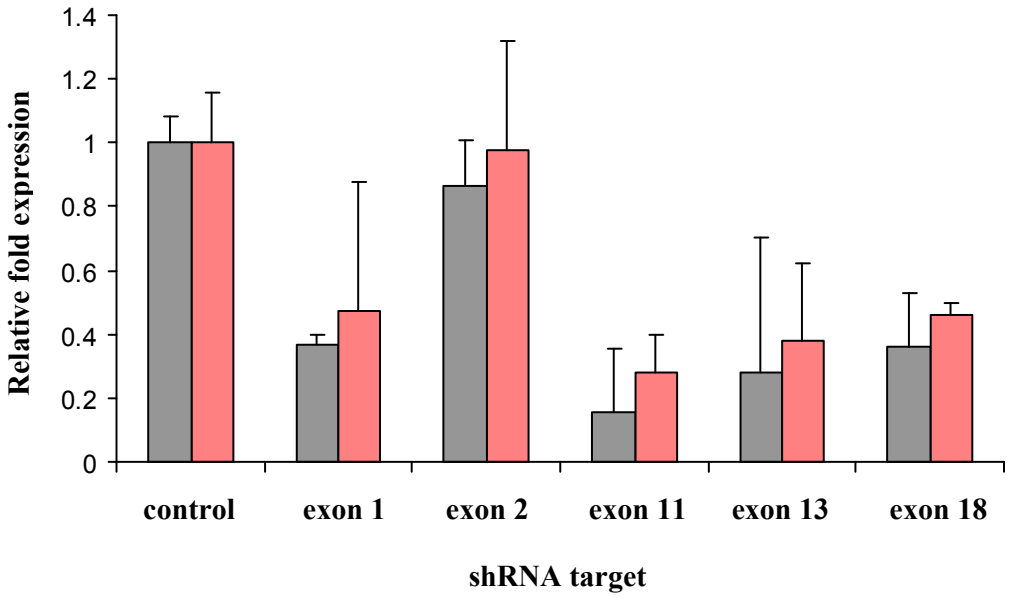
5.4 RESULTS

5.4.1 TCF7L2 knockdown in myeloid leukaemia cell lines

To evaluate the physiological relevance of TCF7L2 overexpression in leukaemia cells, five shRNA clones were assayed for suppression of endogenous TCF7L2 expression in the K562 cell line. Each shRNA lentiviral construct (Table 5.3) was transduced into the K562 pBARVubiR reporter line and stable gene knockdowns were achieved by puromycin selection (see 5.3.2.1- 5.3.2.4). The effects of gene knockdown was assessed by qRT-PCR and western blotting as described in 5.3.2.5.

As shown in Figure 5.3, all but one of the shRNA constructs were able to bring about *TCF7L2* mRNA gene knockdown by more than 50% compared to control. Gene knockdown was not achieved by targeting exon 2, as validated by two sets of primers. Downregulation of *TCF7L2* mRNA expression by shRNAs targeting exon 1, 11, 13 and 18 also led to reduced protein levels as seen in the nuclear extracts (Figure 5.4). Consistent with the qRT-PCR data, inhibition was not achieved at the protein level by shRNA targeting exon 2. Therefore the shRNA clone targeting exon 2 of TCF7L2 was omitted from further analysis.

These results provide evidence that stable production of shRNA in K562-pBAR-VubiR cell allows efficient knockdown for most exons targeted.



Legend:-

RQ primer exon 7/9

RQ primer exon 18

Figure 5.3: Validation of shRNA knockdown by quantitation of *TCF7L2* mRNA expression detected by qRT-PCR. Fold change plot showing mRNA expression levels of *TCF7L2* in K562-pBAR-VubiR cells transduced with shRNA clones targeting TCF7L2 at day 14 post transduction. *TCF7L2* mRNA was amplified using exon 7/9 and exon 18 primer pairs. The fold expression level shown for each shRNA target is relative to the control. Data represent the mean ± SD from duplicate samples.

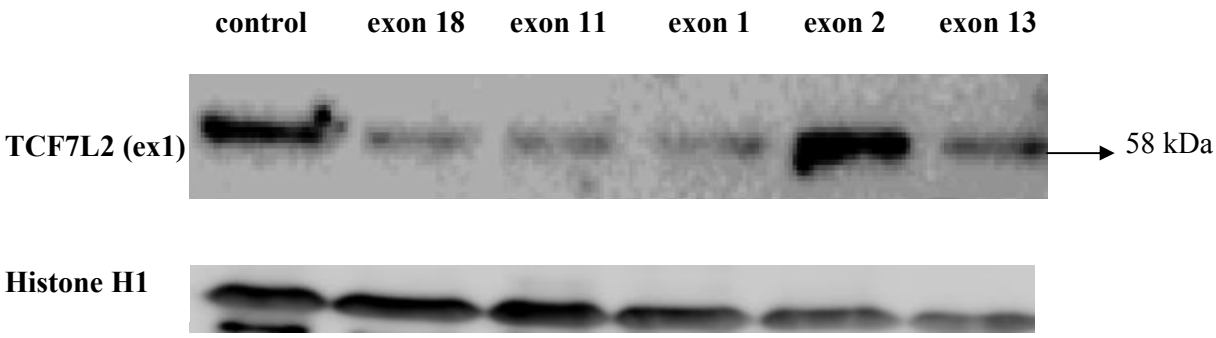


Figure 5.4: Validation of shRNA knockdown by protein expression as detected by western blotting. Blot showing nuclear TFC7L2 protein expression levels of shRNA transduced K562-pBAR-VubiR at day 14 post transduction. Nuclear protein extracts were prepared from indicated samples and the blot was probed with monoclonal TCF7L2 antibody (exon 1 epitope). Loading control used was Histone H1.

5.4.2 TCF7L2 knockdown represses TCF-dependent reporter activity in myeloid leukaemia cells

To investigate the effect of TCF7L2 knockdown on WNT signalling, TCF-dependent transcriptional activity using BAR reporter constructs was measured. As above, the K562-pBAR-VubiR cells expressing BAR reporter were transduced with TCF7L2 shRNA constructs (see 5.4.1) and stimulated with either BIO or WNT3A agonists (agents which stabilise the levels of β-catenin and thereby promote TCF-dependent-transcription). The differences between GFP signals were measured by histogram subtraction as described in section 2.9.2.

As shown in Figure 5.5 (and summarised in Figure 5.6) BIO induced TCF-dependent transcriptional activity was suppressed by up to 60% relative to control activity. Unexpectedly, shRNA targeting exon 18 exhibited little effect on TCF transcriptional activity.

The effect of the natural WNT agonist, WNT3A was next examined. These findings suggested that all shRNA clones suppressed the baseline reporter activity in this cell, yielding more than 60% transcriptional reduction relative to the control (see Figure 5.7). In contrast to the effect seen in BIO stimulation, exon 18 exhibited strongest inhibition on reporter activity after WNT3A stimulation.

In summary, transcription induced by β-catenin stabilisation was inhibited by the loss of TCF7L2 gene expression in leukaemic cells, suggesting that canonical WNT signalling is at least partially dependent on TCF7L2 expression in these cells.

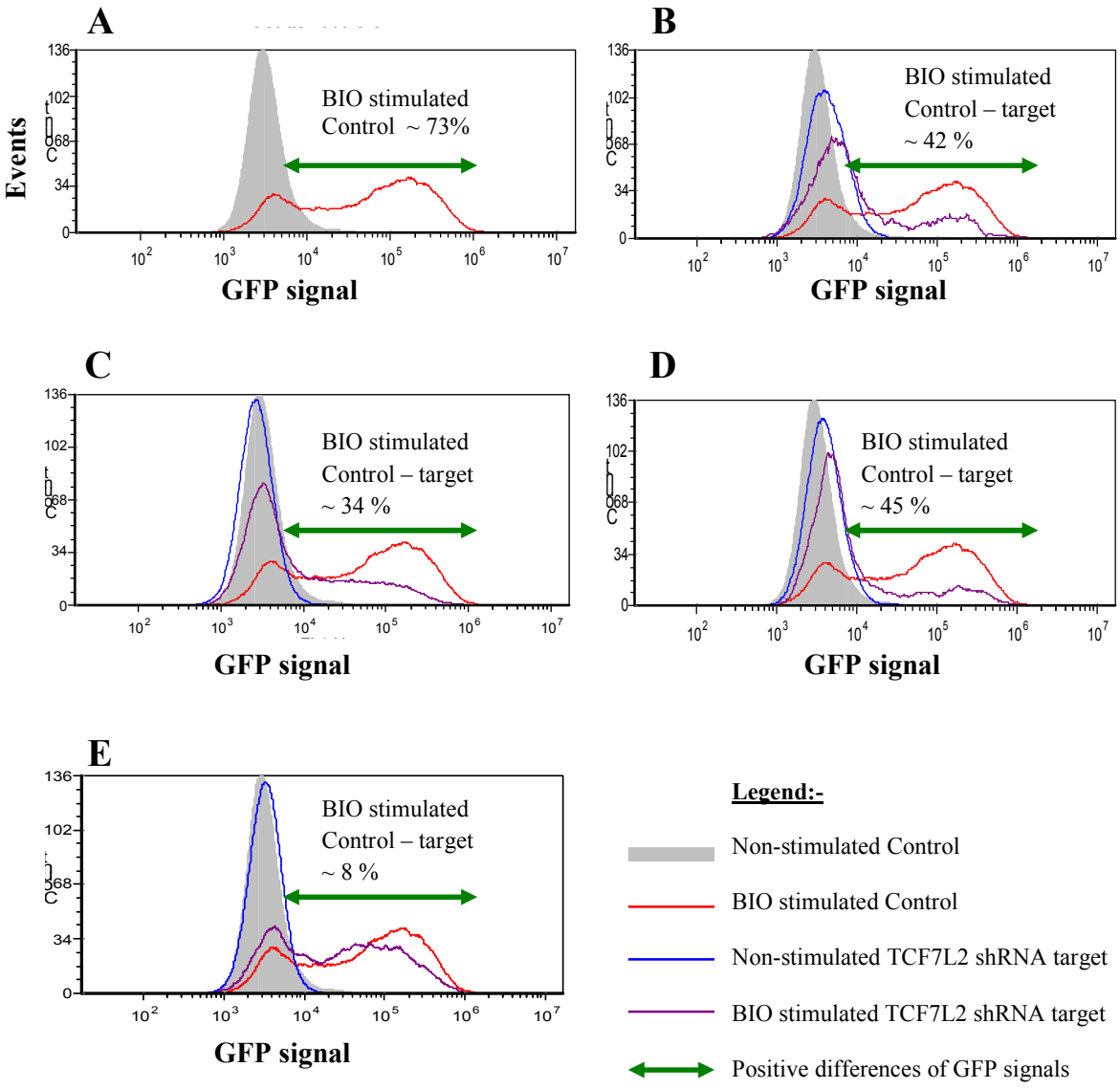


Figure 5.5: Histogram analysis of shRNA transduced K562-pBAR-VubiR reporter cell lines in response to BIO treatment. Cells were seeded on a 48-well plate treated with 2.5 µM BIO for 24 hours. The positive difference of GFP fluorescence signal was calculated for each construct by subtracting the fluorescence of non-stimulated cells from that of BIO stimulated cells. Histograms overlay and values for positive differences are shown for **(A)** Control only **(B)** Control *vs* shRNA exon 1 **(C)** Control *vs* shRNA exon 11 **(D)** Control *vs* shRNA exon 13 **(E)** Control *vs* shRNA exon 18. Data derived from cells harvested at day 10 post transduction.

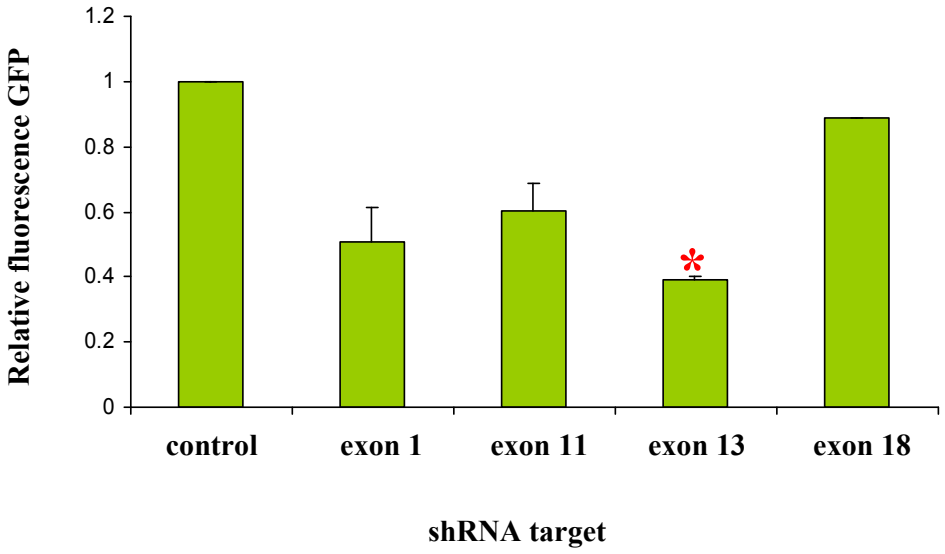


Figure 5.6: TCF-reporter activity of shRNA transduced K562-pBAR-VubiR reporter cell lines in response to BIO treatment. Transduced cells were seeded on a 48-well plate treated with 2.5 μM BIO for 24 hours. The positive difference of GFP fluorescence signal was calculated for each construct by subtracting non-stimulated cells from BIO stimulated cells (see 5.3.4). The fold GFP fluorescence (positive difference) shown for each shRNA target is relative to the Control. Data represent the mean ± SD from two independent experiments at day 10. Statistical significance was calculated using Student's T-test. * $P <$ 0.05.

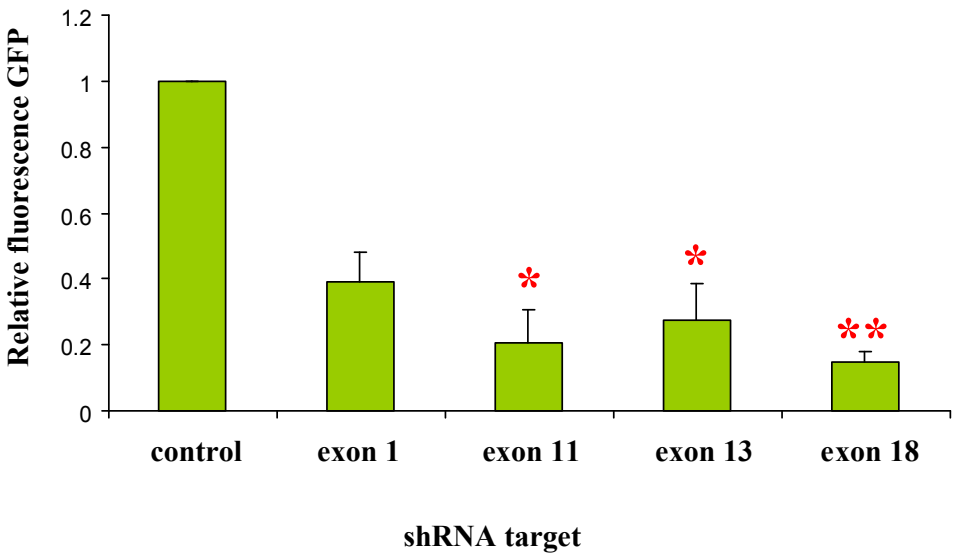


Figure 5.7: TCF-reporter activity of shRNA transduced K562-pBAR-VubiR reporter cell lines in response to WNT3A treatment. Transduced cells were seeded in a 48-well plate treated with 1 µg/ml WNT3A for 24 hours. The positive difference of GFP fluorescence signal was calculated for each construct by subtracting non-stimulated cells from WNT3A stimulated cells (see 5.3.4). The fold GFP fluorescence (positive difference) shown for each shRNA target is relative to the control. Data represent the mean ± SD from two independent experiments at day 10. Statistical significance was calculated using Student's T-test. * $P < 0.05$, ** $P < 0.01$.

5.4.3 TCF7L2 knockdown inhibits growth and survival of myeloid leukaemia cell lines

Previous reports have shown that mice deficient in TCF7L2 develop normally, but die shortly after birth due to the absence of cycling epithelial progenitor cells of the small intestine (Korinek *et al*, 1998a). Similarly, TCF7L2 $^{-/-}$ mutant zebrafish also display a defect of proliferation in the intestine leading to death (Muncan *et al*, 2007). This section therefore investigated whether depletion of TCF7L2 would impair growth and survival of leukaemic cells. Viability was assessed by flow cytometry using 7-AAD (method in 5.3.2.6) and cell counts were assessed as in 2.3.3.

As shown in Figure 5.8, TCF7L2 knockdown by shRNA exon 1 and 13 induced a stronger inhibition of cell growth as compared to shRNA exon 11 and 18 knockdown in K562-pBAR-VubiR cells. Impaired growth was also prominent for U937-pBAR-VubiR (Supplementary 5.1).

To examine whether TCF7L2 shRNA-mediated growth inhibition is associated with apoptosis, cells were stained with 7-AAD. TCF7L2 shRNA reduced the number of living cells as compared to control shRNA. Consistent with the growth data, TCF7L2 shRNA targeting exon 1 and 13 were more effective at inducing apoptosis as compared to exon 11 and 18 (see Figure 5.9).

In summary, TCF7L2 knockdown was found to inhibit proliferation and promote apoptosis indicating the involvement of TCF7L2 in maintaining the survival of myeloid leukaemia cells.

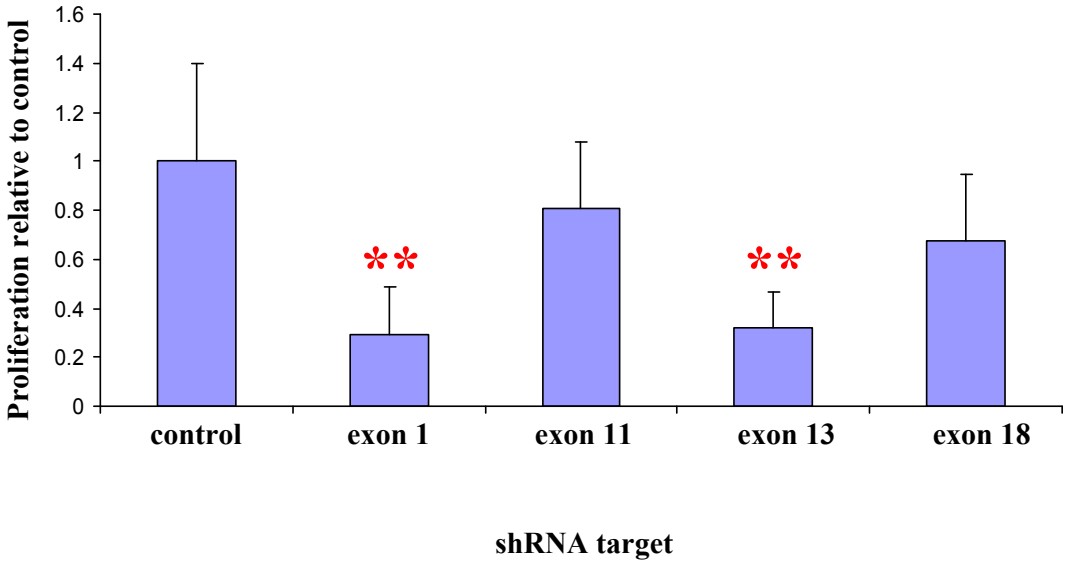


Figure 5.8: Proliferative activity of shRNA transduced K562-pBAR-VubiR reporter cell lines. Proliferation was counted as cell number /ml (x10^5). Relative proliferation shown for each shRNA target is relative to the control. Data represent the mean ± SD from seven time points. Statistical significance was calculated using Student's T-test. ***P* < 0.01

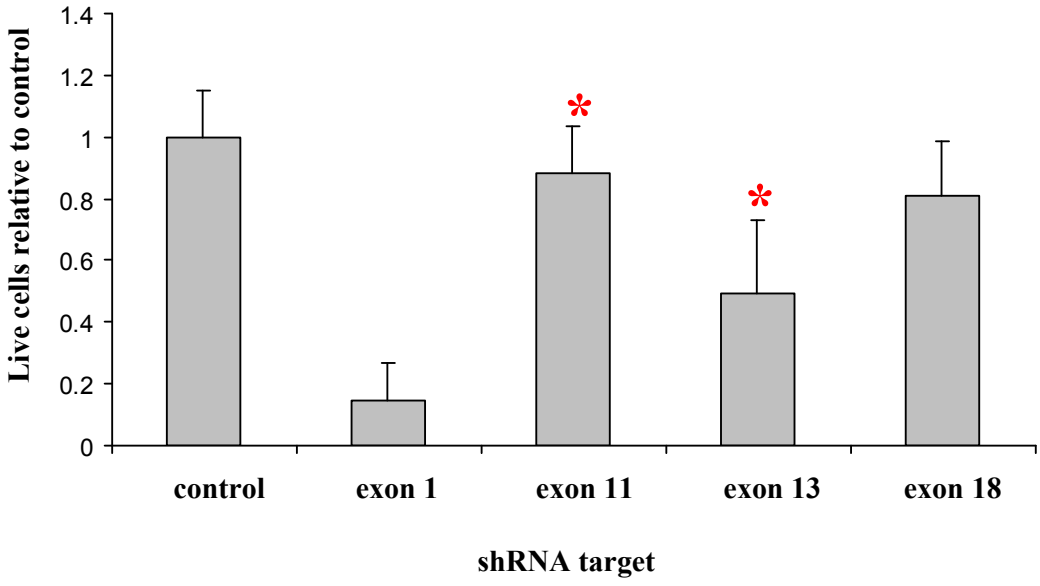


Figure 5.9: Effect of TFC7L2 shRNA on cell viability. The gated live cells shown for each shRNA transduced K562-pBAR-VubiR cell line is shown relative to control. Data represent the mean ± SD from three time points. Statistical significance was calculated using the Student's T-test. * $P < 0.05$, ** $P < 0.01$.

5.4.4 Over-expression of TCF7L2 isoforms repress TCF-reporter activity in myeloid cells

Chapter 4 (4.4.3) identified the protein isoforms that are expressed in AML patients. To explore the role of TCF7L2 over-expression, the sequences of the three TCF7L2 isoforms (78kD, 58kD and 56kD) were generated and retroviral expression constructs were created by subcloning into the PINCO-DsRed vector (detailed in section 5.3.3). Transduced K562-pBAR-VubiR and HeLa-pBAR-VS cells were derived and exogeneous protein expression was determined as described in 5.3.3.2. Western blotting data (Figure 5.10) shows that the TCF7L2-72kDa construct produced a protein with much higher MW of approximately 78kDa, probably due to the in-frame MYC tag coding sequence. The TCF7L2-58kDa construct produced the predicted overexpression at 58kDa compared to the parental control. The shortest construct, TCF7L2-56kDa, generated overexpression at 56kDa MW but also at the 72kDa and 58kDa MW levels, which could be due to stabilising effects on the expression of the other isoforms. This protein expression data shows that it is possible to stably overexpress TCF7L2 protein isoforms in K562-pBAR-VubiR cells.

In the previous section (as described in 5.4.2), TCF7L2 knockdown was shown to cause repression of TCF transcriptional reporter activity. To establish the effect on transcriptional activity of overexpression of TCF7L2, K562-pBAR-VubiR and HeLa-pBAR-VS cells were transfected with retroviral expression constructs encoding three TCF7L2 isoforms; 72kDa, 58kDa and 56kDa and EV as control. Transduced cells were then treated with the WNT agonists, BIO and WNT3A (described in 5.3.4).

No change in constitutive reporter activity was seen for any of the three isoforms analysed (data not shown). Surprisingly overexpression suppressed the response to WNT agonists for TCF7L2 72kDa, 58kDa and 56kDa in myeloid cells (Figure 5.11A-B). To examine if this response might be restricted to the context of myeloid cells, an epithelial cell reporter line, HeLa-pBAR-VS, was transduced with TCF7L2-72kDa and 56kDa constructs and assayed for reporter activity similarly to K562-pBARV-UbiR. Again, in this case, no significant changes of relative reporter activity were seen between the control and TFC7L2 transduced HeLa reporter cells (see Figure 5.11C). Though the effect on reporter activity in myeloid cells was similar to that observed with TFC7L2 knockdown, in contrast to the effect of knockdown, no significant difference was observed in terms of cell growth and survival due to TCF7L2 overexpression (data not shown). In conclusion overexpression of

all TCF7L2 isoforms tested suppressed WNT agonist responses in myeloid cells but not in epithelial cells.

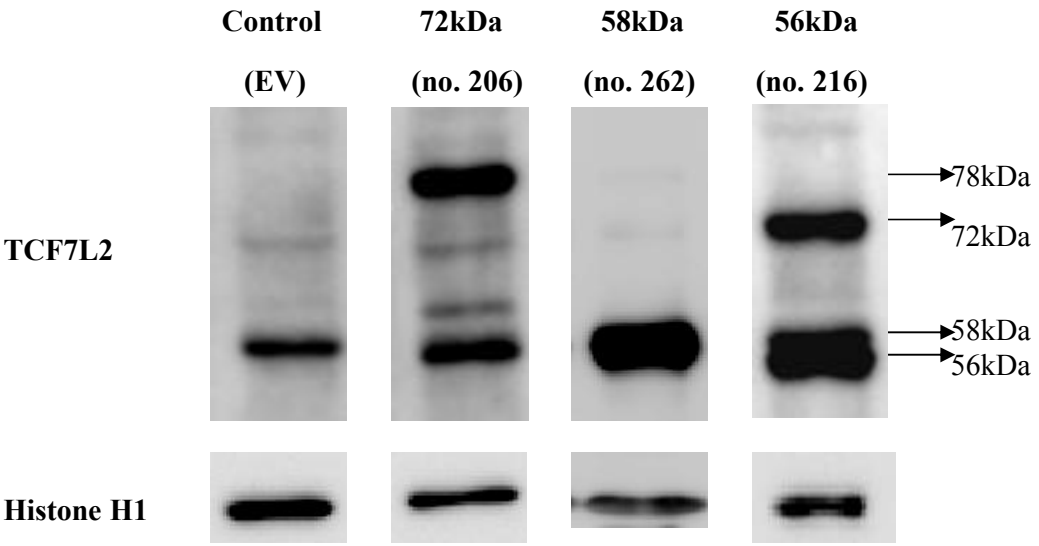


Figure 5.10: Validation of retroviral overexpression of TCF7L2 protein as detected by western blotting. Blot shows nuclear TCF7L2 expression levels in retrovirally transduced K562-pBAR-VubiR at day 28 post transduction. The blot was probed with monoclonal TCF7L2 antibody (exon11 epitope). The loading control used was Histone H1. EV (empty vector): PINCO-DsRed vector backbone plasmid, no.206: PINCO-DsRed TCF7L2-72kDa plasmid, no.262: PINCO-DsRed TCF7L2-58kDa, no.216: PINCO-DsRed TCF7L2-56kDa.

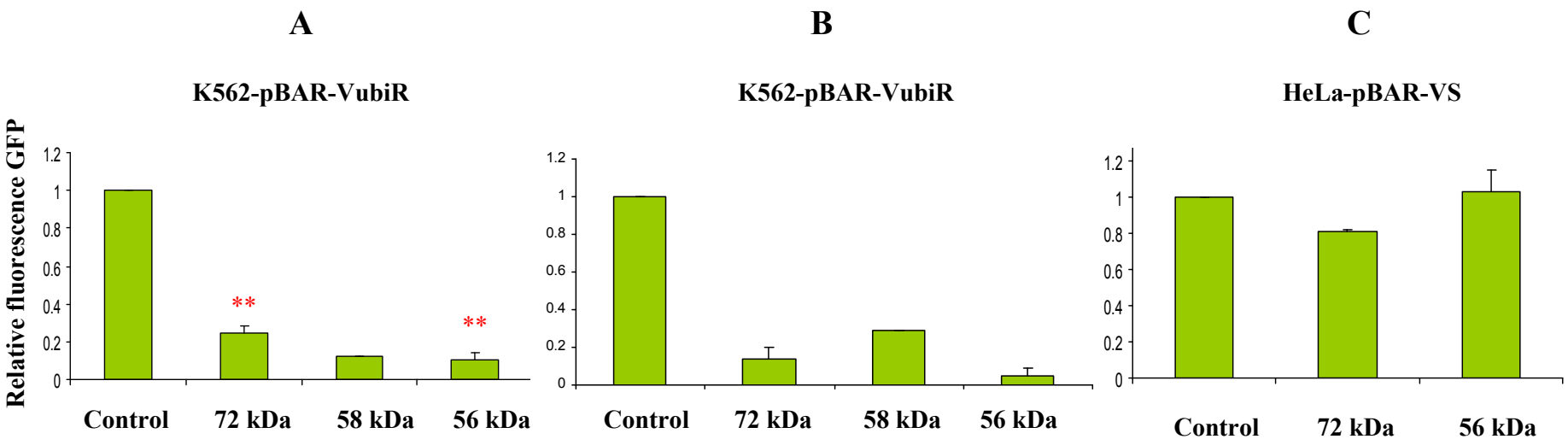


Figure 5.11: Effect of TCF7L2 overexpression on TCF-reporter activity in response to BIO and WNT3A treatment. TCF7L2-72kDa, 58kDa and 56kDa isoforms were overexpressed in stable TCF reporter cell lines. **(A)** K562-pBAR-VubiR treated with 2.5 µM BIO for 24 hours **(B)** K562-pBAR-VubiR treated with 1 µg/ml WNT3A for 24 hours **(C)** HeLa-pBAR-VS cells treated with 2.5µM BIO for 24 hours. Positive differences of GFP fluorescence were calculated for each construct by subtracting non-stimulated cells from BIO-stimulated cells (see 5.3.4). The fold GFP fluorescence (positive difference) for each TCF7L2 isoform overexpressing cell is relative to EV Control. Data represent the mean ± SD from two time points. Statistical significance was calculated using Student's T-test. ** $P < 0.01$.

5.4.5 Overexpression of TCF7L2 induces differentiation of monocytic lineage in normal haematopoietic cells

To assess whether TCF7L2 overexpression affects normal haematopoietic development, normal human $CD34^+$ HPC were transduced with 72kDa and 56kDa TCF7L2 isoforms and their subsequent development was assessed by immunophenotypic assay.

Normal $CD34^+$ HPC were purified from neonatal CB and transduced with the retroviral expression vectors as detailed in 5.3.5.1. Verification of transduction efficiency was determined by flow cytometric detection of the DsRed marker starting on day 3 until day 20 of culture. Transduction frequencies at day 6 are shown in Figure 5.12. Though the transduction frequencies of the TCF7L2 overexpressing cultures were relatively low compared to that achieved with EV, it was sufficient to carry out immunophenotypic analysis and was consistent with the relative titre of the virus used for transduction.

Following infection (from day 3), the three sub-populations of lineage committed haematopoietic cells (monocytes, granulocytes and erythrocytes) were identified using the lineage discriminators, CD13 and CD36. By day 9, it was evident by 4-colour flow cytometric analysis that the TCF7L2 transduced cultures consisted predominantly of monocytes of which > 50% were of monocytic phenotype ($CD13^+CD36^+$) as shown by dot plot (Figure 5.13) and lineage frequency (Figure 5.14). The fluorescence intensity of lineage marker expression shows not only that there is a higher frequency of monocytic cells but there are also higher levels of CD14 implying accelerated monocytic development (Figure 5.15); whereas the frequency of granulocytic lineage cells was reduced as was the level of differentiation marker expression (CD15) implying retarded development (Figure 5.16).

Taken together, these results demonstrate that overexpression of TCF7L2 isoforms via WNT signalling is sufficient to promote monocytic differentiation.

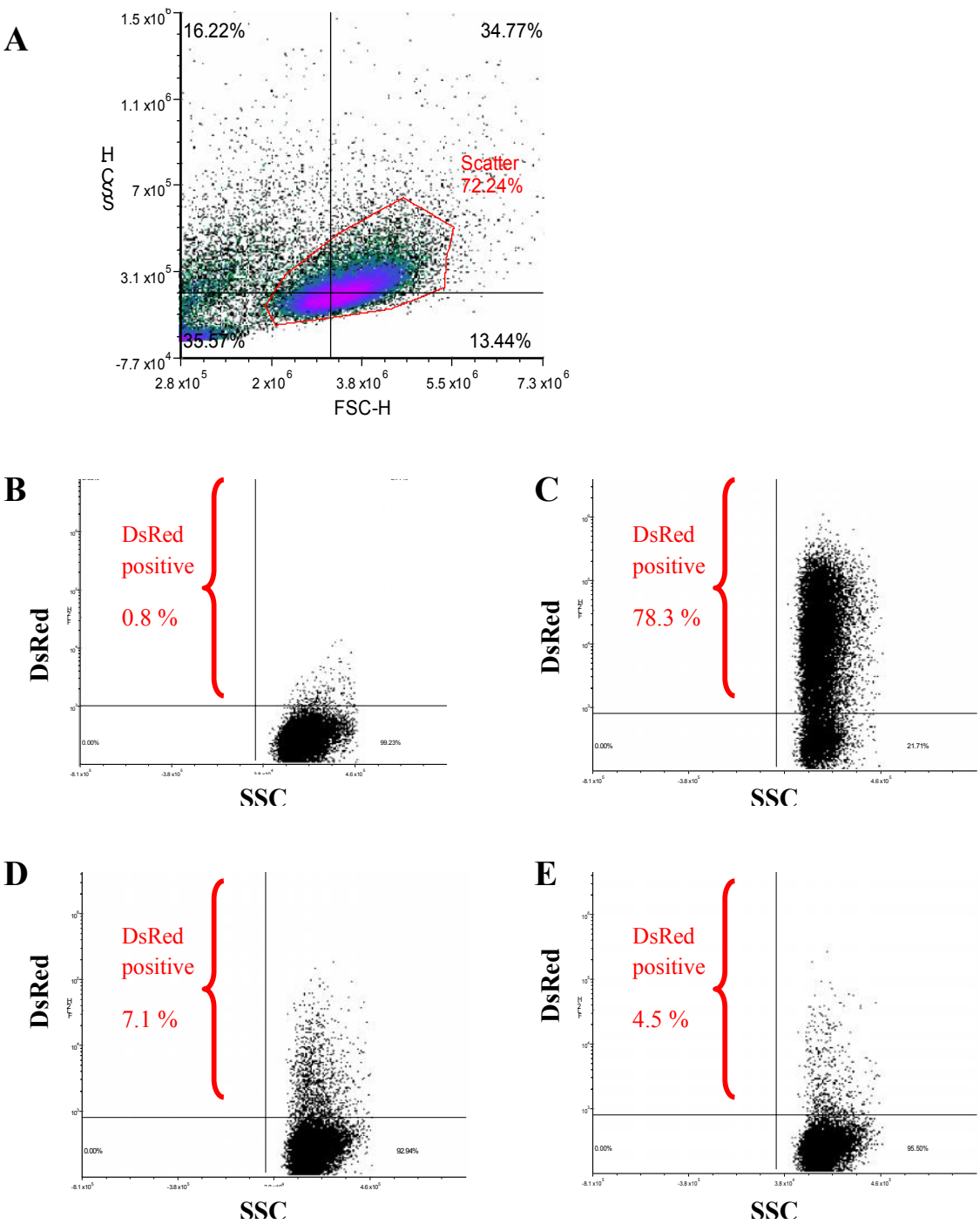


Figure 5.12: Bivariate dot plots showing typical DsRed positivity of TCF7L2 transduced CD34$^+$ cells (A) A scatter gate has been drawn to exclude debris high forward scatter events (cell clumps). This scatter region was then used to gate cells based on DsRed fluorescence **(B)** The gate has been set to define the DsRed positive cells. Similar gating was applied to determine DsRed positive regions for **(C)** EV Control **(D)** TCF7L2-56kDa transduced CD34$^+$ **(E)** TCF7L2-72kDa transduced CD34$^+$. The data represent cultures harvested on day 6.

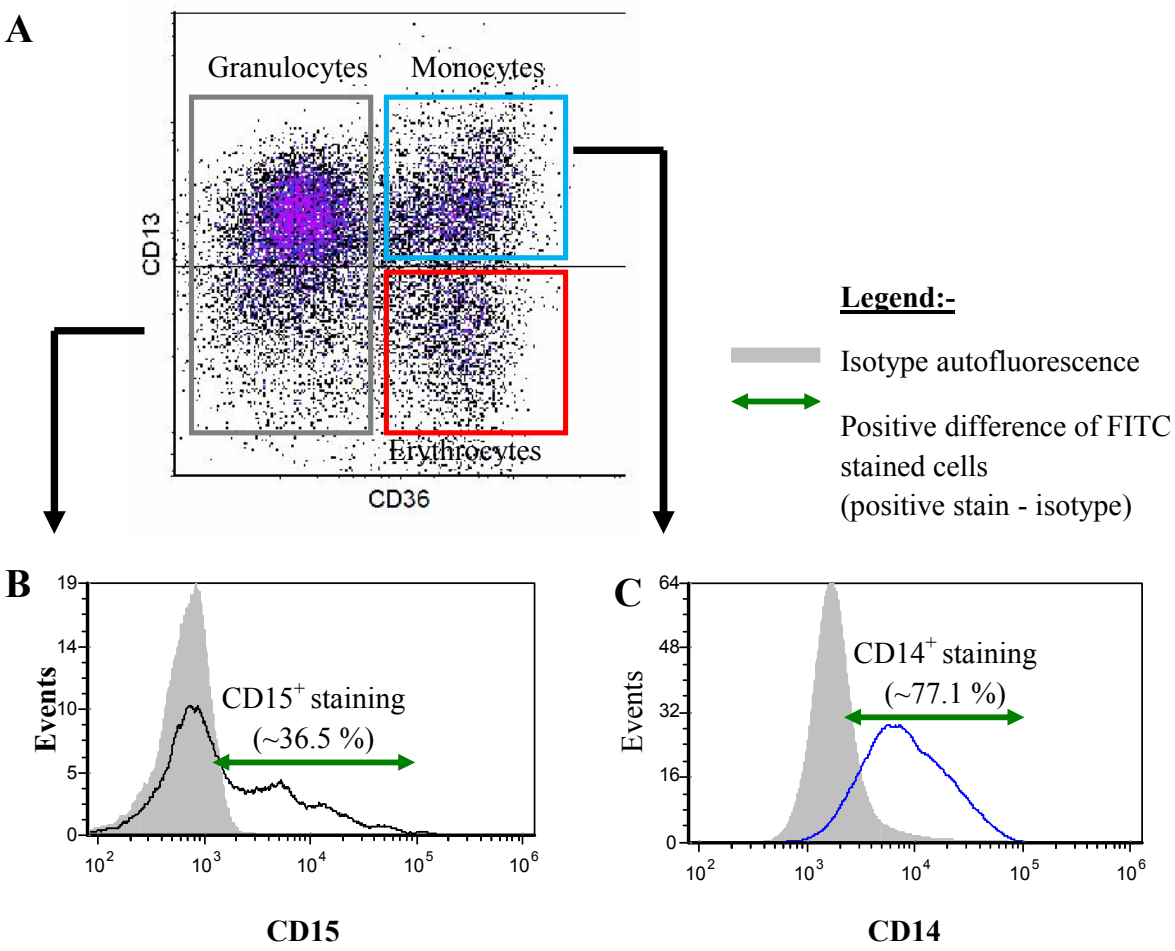


Figure 5.13: Representative dot plot and histograms showing immunophenotypic analysis of lineage differentiation of TCF7L2 transduced CD34$^+$ cells. (A) CD13 and CD36 discriminate monocyte, granulocyte and erythroid lineages within DsRed positive gated cells. Each subpopulation has been further examined for cell surface marker expression. **(B)** Example data for the granulocyte subpopulation showing expression the granulocytic marker, CD15, and **(C)** for the monocytic marker using CD14 on the monocytic subpopulation. Isotype staining: filled histogram; marker staining: open histogram. Data represent cultures harvested at day 9.

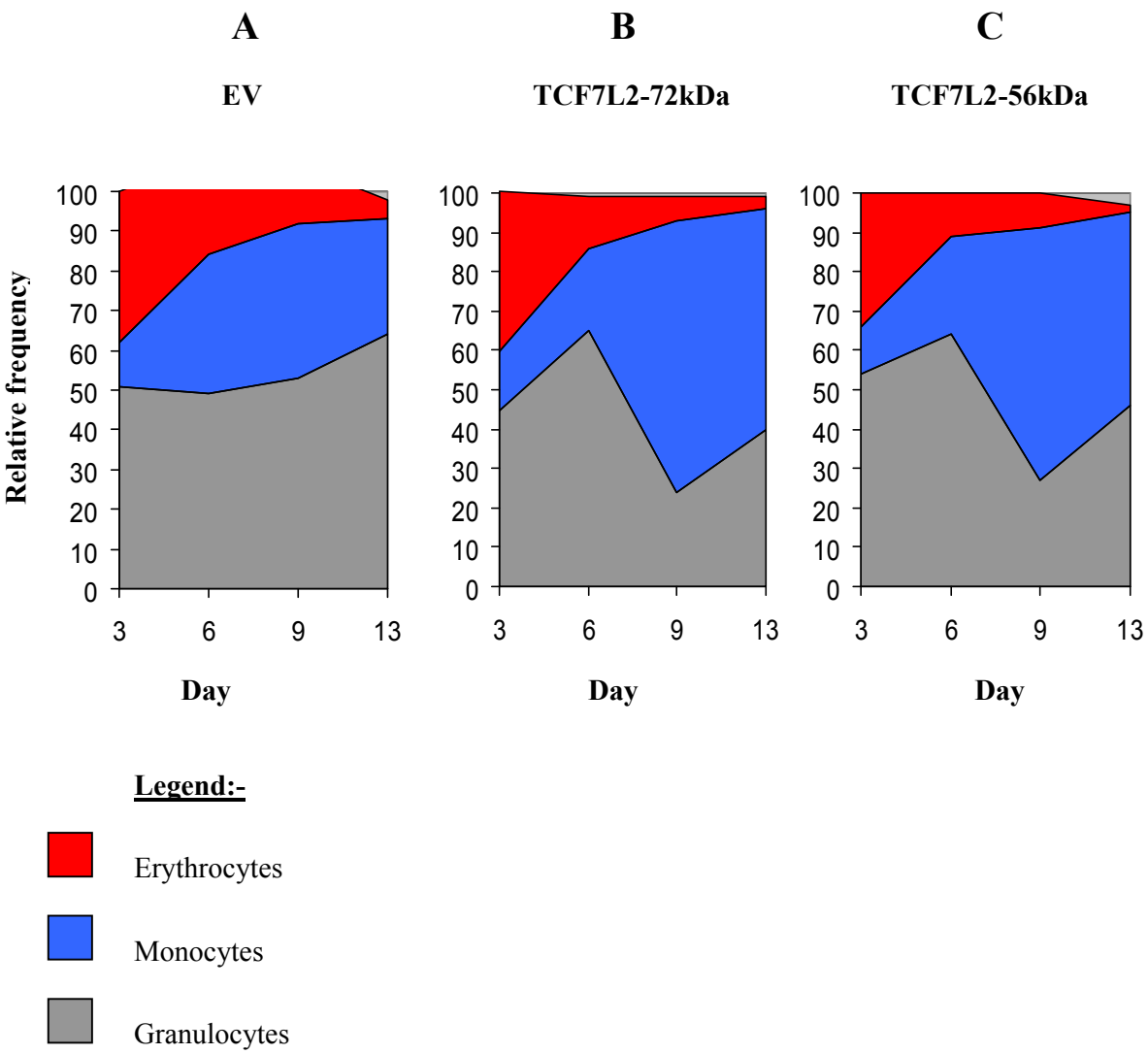


Figure 5.14: Relative frequency of erythroid, monocytic and granulocytic lineage cells in transduced CD34$^+$ HPC. Comparison of relative lineage frequency over 13 days post transduction as detected by CD13 and CD36 expression in CD34$^+$ HPC transduced with **(A)** EV **(B)** TCF7L2-72kDa **(C)** TCF7L2-56kDa.

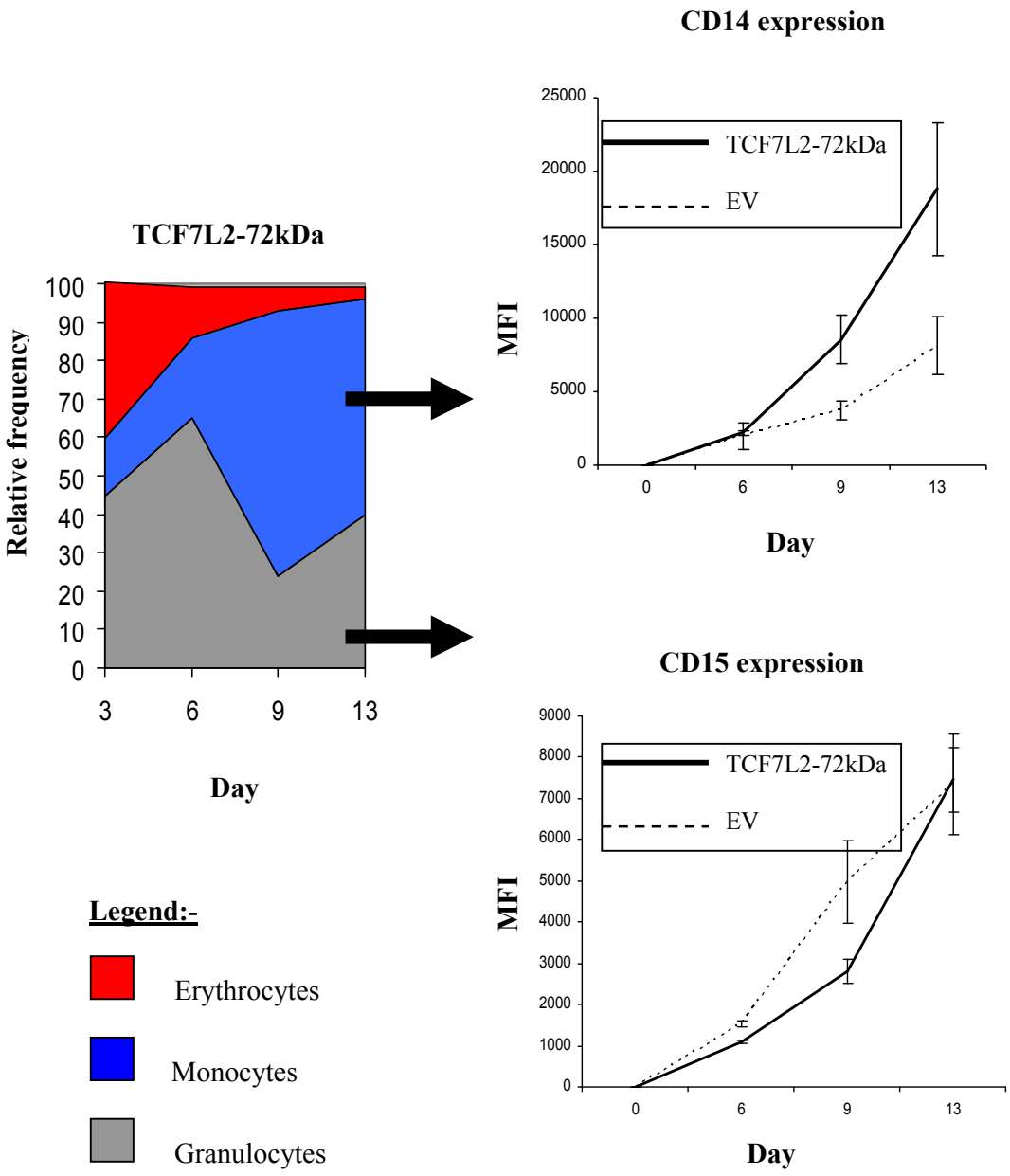


Figure 5.15: Intensity of specific lineage marker expression in CD34$^+$ HPC transduced with TCF7L2-72kDa. The levels of marker expression over 13 days post transduction are shown for **(A)** monocytic lineage (CD14$^+$) and **(B)** granulocytic lineage (CD15$^+$). Data represent the mean ± SD from two experiments. MFI = Mean Fluorescence Intensity.

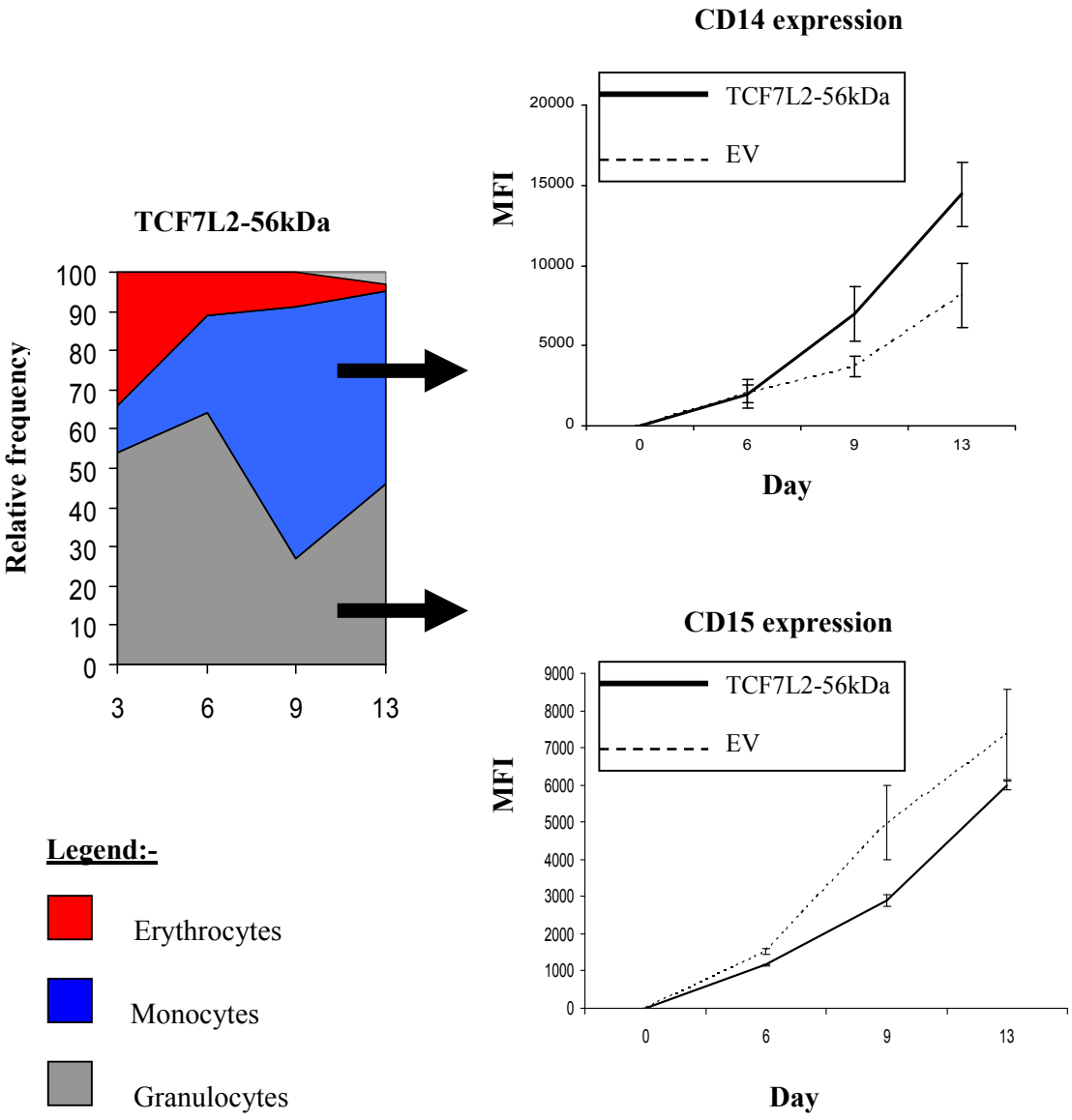


Figure 5.16: Intensity of specific lineage marker expression in CD34$^+$ HPC transduced with TCF7L2-56kDa. The levels of marker expression over 13 days post transduction are shown for **(A)** monocytic lineage (CD14$^+$) and **(B)** granulocytic lineage (CD15$^+$). Data represent the mean ± SD from two experiments. MFI = Mean Fluorescence Intensity.

5.5 DISCUSSION

5.5.1 TCF7L2 is required for leukaemia cell survival and growth

The importance of TCF7L2 expression on the growth and survival of myeloid cells was established using shRNA. Downregulation of endogenously expressed TCF7L2 mRNA and protein were shown to suppress both the growth of myeloid cells and thus transcriptional activity in response to β-catenin stabilisation. This data is in line with previous reports which suggest that TCF7L2 might be a key factor in WNT signalling involved in cancer cell proliferation. TCF7L2 is known to be normally expressed in intestinal stem cells. Transgenic mice lacking this protein ($TCF7L2^{-/-}$) have a significantly reduced proliferative activity in the intervillus regions along with reduced numbers of epithelial cells which die within 24 hours of birth (Korinek *et al*, 1998b). Conditional murine models ($TCF7L2^{LoxP/LoxP}$ knockout embryos) also demonstrated a phenotype similar to that of the classical $TCF7L2^{-/-}$ knockout mice, indicated by the total absence of proliferative cells in the small intestinal crypt (van Es *et al*, 2012). Further, a defect of cell proliferation was also reported in $TCF7L2^{-/-}$ zebrafish intestine (Muncan *et al*, 2007). In colon carcinoma, expression of dominant-negative TCF7L2 caused severe growth arrest (as measured by colony-forming assay and viable cell counting of HCT116 cells) due to cell cycle arrest in G1. The growth of these cells was reported to be strongly dependent on β-catenin / TCF7L2 mediated expression of cyclin D1 (Tetsu & McCormick, 1999).

Loss of β-catenin / TCF7L2 interaction and the reduced expression of cyclin D1 as seen in colon carcinoma, causing cell cycle arrest, could also be relevant to myeloid cells. Cyclin D1 is one of the direct TCF7L2 target genes known to have a TCF response element within its promoter region. Indeed, inhibition of WNT signalling using small molecule β-catenin / LEF1 inhibitors (eg: CGP049090 and PFK115-584) selectively induced cell death in AML cell lines and primary AML blasts within 24 hours and led to decreased protein expression of cyclin D1 and c-MYC contributing to a loss of proliferative activity (Minke *et al*, 2009). The induction of apoptosis promoted by TCF7L2 shRNA in this study is in accord with the induced apoptosis and cell cycle arrest effect of antagonists of the β-catenin / TCF7L2 complex in HepG2 and Huh7 hepatocellular carcinoma cells. All antagonists also suppressed *in vivo* tumour growth in a HepG2 xenograft model, associated with reduced *C-MYC, cyclin D1* and *survivin* expression (Wei *et al*, 2010). The impact of induced apoptosis

in colorectal cancer cells has been reported to be even stronger due to loss of TCF7L2 as compared to β-catenin (Xie *et al*, 2012), supporting a role for TCF7L2 in the survival of cancer cells. Nevertheless in this study, the role of individual isoforms is unknown because there has been no attempt to knockdown individual isoforms in any context. Most of the work on TCF7L2 in proliferation and survival has been carried out in the colorectal cells and there has been no other investigation of its role in the context of haematopoietic cells except where its role is inferred from the use of WNT signalling inhibitors which may also inhibit the activity of other TCF family members (Leung *et al*, 2002; Bao *et al*, 2012).

In contrast to the tumour promoting ability, previous studies have shown that knockdown of TCF7L2 in colorectal cell lines using TCF7L2 RNAi (Tang *et al*, 2008) could actually produce an increase in cell growth. Heterozygous TCF7L2 haploinsufficiency resulted in increased tumour formation in the colon of APC mutant mice suggesting that TCF7L2 may also function as a tumour suppressor (Angus-Hill *et al*, 2011). Angus-Hill *et al* hypothesised that TCF / LEF family members may function in distinct and opposing roles rather than overlapping roles and are essential for maintaining normal epithelial cell turnover. This was supported by the increased expression of LEF1 and TCF7 in TCF7L2 haploinsufficiency mutant APC colon tumours mentioned above. Additionally, TCF7L2 and LEF1 are also differentially regulated during colon tumour progression (Kriegl *et al*, 2010) and melanoma phenotype switch (Eichhoff *et al*, 2011) showing both are phenotype-specific and inversely correlated, with a function opposing each other. Another possible explanation for the tumour phenotype is the synergy between its homolog TCF7 and APC. TCF7 can act as a feedback transcriptional repressor of β-catenin / TCF7L2 target genes and disruption of this negative feedback loop would allow the formation of epithelial tumours similar to that seen with the loss of APC. Transcriptional activation of target genes such as c-MYC and cyclin D1 by β-catenin / TCF7L2 is thus counteracted by TCF7 as repressor (Roose *et al*, 1999). The knock-down data obtained from this study supports a role for TCF7L2 in proliferation of haematopoietic cells (rather than as a tumour suppressor).

5.5.2 Effect of TCF7L2 overexpression on canonical WNT signalling

Overexpression of TCF7L2 either in combination with β-catenin or singly elicits enhancement of a transactivation response in several cancer cell lines (Korinek *et al*, 1997; Tsedensodnom *et al*, 2011). From this study, no activation was produced by TCF7L2 isoform overexpression (72kDa, 58kDa and 56kDa). Stable reporter lines expressing the lentiviral BAR reporter were generated and the effects on TCF reporter activity of overexpression of different TFC7L2 isoforms were assayed on this background. In this study, β-catenin was stabilised by the action of the GSK3β inhibitor (BIO) or by WNT3A ligand binding. Stimulation of the WNT pathway by both agonists along with TCF7L2 overexpression showed attenuated promoter activities in the BAR reporter lines (as described in 5.4.4). A plausible explanation for this effect is that there might be insufficient β-catenin being translocated into the nucleus to bind with the excess TFC7L2 resulting in the blunted response as measured by the BAR reporter. Previously, Morgan *et al* has reported similar findings in which overexpression of catenins alone is insufficient to promote their transcriptional activity, which requires their translocation and retention in the nucleus (Morgan *et al*, 2013). Alternatively, nuclear translocation of TCF7L2 could have been impaired. However, in this study, western blotting analysis confirmed the nuclear translocation of TCF7L2 in the cell lines tested. Ectopic expression of TCF7L2 seems to overload the nucleus with TCF7L2 that potentially might bind free nuclear β-catenin promoting a loss of function effect. These data imply that there is an optimal level of TCF7L2 and that by overexpressing on a background of already high expression, the ability of these cells to respond to WNT agonists may be compromised.

TCF7L2 has a known role as a transcriptional repressor. A repression effect as seen with TCF7L2-72kDa has also been documented in colon cancer with mutated APC; DLD1 cells (Tang *et al*, 2008). Tang *et al* expressed the TCF7L2-72kDa in colorectal cancer and concluded that the C-clamp sequence of TCF7L2 contributes to its repressor function and that CtBP may regulate a tumour suppressive function by contributing to the transcriptional repression of TCF7L2 as well as by antagonising the activity of the β-catenin complex (Chinnadurai, 2002; Hamada & Bienz, 2004). In contrast, constructs expressing shorter TCF7L2 isoforms (58kDa and 56kDa) have shown augmentation of the reporter signal, as compared to the 72kDa isoform, upon β-catenin stabilisation in the context of colorectal and hepatocarcinoma tumours (Tang *et al*, 2008; Tsedensodnom *et al*, 2011). However, in this study, the overexpression of all three TCF7L2 isoforms suppressed WNT agonist responses

in myeloid cells. Whereas in the epithelial cell line, HeLa, overexpression of TCF7L2-72kDa had little effect and other studies have suggested that this is because HeLa cells already express sufficient levels of TCF to support β-catenin-dependent activation and their findings show that full-length TCF7L2-72kDa increased activation only slightly (Tetsu & McCormick, 1999). Taken together these data suggest that the effects of TCF7L2 overexpression are highly context dependent.

Another factor that could impair TCF transactivation is post-translational modifications that include association of TCF / LEF with acetyltransferases and deacetylases. Acetylation of TCF7L2-72kDa by cAMP-response-element binding protein (CREB)-binding protein (CBP) protein has been reported to inhibit transcriptional activation as seen in HEK293 cells. Furthermore, the shorter TCF7L2-56kDa or TCF7L2-58kDa isoforms are not acetylated by CBP (Elfert *et al*, 2013). Genetically, the CBP acetylase behaves as a repressor of WNT signalling preferentially by inducing acetylation of the TCF7L2-72kDa isoform (Waltzer & Bienz, 1998). Consistent with this, CBP acetylates the catenin-binding domain in TCF, thereby weakening its interaction with β-catenin and its DNA binding ability as compared to the shorter isoforms therefore giving a repression effect. These observations add to the notion that different TCF7L2 isoforms may have different roles as transcriptional repressors or activators, also depending on its binding partner proteins. The uniform effect of the different isoforms in this study suggests that in this context acetylation is not a factor in inhibiting transcriptional responses.

One caveat to the TCF reporter data is that most TCF reporters incorporate only minimal promoters (as for the pBAR series) and that in fact, most cellular promoters are not exclusively dependent on TCF activation, therefore reporter activity may not necessarily be predictive of effects on WNT responsive genes because it may be co-regulated by a variety of other factors. More significantly, natural TCF activated enhancers require additional activators such as CBP and p300 that may either link activator proteins to the basal transcription machinery or alter chromatin structure through their intrinsic activities for proper expression (Carlsson *et al*, 1993; Hecht & Kemler, 2000; Darken & Wilson, 2001).

5.5.3 TCF7L2 in haematopoietic development

Few studies have addressed the role of TCF7L2 in haematopoietic development although a number of studies have described roles of various other components of WNT signalling in erythroid and myeloid differentiation, as reviewed previously (Staal & Luis, 2010). Findings from this study indicated that an increase in TCF7L2 (in the absence of WNT agonists) is sufficient to promote the relative expansion of the monocytic lineage. At the same time it promoted the expression of monocyte differentiation antigen expression on the monocyte subpopulation. It is possible that high TCF7L2 expression alters monocyte differentiation through its transactivation of target genes such as PU.1 that have been shown previously to regulate different stages of monocytes differentiation as reviewed by (Valledor *et al*, 1998; Auffray *et al*, 2009).

Monocytes themselves are heterogeneous, with variable nuclear morphology, granularity and differential expression of CD14 and CD16 surface markers (Gordon & Taylor, 2005). Monocytes can differentiate into macrophages or dendritic cells, however it is still not known whether TCF7L2 regulates monocyte commitment or monocyte differentiation. Yet, studies have shown that TCF7L2 expression is consistently up-regulated in primary macrophages (Lehtonen *et al*, 2007), and its protein expression is induced upon macrophage differentiation in myeloid cell lines (Baek *et al*, 2009). This is consistent with high TCF7L2 being sufficient to promote monocyte differentiation or suppress granulocyte differentiation during GM-CSF directed bi-lineage terminal differentiation of FDB1 cells (Brown *et al*, 2012).

Following acute injury of HSC, TCF7L2 has been shown to be involved in the regeneration of cells by its combined action with SMAD1 to regulate erythroid differentiation (Trompouki *et al*, 2011) indicating further functional diversity of TCF7L2. Additionally, genome studies using ChIP sequencing analysis in several types of solid tumour cells revealed significant TCF7L2 binding site enrichment for other TFs, including HNF4α, FOXA2 and the GATA3 (Frietze *et al*, 2012), indicating an important relationship between TCF7L2 and other TFs in lineage regulation.

Enforced WNT signalling activation could also led to loss of myeloid lineage commitment and disruption of repopulating stem cell activity (Kirstetter *et al*, 2006), indicating the presence of WNT dosage-dependent effects. This notion is supported by work done in APC mutant mice that demonstrated the intermediate amounts of WNT

signals are sufficient to enhance myeloid differentiation while only low level of WNT signals are required to sustain normal haematopoiesis (Luis *et al*, 2011).

In summary, overexpression of TCF7L2 does not necessarily promote transcription in myeloid cell lines in the context of the minimal reporter system used. Endogeneous TCF7L2 expression is required for myeloid cell line survival supporting its role as a tumour promoting gene. In the case of normal cells, enforced expression of TCF7L2, potentiates monocytic differentiation compared to other lineages suggesting a role in regulating cell fate.

6 - General Discussion

Genome-wide expression analysis directly comparing the expression profile of normal HPC and from patients with AML in this study has identified 1595 differentially expressed genes common to different subtypes of AML (Figure 3.9). Using further downstream analysis, several pathways, including WNT signalling were identified to be dysregulated in AML; WNT is well known for its role in cancer development and stem cell biology. Furthermore, by using protein enrichment bioinformatic tools to compare the WNT genes corresponding to up and downregulated expression, a list of putative common TFs which are connected to other WNT genes were identified.

Transcriptome microarray GEP has become a frontline standard approach to identify changes in gene expression involved in both physiological and pathophysiological processes. Chapter 3 initially identified WNT as a pathway associated with dysregulated gene expression in AML blasts. As part of this pathway, the study showed for the first time that *TCF7L2* was the most dysregulated TF in AML. The *TCF7L2* gene as discussed previously is known to be susceptible to alternative splicing. This study also examined the mRNA splice products and the protein expression of these isoforms in AML. Although it is noteworthy that this pan-genome 3'IVT microarray is often described as being able to scan the whole transcriptome, they are not well suited to detect expression of isoforms generated by alternative mRNA splicing produced by *TCF7L2*. Novel strategies are now available to detect splicing based on the expression of known or predicted splice variants at the RNA level and microarrays are evolving towards the inclusion of splice-related content of the gene. The combination of exon and junction specific probe-sets will provide a robust approach to monitor known splice events. More recently, several groups have reported the used of more extensively designed probes around splicing events on a genome-wide scale (Castle *et al*, 2003; Le *et al*, 2004; Fehlbaum *et al*, 2005). This approach provided a comprehensive profile of alternatively spliced apoptotic genes in Imatinib-treated K562 cells (Liu *et al*, 2012).

In addition to the TCF7L2 isoforms identified and presented in Chapter 4, a previous study utilising embryonic cells has identified a further splice event within TCF7L2 (Vacik *et al*, 2011). Vacik *et al* discovered that a dnTCF7L2 transcript can be generated from an alternative first exon that lies hidden inside the fifth intron of the gene encoding TCF7L2, yielding a new 35 kDa isoform. Functionally, dnTCF7L2 can displace TCF7L2 in the nucleus and because this protein lacks a β-catenin binding site, it remains a transcriptional repressor even in the presence of a WNT signal and the availability of β-catenin. This isoform would not have been amplified by the PCR strategy used in this study but a TCF7L2 isoform of this molecular weight was not detected at the protein level.

TCF7L2 activity can also be influenced by alternative splicing giving rise to proteins that have different susceptibilities to post translational modification. Acetylation of TCF7L2 has been reported to be restricted to C-tail-E splice variants and requires the simultaneous presence of β-catenin. The fact that acetylation affects only a subset of TCF7L2 splice variants and is mediated preferentially by CBP suggests that the conditional acetylation of TCF7L2 is a novel regulatory mechanism that diversifies the transcriptional output of TCF / β-catenin signalling in response to changing intracellular signalling (Elfert *et al*, 2013). TCF7L2 activity can also be influenced by sumoylation which leads to its activation and possibly also affects its stability (Yamamoto *et al*, 2003). Sumoylation of TCF7L2 has however not been studied in haematopoietic cells.

Data from Chapter 5 suggests that TCF7L2 could serve a role in normal haematopoietic lineage fate; its overexpression potentiated monocytic development in haematopoietic progenitor cells. From GEP analysis done in AML however, the highest mRNA expression of TCF7L2 was found mostly within the undifferentiated AML (M0/M1) rather than AML with monocytic features (M4, M5) (Supplementary 6.1). Validation of this finding was not performed at the protein level in M1 patients. However, data from Chapter 4 demonstrates a correlation between the levels of TCF7L2 protein and mRNA in primary human AML blasts, thereby supporting the Affymetrix GEP data.

The function of TCF7L2 in cancer is widely determined from studies in colon carcinoma but very few have been tested in AML. This current study found that knockdown of TCF7L2 suppressed myeloid cancer growth. However, previous studies found that in colorectal cancer where TCF7L2 is inactivated, the cells grew more rapidly in culture, and TCF7L2 is therefore thought to function as a tumour suppressor instead of

promoting cancer (Tang *et al*, 2008; Angus-Hill *et al*, 2011). This strongly suggests dual roles of TCF7L2 that remain uninvestigated in the context of haematological malignancies.

It also remains unknown whether different isoforms of TCF7L2 differentially regulate target genes in haematopoietic cells. In further studies, it would be important to demonstrate evidence of direct regulation of WNT target gene expression by TCF7L2 in the context of myeloid cells. The search for TCF7L2 target genes has not been extensively studied in the context of haematopoiesis and a combination of different approaches are required to achieve this; e.g. the assessment of gene expression changes in HPCs with or without WNT pathway stimulation combined with chromatin immunoprecipitation binding assay (Trompouki *et al*, 2011). The regulatory region through which TCF7L2 expression is activated is among the most conserved DNA sequences across vertebrates. Through *in silico* analysis (Park *et al*, 2013) and DNA arrays (Hatzis *et al*, 2008), putative TCF7L2 binding elements have been identified in the promoter regions of several direct TCF7L2 target genes (eg: *c-MYC, ABCD1*) that might be useful to monitor the downstream output of TCF7L2 / β-catenin activation or repression; this needs to be established in the context of myeloid haematopoiesis. It will be important to evaluate the promoter activity under multiple criteria such as how do the target promoters respond to WNT agonist or β-catenin mediated signals and whether there are cooperative interactions with other TCF homologs (e.g. TCF7 and TCF7L1).

If activation of WNT signalling due to high TCF7L2 expression can be confirmed in a larger cohort of primary AML cells then this protein may represent a potential therapeutic target. Previous studies have found ways to block WNT signalling, both upstream and downstream of APC as reviewed previously (Anastas & Moon, 2013). However, WNT signalling effectors within the nucleus have been difficult to target. Improved drug discovery platforms have helped to discover new molecules that can alter the WNT pathway including small molecule compounds that target specifically the TCF7L2 / β-catenin complex (eg: iCRT3, iCRT5, BC21, NC043, PKF118-310) (Tian *et al*, 2012) and compounds that target TCF / β-catenin-dependent transcription (eg: OSU030123, 6-dihydroxyflavone) (Baryawno *et al*, 2010; Ewan *et al*, 2010). These have been demonstrated to inhibit growth of mainly colorectal cancer cells and other solid tumours (eg: medulloblastoma, breast cancer). Their efficacy in leukaemia still needs to be addressed.

7 - References

Aberle,H., Bauer,A., Stappert,J., Kispert,A., & Kemler,R. (1997) beta-catenin is a target for the ubiquitin-proteasome pathway. *EMBO J.*, **16**, 3797-3804.

Allison,D.B., Cui,X., Page,G.P., & Sabripour,M. (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nat.Rev.Genet.*, **7**, 55-65.

Alter,O., Brown,P.O., & Botstein,D. (2000) Singular value decomposition for genome-wide expression data processing and modeling. *Proc.Natl.Acad.Sci.U.S.A*, **97**, 10101-10106.

Anastas,J.N. & Moon,R.T. (2013) WNT signalling pathways as therapeutic targets in cancer. *Nat.Rev.Cancer*, **13**, 11-26.

Angus-Hill,M.L., Elbert,K.M., Hidalgo,J., & Capecchi,M.R. (2011) T-cell factor 4 functions as a tumor suppressor whose disruption modulates colon cell proliferation and tumorigenesis. *Proc.Natl.Acad.Sci.U.S.A*, **108**, 4914-4919.

Arce,L., Yokoyama,N.N., & Waterman,M.L. (2006) Diversity of LEF/TCF action in development and disease. *Oncogene*, **25**, 7492-7504.

Asou,H., Tashiro,S., Hamamoto,K., Otsuji,A., Kita,K., & Kamada,N. (1991) Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation. *Blood*, **77**, 2031-2036.

Atcha,F.A., Munguia,J.E., Li,T.W., Hovanes,K., & Waterman,M.L. (2003) A new beta-catenin-dependent activation domain in T cell factor. *J.Biol.Chem.*, **278**, 16169-16175.

Atcha,F.A., Syed,A., Wu,B., Hoverter,N.P., Yokoyama,N.N., Ting,J.H. and others (2007) A unique DNA binding domain converts T-cell factors into strong Wnt effectors. *Mol.Cell Biol.*, **27**, 8352-8363.

Avigdor,A., Goichberg,P., Shivtiel,S., Dar,A., Peled,A., Samira,S. and others (2004) CD44 and hyaluronic acid cooperate with SDF-1 in the trafficking of human CD34+ stem/progenitor cells to bone marrow. *Blood*, **103**, 2981-2989.

Baek,Y.S., Haas,S., Hackstein,H., Bein,G., Hernandez-Santana,M., Lehrach,H. and others (2009) Identification of novel transcriptional regulators involved in macrophage differentiation and activation in U937 cells. *BMC.Immunol.*, **10**, 18.

Banky,B., Raso-Barnett,L., Barbai,T., Timar,J., Becsagh,P., & Raso,E. (2012) Characteristics of CD44 alternative splice pattern in the course of human colorectal adenocarcinoma progression. *Mol.Cancer*, **11**, 83.

Barker,N., Huls,G., Korinek,V., & Clevers,H. (1999) Restricted high level expression of Tcf-4 protein in intestinal and mammary gland epithelium. *Am.J.Pathol.*, **154**, 29-35.

Barker,N., Morin,P.J., & Clevers,H. (2000) The Yin-Yang of TCF/beta-catenin signaling. *Adv.Cancer Res.*, **77**, 1-24.

Barth,A.I., Nathke,I.S., & Nelson,W.J. (1997) Cadherins, catenins and APC protein: interplay between cytoskeletal complexes and signaling pathways. *Curr.Opin.Cell Biol.*, **9**, 683-690.

Baryawno,N., Sveinbjornsson,B., Eksborg,S., Chen,C.S., Kogner,P., & Johnsen,J.I. (2010) Small-molecule inhibitors of phosphatidylinositol 3-kinase/Akt signaling inhibit Wnt/beta-catenin pathway cross-talk and suppress medulloblastoma growth. *Cancer Res.*, **70**, 266-276.

Behre,G., Zhang,P., Zhang,D.E., & Tenen,D.G. (1999) Analysis of the modulation of transcriptional activity in myelopoiesis and leukemogenesis. *Methods*, **17**, 231-237.

Behrens,J., Jerchow,B.A., Wurtele,M., Grimm,J., Asbrand,C., Wirtz,R. and others (1998) Functional interaction of an axin homolog, conductin, with beta-catenin, APC, and GSK3beta. *Science*, **280**, 596-599.

Bemmo,A., Benovoy,D., Kwan,T., Gaffney,D.J., Jensen,R.V., & Majewski,J. (2008) Gene expression and isoform variation analysis using Affymetrix Exon Arrays. *BMC.Genomics*, **9**, 529.

Benito,M., Parker,J., Du,Q., Wu,J., Xiang,D., Perou,C.M. and others (2004) Adjustment of systematic microarray data biases. *Bioinformatics.*, **20**, 105-114.

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

Bennett,L.B., Taylor,K.H., Arthur,G.L., Rahmatpanah,F.B., Hooshmand,S.I., & Caldwell,C.W. (2010) Epigenetic regulation of WNT signaling in chronic lymphocytic leukemia. *Epigenomics.*, **2**, 53-70.

Berger,M.F., Philippakis,A.A., Qureshi,A.M., He,F.S., Estep,P.W., III, & Bulyk,M.L. (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat.Biotechnol.*, **24**, 1429-1435.

Beslu,N., Krosl,J., Laurin,M., Mayotte,N., Humphries,K.R., & Sauvageau,G. (2004) Molecular interactions involved in HOXB4-induced activation of HSC self-renewal. *Blood*, **104**, 2307-2314.

Betz,B.L. & Hess,J.L. (2010) Acute myeloid leukemia diagnosis in the 21st century. *Arch.Pathol.Lab Med.*, **134**, 1427-1433.

Biechele,T.L. & Moon,R.T. (2008) Assaying beta-catenin/TCF transcription with beta-catenin/TCF transcription-based reporter constructs. *Methods Mol.Biol.*, **468**, 99-110.

Biechele,T.L., Adams,A.M., & Moon,R.T. (2009) Transcription-based reporters of Wnt/beta-catenin signaling. *Cold Spring Harb.Protoc.*, **2009**, db.

Blair,A., Hogge,D.E., & Sutherland,H.J. (1998) Most acute myeloid leukemia progenitor cells with long-term proliferative ability in vitro and in vivo have the phenotype CD34(+)/CD71(-)/HLA-DR-. *Blood*, **92**, 4325-4335.

Boj,S.F., van Es,J.H., Huch,M., Li,V.S., Jose,A., Hatzis,P. and others (2012) Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. *Cell*, **151**, 1595-1607.

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

Bradstock,K., Matthews,J., Benson,E., Page,F., & Bishop,J. (1994) Prognostic value of immunophenotyping in acute myeloid leukemia. Australian Leukaemia Study Group. *Blood*, **84**, 1220-1225.

Briddell,R.A., Brandt,J.E., Leemhuis,T.B., & Hoffman,R. (1992) Role of cytokines in sustaining long-term human megakaryocytopoiesis in vitro. *Blood*, **79**, 332-337.

Brown,A.L., Salerno,D.G., Sadras,T., Engler,G.A., Kok,C.H., Wilkinson,C.R. and others (2012) The GM-CSF receptor utilizes beta-catenin and Tcf4 to specify macrophage lineage differentiation. *Differentiation*, **83**, 47-59.

Bruno,E., Briddell,R.A., Cooper,R.J., & Hoffman,R. (1991) Effects of recombinant interleukin 11 on human megakaryocyte progenitor cells. *Exp.Hematol.*, **19**, 378-381.

Burgess,A.W. & Metcalf,D. (1980) The nature and action of granulocyte-macrophage colony stimulating factors. *Blood*, **56**, 947-958.

Burnett,A.K., Grimwade,D., Solomon,E., Wheatley,K., & Goldstone,A.H. (1999) Presenting white blood cell count and kinetics of molecular remission predict prognosis in acute promyelocytic leukemia treated with all-trans retinoic acid: result of the Randomized MRC Trial. *Blood*, **93**, 4131-4143.

Burnett,A.K. (2012) Treatment of acute myeloid leukemia: are we making progress? *Hematology.Am.Soc.Hematol.Educ.Program.*, **2012**, 1-6.

Cadigan,K.M. & Nusse,R. (1997) Wnt signaling: a common theme in animal development. *Genes Dev.*, **11**, 3286-3305.

Calin,G.A., Liu,C.G., Ferracin,M., Hyslop,T., Spizzo,R., Sevignani,C. and others (2007) Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell*, **12**, 215-229.

Carlsson,P., Waterman,M.L., & Jones,K.A. (1993) The hLEF/TCF-1 alpha HMG protein contains a context-dependent transcriptional activation domain that induces the TCR alpha enhancer in T cells. *Genes Dev.*, **7**, 2418-2430.

Casas,S., Ollila,J., Aventin,A., Vihinen,M., Sierra,J., & Knuutila,S. (2003) Changes in apoptosis-related pathways in acute myelocytic leukemia. *Cancer Genet.Cytogenet.*, **146**, 89-101.

Castano,J., Raurell,I., Piedra,J.A., Miravet,S., Dunach,M., & Garcia de,H.A. (2002) Beta-catenin N- and C-terminal tails modulate the coordinated binding of adherens junction proteins to beta-catenin. *J.Biol.Chem.*, **277**, 31541-31550.

Castle,J., Garrett-Engele,P., Armour,C.D., Duenwald,S.J., Loerch,P.M., Meyer,M.R. and others (2003) Optimization of oligonucleotide arrays and RNA amplification protocols for analysis of transcript structure and alternative splicing. *Genome Biol.*, **4**, R66.

Castrop,J., Verbeek,S., Hofhuis,F., & Clevers,H. (1995) Circumvention of tolerance for the nuclear T cell protein TCF-1 by immunization of TCF-1 knock-out mice. *Immunobiology*, **193**, 281-287.

Cavallo,R.A., Cox,R.T., Moline,M.M., Roose,J., Polevoy,G.A., Clevers,H. and others (1998) Drosophila Tcf and Groucho interact to repress Wingless signalling activity. *Nature*, **395**, 604-608.

Chang,H.R., Cheng,T.L., Liu,T.Z., Hu,H.S., Hsu,L.S., Tseng,W.C. and others (2006) Genetic and cellular characterizations of human TCF4 with microsatellite instability in colon cancer and leukemia cell lines. *Cancer Lett.*, **233**, 165-171.

Chen,C., Liu,Y., Lu,C., Cross,J.R., Morris,J.P., Shroff,A.S. and others (2013) Cancer-associated IDH2 mutants drive an acute myeloid leukemia that is susceptible to Brd4 inhibition. *Genes Dev.*, **27**, 1974-1985.

Cheshier,S.H., Morrison,S.J., Liao,X., & Weissman,I.L. (1999) In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc.Natl.Acad.Sci.U.S.A*, **96**, 3120-3125.

Cheson,B.D., Bennett,J.M., Kopecky,K.J., Buchner,T., Willman,C.L., Estey,E.H. and others (2003) Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia. *J.Clin.Oncol.*, **21**, 4642-4649.

Choi,J.K., Yu,U., Kim,S., & Yoo,O.J. (2003) Combining multiple microarray studies and modeling interstudy variation. *Bioinformatics.*, **19 Suppl 1**, i84-i90.

Chou,W.C., Tang,J.L., Wu,S.J., Tsay,W., Yao,M., Huang,S.Y. and others (2007) Clinical implications of minimal residual disease monitoring by quantitative polymerase chain reaction in acute myeloid leukemia patients bearing nucleophosmin (NPM1) mutations. *Leukemia*, **21**, 998-1004.

Chung,E.J., Hwang,S.G., Nguyen,P., Lee,S., Kim,J.S., Kim,J.W. and others (2002) Regulation of leukemic cell adhesion, proliferation, and survival by beta-catenin. *Blood*, **100**, 982-990.

Clevers,H. & van de Wetering,M. (1997) TCF/LEF factor earn their wings. *Trends Genet.*, **13**, 485-489.

Clevers,H. (2006) Wnt/beta-catenin signaling in development and disease. *Cell*, **127**, 469-480.

Cobas,M., Wilson,A., Ernst,B., Mancini,S.J., MacDonald,H.R., Kemler,R. and others (2004) Beta-catenin is dispensable for hematopoiesis and lymphopoiesis. *J.Exp.Med.*, **199**, 221-229.

Cocco,L., Manzoli,L., Palka,G., & Martelli,A.M. (2005) Nuclear phospholipase C beta1, regulation of the cell cycle and progression of acute myeloid leukemia. *Adv.Enzyme Regul.*, **45**, 126-135.

Collins,S.J., Ruscetti,F.W., Gallagher,R.E., & Gallo,R.C. (1978) Terminal differentiation of human promyelocytic leukemia cells induced by dimethyl sulfoxide and other polar compounds. *Proc.Natl.Acad.Sci.U.S.A*, **75**, 2458-2462.

Cong,F., Schweizer,L., & Varmus,H. (2004) Wnt signals across the plasma membrane to activate the beta-catenin pathway by forming oligomers containing its receptors, Frizzled and LRP. *Development*, **131**, 5103-5115.

Congdon,K.L., Voermans,C., Ferguson,E.C., DiMascio,L.N., Uqoezwa,M., Zhao,C. and others (2008) Activation of Wnt signaling in hematopoietic regeneration. *Stem Cells*, **26**, 1202-1210.

Connor,A.E., Baumgartner,R.N., Baumgartner,K.B., Kerber,R.A., Pinkston,C., John,E.M. and others (2012) Associations between TCF7L2 polymorphisms and risk of breast cancer among Hispanic and non-Hispanic white women: the Breast Cancer Health Disparities Study. *Breast Cancer Res.Treat.*, **136**, 593-602.

Cuilliere-Dartigues,P., El-Bchiri,J., Krimi,A., Buhard,O., Fontanges,P., Flejou,J.F. and others (2006) TCF-4 isoforms absent in TCF-4 mutated MSI-H colorectal cancer cells colocalize with nuclear CtBP and repress TCF-4-mediated transcription. *Oncogene*, **25**, 4441-4448.

da Cunha,C.B., Oliveira,C., Wen,X., Gomes,B., Sousa,S., Suriano,G. and others (2010) De novo expression of CD44 variants in sporadic and hereditary gastric cancer. *Lab Invest*, **90**, 1604-1614.

Daniels,D.L. & Weis,W.I. (2005) Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nat.Struct.Mol.Biol.*, **12**, 364-371.

Darken,R.S. & Wilson,P.A. (2001) Axis induction by wnt signaling: Target promoter responsiveness regulates competence. *Dev.Biol.*, **234**, 42-54.

Dash,A. & Gilliland,D.G. (2001) Molecular genetics of acute myeloid leukaemia. *Best.Pract.Res.Clin.Haematol.*, **14**, 49-64.

Ding,L., Ley,T.J., Larson,D.E., Miller,C.A., Koboldt,D.C., Welch,J.S. and others (2012) Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*, **481**, 506-510.

Duncan,A.W., Rattis,F.M., DiMascio,L.N., Congdon,K.L., Pazianos,G., Zhao,C. and others (2005) Integration of Notch and Wnt signaling in hematopoietic stem cell maintenance. *Nat.Immunol.*, **6**, 314-322.

Dupuy,A. & Simon,R.M. (2007) Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J.Natl.Cancer Inst.*, **99**, 147-157.

Duval,A., Iacopetta,B., Ranzani,G.N., Lothe,R.A., Thomas,G., & Hamelin,R. (1999) Variable mutation frequencies in coding repeats of TCF-4 and other target genes in colon, gastric and endometrial carcinoma showing microsatellite instability. *Oncogene*, **18**, 6806-6809.

Duval,A., Rolland,S., Tubacher,E., Bui,H., Thomas,G., & Hamelin,R. (2000) The human T-cell transcription factor-4 gene: structure, extensive characterization of alternative splicings, and mutational analysis in colorectal cancer cell lines. *Cancer Res.*, **60**, 3872-3879.

Dykstra,B., Kent,D., Bowie,M., McCaffrey,L., Hamilton,M., Lyons,K. and others (2007) Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell*, **1**, 218-229.

Eastman,Q. & Grosschedl,R. (1999) Regulation of LEF-1/TCF transcription factors by Wnt and other signals. *Curr.Opin.Cell Biol.*, **11**, 233-240.

Ebert,B.L. (2010) Genetic deletions in AML and MDS. *Best.Pract.Res.Clin.Haematol.*, **23**, 457-461.

Eichhoff,O.M., Weeraratna,A., Zipser,M.C., Denat,L., Widmer,D.S., Xu,M. and others (2011) Differential LEF1 and TCF4 expression is involved in melanoma cell phenotype switching. *Pigment Cell Melanoma Res.*, **24**, 631-642.

Eisennhart,C. (1947) The assumptions underlying the analysis of variance. *Biometrics*, **3**, 21.

Elfert,S., Weise,A., Bruser,K., Biniossek,M.L., Jagle,S., Senghaas,N. and others (2013) Acetylation of human TCF4 (TCF7L2) proteins attenuates inhibition by the HBP1 repressor and induces a conformational change in the TCF4::DNA complex. *PLoS.One.*, **8**, e61867.

Estey,E. & Dohner,H. (2006) Acute myeloid leukaemia. *Lancet*, **368**, 1894-1907.

Ewan,K., Pajak,B., Stubbs,M., Todd,H., Barbeau,O., Quevedo,C. and others (2010) A useful approach to identify novel small-molecule inhibitors of Wnt-dependent transcription. *Cancer Res.*, **70**, 5963-5973.

Fang,M., Li,J., Blauwkamp,T., Bhambhani,C., Campbell,N., & Cadigan,K.M. (2006) C-terminal-binding protein directly activates and represses Wnt transcriptional targets in Drosophila. *EMBO J.*, **25**, 2735-2745.

Faubert,A., Chagraoui,J., Mayotte,N., Frechette,M., Iscove,N.N., Humphries,R.K. and others (2008) Complementary and independent function for Hoxb4 and Bmi1 in HSC activity. *Cold Spring Harb.Symp.Quant.Biol.*, **73**, 555-564.

Fehlbaum,P., Guihal,C., Bracco,L., & Cochet,O. (2005) A microarray configuration to quantify expression levels and relative abundance of splice variants. *Nucleic Acids Res.*, **33**, e47.

Fenaux,P. & Detourmignies,L. (1994) Therapy-related acute promyelocytic leukaemia. *Br.J.Haematol.*, **87**, 445.

Ferrell,C.M., Dorsam,S.T., Ohta,H., Humphries,R.K., Derynck,M.K., Haqq,C. and others (2005) Activation of stem-cell specific genes by HOXA9 and HOXA10 homeodomain proteins in CD34+ human cord blood cells. *Stem Cells*, **23**, 644-655.

Figueroa,M.E., Abdel-Wahab,O., Lu,C., Ward,P.S., Patel,J., Shih,A. and others (2010) Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*, **18**, 553-567.

Fleming,H.E., Janzen,V., Lo,C.C., Guo,J., Leahy,K.M., Kronenberg,H.M. and others (2008) Wnt signaling in the niche enforces hematopoietic stem cell quiescence and is necessary to preserve self-renewal in vivo. *Cell Stem Cell*, **2**, 274-283.

Foran,J.M. (2010) New prognostic markers in acute myeloid leukemia: perspective from the clinic. *Hematology.Am.Soc.Hematol.Educ.Program.*, **2010**, 47-55.

Franco,C.A., Liebner,S., & Gerhardt,H. (2009) Vascular morphogenesis: a Wnt for every vessel? *Curr.Opin.Genet.Dev.*, **19**, 476-483.

Frietze,S., Wang,R., Yao,L., Tak,Y.G., Ye,Z., Gaddis,M. and others (2012) Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. *Genome Biol.*, **13**, R52.

Gan,X.Q., Wang,J.Y., Xi,Y., Wu,Z.L., Li,Y.P., & Li,L. (2008) Nuclear Dvl, c-Jun, beta-catenin, and TCF form a complex leading to stabilization of beta-catenin-TCF interaction. *J.Cell Biol.*, **180**, 1087-1100.

Godin,I., Dieterlen-Lievre,F., & Cumano,A. (1995) Emergence of multipotent hemopoietic cells in the yolk sac and paraaortic splanchnopleura in mouse embryos, beginning at 8.5 days postcoitus. *Proc.Natl.Acad.Sci.U.S.A*, **92**, 773-777.

Goessling,W., North,T.E., Lord,A.M., Ceol,C., Lee,S., Weidinger,G. and others (2008) APC mutant zebrafish uncover a changing temporal requirement for wnt signaling in liver development. *Dev.Biol.*, **320**, 161-174.

Goessling,W., North,T.E., Loewer,S., Lord,A.M., Lee,S., Stoick-Cooper,C.L. and others (2009) Genetic interaction of PGE2 and Wnt signaling regulates developmental specification of stem cells and regeneration. *Cell*, **136**, 1136-1147.

Golan,T., Yaniv,A., Bafico,A., Liu,G., & Gazit,A. (2004) The human Frizzled 6 (HFz6) acts as a negative regulator of the canonical Wnt. beta-catenin signaling cascade. *J.Biol.Chem.*, **279**, 14879-14888.

Goodell,M.A., Rosenzweig,M., Kim,H., Marks,D.F., DeMaria,M., Paradis,G. and others (1997) Dye efflux studies suggest that hematopoietic stem cells expressing low or undetectable levels of CD34 antigen exist in multiple species. *Nat.Med.*, **3**, 1337-1345.

Gordon,M.D. & Nusse,R. (2006) Wnt signaling: multiple pathways, multiple receptors, and multiple transcription factors. *J.Biol.Chem.*, **281**, 22429-22433.

Gordon,S. & Taylor,P.R. (2005) Monocyte and macrophage heterogeneity. *Nat.Rev.Immunol.*, **5**, 953-964.

Grant,S.F. (2012) Understanding the elusive mechanism of action of TCF7L2 in metabolism. *Diabetes*, **61**, 2657-2658.

Greenawalt,D.M., Sieberts,S.K., Cornelis,M.C., Girman,C.J., Zhong,H., Yang,X. and others (2012) Integrating genetic association, genetics of gene expression, and single nucleotide polymorphism set analysis to identify susceptibility Loci for type 2 diabetes mellitus. *Am.J.Epidemiol.*, **176**, 423-430.

Greenbaum,A., Hsu,Y.M., Day,R.B., Schuettpelz,L.G., Christopher,M.J., Borgerding,J.N. and others (2013) CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. *Nature*, **495**, 227-230.

Greenberg,P.L., Lee,S.J., Advani,R., Tallman,M.S., Sikic,B.I., Letendre,L. and others (2004) Mitoxantrone, etoposide, and cytarabine with or without valspodar in patients with relapsed or refractory acute myeloid leukemia and high-risk myelodysplastic syndrome: a phase III trial (E2995). *J.Clin.Oncol.*, **22**, 1078-1086.

Gregory,M.A., Phang,T.L., Neviani,P., Alvarez-Calderon,F., Eide,C.A., O'Hare,T. and others (2010) Wnt/Ca2+/NFAT signaling maintains survival of Ph+ leukemia cells upon inhibition of Bcr-Abl. *Cancer Cell*, **18**, 74-87.

Grignani,F., Kinsella,T., Mencarelli,A., Valtieri,M., Riganelli,D., Grignani,F. and others (1998) High-efficiency gene transfer and selection of human hematopoietic progenitor cells with a hybrid EBV/retroviral vector expressing the green fluorescence protein. *Cancer Res.*, **58**, 14-19.

Grimwade,D., Walker,H., Oliver,F., Wheatley,K., Harrison,C., Harrison,G. and others (1998) The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood*, **92**, 2322-2333.

Grimwade,D. & Hills,R.K. (2009) Independent prognostic factors for AML outcome. *Hematology.Am.Soc.Hematol.Educ.Program.*, 385-395.

Grimwade,D., Hills,R.K., Moorman,A.V., Walker,H., Chatters,S., Goldstone,A.H. and others (2010) Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood*, **116**, 354-365.

Guo,Z., Dose,M., Kovalovsky,D., Chang,R., O'Neil,J., Look,A.T. and others (2007) Beta-catenin stabilization stalls the transition from double-positive to single-positive stage and predisposes thymocytes to malignant transformation. *Blood*, **109**, 5463-5472.

Habas,R., Kato,Y., & He,X. (2001) Wnt/Frizzled activation of Rho regulates vertebrate gastrulation and requires a novel Formin homology protein Daam1. *Cell*, **107**, 843-854.

Haferlach,T., Kohlmann,A., Klein,H.U., Ruckert,C., Dugas,M., Williams,P.M. and others (2009) AML with translocation t(8;16)(p11;p13) demonstrates unique cytomorphological, cytogenetic, molecular and prognostic features. *Leukemia*, **23**, 934-943.

Harris,N.L., Jaffe,E.S., Diebold,J., Flandrin,G., Muller-Hermelink,H.K., Vardiman,J. and others (1999) World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997. *J.Clin.Oncol.*, **17**, 3835-3849.

Hatzis,P., van der Flier,L.G., van Driel,M.A., Guryev,V., Nielsen,F., Denissov,S. and others (2008) Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells. *Mol.Cell Biol.*, **28**, 2732-2744.

He,X.C., Zhang,J., Tong,W.G., Tawfik,O., Ross,J., Scoville,D.H. and others (2004) BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. *Nat.Genet.*, **36**, 1117-1121.

Hecht,A. & Kemler,R. (2000) Curbing the nuclear activities of beta-catenin. Control over Wnt target gene expression. *EMBO Rep.*, **1**, 24-28.

Hecht,A. & Stemmler,M.P. (2003) Identification of a promoter-specific transcriptional activation domain at the C terminus of the Wnt effector protein T-cell factor 4. *J.Biol.Chem.*, **278**, 3776-3785.

Held,W., Clevers,H., & Grosschedl,R. (2003) Redundant functions of TCF-1 and LEF-1 during T and NK cell development, but unique role of TCF-1 for Ly49 NK cell receptor acquisition. *Eur.J.Immunol.*, **33**, 1393-1398.

Hidalgo,A., Sanz-Rodriguez,F., Rodriguez-Fernandez,J.L., Albella,B., Blaya,C., Wright,N. and others (2001) Chemokine stromal cell-derived factor-1alpha modulates VLA-4 integrin-dependent adhesion to fibronectin and VCAM-1 on bone marrow hematopoietic progenitor cells. *Exp.Hematol.*, **29**, 345-355.

Hirabayashi,Y., Itoh,Y., Tabata,H., Nakajima,K., Akiyama,T., Masuyama,N. and others (2004) The Wnt/beta-catenin pathway directs neuronal differentiation of cortical neural precursor cells. *Development*, **131**, 2791-2801.

Hong,C.F., Chou,Y.T., Lin,Y.S., & Wu,C.W. (2009) MAD2B, a novel TCF4-binding protein, modulates TCF4-mediated epithelial-mesenchymal transdifferentiation. *J.Biol.Chem.*, **284**, 19613-19622.

Hoppler,S. & Kavanagh,C.L. (2007) Wnt signalling: variety at the core. *J.Cell Sci.*, **120**, 385-393.

Hsu,W., Zeng,L., & Costantini,F. (1999) Identification of a domain of Axin that binds to the serine/threonine protein phosphatase 2A and a self-binding domain. *J.Biol.Chem.*, **274**, 3439-3445.

Hsu,Y.C. & Fuchs,E. (2012) A family business: stem cell progeny join the niche to regulate homeostasis. *Nat.Rev.Mol.Cell Biol.*, **13**, 103-114.

Hu,Z., Fan,C., Oh,D.S., Marron,J.S., He,X., Qaqish,B.F. and others (2006) The molecular portraits of breast tumors are conserved across microarray platforms. *BMC.Genomics*, **7**, 96.

Hurlstone,A. & Clevers,H. (2002) T-cell factors: turn-ons and turn-offs. *EMBO J.*, **21**, 2303-2311.

Hwang,S.G., Lee,H.C., Trepel,J.B., & Jeon,B.H. (2002) Anticancer-drug-induced apoptotic cell death in leukemia cells is associated with proteolysis of beta-catenin. *Leuk.Res.*, **26**, 863-871.

Iakoucheva,L.M., Kimzey,A.L., Masselon,C.D., Smith,R.D., Dunker,A.K., & Ackerman,E.J. (2001) Aberrant mobility phenomena of the DNA repair protein XPA. *Protein Sci.*, **10**, 1353-1362.

Ihaka,R. & Gentleman,R. (1996) R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299-314.

Ioannidis,J.P., Allison,D.B., Ball,C.A., Coulibaly,I., Cui,X., Culhane,A.C. and others (2009) Repeatability of published microarray gene expression analyses. *Nat.Genet.*, **41**, 149-155.

Ishitani,T., Ninomiya-Tsuji,J., Nagai,S., Nishita,M., Meneghini,M., Barker,N. and others (1999) The TAK1-NLK-MAPK-related pathway antagonizes signalling between beta-catenin and transcription factor TCF. *Nature*, **399**, 798-802.

Ishitani,T., Kishida,S., Hyodo-Miura,J., Ueno,N., Yasuda,J., Waterman,M. and others (2003) The TAK1-NLK mitogen-activated protein kinase cascade functions in the Wnt-5a/Ca(2+) pathway to antagonize Wnt/beta-catenin signaling. *Mol.Cell Biol.*, **23**, 131-139.

Ishitani,T., Ninomiya-Tsuji,J., & Matsumoto,K. (2003) Regulation of lymphoid enhancer factor 1/T-cell factor by mitogen-activated protein kinase-related Nemo-like kinase-dependent phosphorylation in Wnt/beta-catenin signaling. *Mol.Cell Biol.*, **23**, 1379-1389.

Isken,F., Steffen,B., Merk,S., Dugas,M., Markus,B., Tidow,N. and others (2008) Identification of acute myeloid leukaemia associated microRNA expression patterns. *Br.J.Haematol.*, **140**, 153-161.

Jaatinen,T., Hemmoranta,H., Hautaniemi,S., Niemi,J., Nicorici,D., Laine,J. and others (2006) Global gene expression profile of human cord blood-derived CD133+ cells. *Stem Cells*, **24**, 631-641.

Jan,M., Snyder,T.M., Corces-Zimmerman,M.R., Vyas,P., Weissman,I.L., Quake,S.R. and others (2012) Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci.Transl.Med.*, **4**, 149ra118.

Jan,M. & Majeti,R. (2013) Clonal evolution of acute leukemia genomes. *Oncogene*, **32**, 135-140.

Jeffery,I.B., Higgins,D.G., & Culhane,A.C. (2006) Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC.Bioinformatics.*, **7**, 359.

Jiang,Y., Dunbar,A., Gondek,L.P., Mohan,S., Rataul,M., O'Keefe,C. and others (2009) Aberrant DNA methylation is a dominant mechanism in MDS progression to AML. *Blood*, **113**, 1315-1325.

Jin,L., Hope,K.J., Zhai,Q., Smadja-Joffe,F., & Dick,J.E. (2006) Targeting of CD44 eradicates human acute myeloid leukemic stem cells. *Nat.Med.*, **12**, 1167-1174.

Johnson,W.E., Li,C., & Rabinovic,A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.*, **8**, 118-127.

Jost,E., Schmid,J., Wilop,S., Schubert,C., Suzuki,H., Herman,J.G. and others (2008) Epigenetic inactivation of secreted Frizzled-related proteins in acute myeloid leukaemia. *Br.J.Haematol.*, **142**, 745-753.

Kinsella,T.M. & Nolan,G.P. (1996) Episomal vectors rapidly and stably produce high-titer recombinant retrovirus. *Hum.Gene Ther.*, **7**, 1405-1413.

Kirstetter,P., Anderson,K., Porse,B.T., Jacobsen,S.E., & Nerlov,C. (2006) Activation of the canonical Wnt pathway leads to loss of hematopoietic stem cell repopulation and multilineage differentiation block. *Nat.Immunol.*, **7**, 1048-1056.

Knapper,S. (2011) The clinical development of FLT3 inhibitors in acute myeloid leukemia. *Expert.Opin.Investig.Drugs*, **20**, 1377-1395.

Kohlmann,A., Schoch,C., Schnittger,S., Dugas,M., Hiddemann,W., Kern,W. and others (2003) Molecular characterization of acute leukemias by use of microarray technology. *Genes Chromosomes.Cancer*, **37**, 396-405.

Kohlmann,A., Haschke-Becher,E., Wimmer,B., Huber-Wechselberger,A., Meyer-Monard,S., Huxol,H. and others (2008a) Intraplatform reproducibility and technical precision of gene expression profiling in 4 laboratories investigating 160 leukemia samples: the DACH study. *Clin.Chem.*, **54**, 1705-1715.

Kohlmann,A., Kipps,T.J., Rassenti,L.Z., Downing,J.R., Shurtleff,S.A., Mills,K.I. and others (2008b) An international standardization programme towards the application of gene expression profiling in routine leukaemia diagnostics: the Microarray Innovations in LEukemia study prephase. *Br.J.Haematol.*, **142**, 802-807.

Kohlmann,A., Bullinger,L., Thiede,C., Schaich,M., Schnittger,S., Dohner,K. and others (2010) Gene expression profiling in AML with normal karyotype can predict mutations for molecular markers and allows novel insights into perturbed biological pathways. *Leukemia*, **24**, 1216-1220.

Kokolus,K. & Nemeth,M.J. (2010) Non-canonical Wnt signaling pathways in hematopoiesis. *Immunol.Res.*, **46**, 155-164.

Koreth,J., Schlenk,R., Kopecky,K.J., Honda,S., Sierra,J., Djulbegovic,B.J. and others (2009) Allogeneic stem cell transplantation for acute myeloid leukemia in first complete remission: systematic review and meta-analysis of prospective clinical trials. *JAMA*, **301**, 2349-2361.

Korinek,V., Barker,N., Morin,P.J., van,W.D., de,W.R., Kinzler,K.W. and others (1997) Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. *Science*, **275**, 1784-1787.

Korinek,V., Barker,N., Moerer,P., van,D.E., Huls,G., Peters,P.J. and others (1998a) Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat.Genet.*, **19**, 379-383.

Korinek,V., Barker,N., Willert,K., Molenaar,M., Roose,J., Wagenaar,G. and others (1998b) Two members of the Tcf family implicated in Wnt/beta-catenin signaling during embryogenesis in the mouse. *Mol.Cell Biol.*, **18**, 1248-1256.

Krause,D.S. & Van Etten,R.A. (2007) Right on target: eradicating leukemic stem cells. *Trends Mol.Med.*, **13**, 470-481.

Kriegl,L., Horst,D., Reiche,J.A., Engel,J., Kirchner,T., & Jung,A. (2010) LEF-1 and TCF4 expression correlate inversely with survival in colorectal cancer. *J.Transl.Med.*, **8**, 123.

Krzanowski,W.J. (1980) Mixtures of continuous and categorical variables in discriminant analysis. *Biometrics*, **36**, 493-499.

Kvinlaug,B.T., Chan,W.I., Bullinger,L., Ramaswami,M., Sears,C., Foster,D. and others (2011) Common and overlapping oncogenic pathways contribute to the evolution of acute myeloid leukemias. *Cancer Res.*, **71**, 4117-4129.

Lagasse,E. & Weissman,I.L. (1997) Enforced expression of Bcl-2 in monocytes rescues macrophages and partially reverses osteopetrosis in op/op mice. *Cell*, **89**, 1021-1031.

Lai,S.L., Chien,A.J., & Moon,R.T. (2009) Wnt/Fz signaling and the cytoskeleton: potential roles in tumorigenesis. *Cell Res.*, **19**, 532-545.

Lan,F., Yue,X., Han,L., Shi,Z., Yang,Y., Pu,P. and others (2012) Genome-wide identification of TCF7L2/TCF4 target miRNAs reveals a role for miR-21 in Wnt-driven epithelial cancer. *Int.J.Oncol.*, **40**, 519-526.

Lange,B., Valtieri,M., Santoli,D., Caracciolo,D., Mavilio,F., Gemperlein,I. and others (1987) Growth factor requirements of childhood acute leukemia: establishment of GM-CSF-dependent cell lines. *Blood*, **70**, 192-199.

Lanotte,M., Martin-Thouvenin,V., Najman,S., Balerini,P., Valensi,F., & Berger,R. (1991) NB4, a maturation inducible cell line with t(15;17) marker isolated from a human acute promyelocytic leukemia (M3). *Blood*, **77**, 1080-1086.

Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H. and others (2007) Clustal W and Clustal X version 2.0. *Bioinformatics.*, **23**, 2947-2948.

Larson,R.A., Boogaerts,M., Estey,E., Karanes,C., Stadtmauer,E.A., Sievers,E.L. and others (2002) Antibody-targeted chemotherapy of older patients with acute myeloid leukemia in first relapse using Mylotarg (gemtuzumab ozogamicin). *Leukemia*, **16**, 1627-1636.

Le,B.O., Shu,L., Marchand,M., Neve,B., Paroni,F., Kerr,C.J. and others (2011a) TCF7L2 splice variants have distinct effects on beta-cell turnover and function. *Hum.Mol.Genet.*, **20**, 1906-1915.

Le,K., Mitsouras,K., Roy,M., Wang,Q., Xu,Q., Nelson,S.F. and others (2004) Detecting tissue-specific regulation of alternative splicing as a qualitative change in microarray data. *Nucleic Acids Res.*, **32**, e180.

Lehtonen,A., Ahlfors,H., Veckman,V., Miettinen,M., Lahesmaa,R., & Julkunen,I. (2007) Gene expression profiling during differentiation of human monocytes to macrophages or dendritic cells. *J.Leukoc.Biol.*, **82**, 710-720.

Leong,H.S., Yates,T., Wilson,C., & Miller,C.J. (2005) ADAPT: a database of affymetrix probesets and transcripts. *Bioinformatics.*, **21**, 2552-2553.

Ley,T.J., Mardis,E.R., Ding,L., Fulton,B., McLellan,M.D., Chen,K. and others (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*, **456**, 66-72.

Li,C. (2008) Automating dChip: toward reproducible sharing of microarray data analysis. *BMC.Bioinformatics.*, **9**, 231.

Licchesi,J.D., Westra,W.H., Hooker,C.M., Machida,E.O., Baylin,S.B., & Herman,J.G. (2008) Epigenetic alteration of Wnt pathway antagonists in progressive glandular neoplasia of the lung. *Carcinogenesis*, **29**, 895-904.

Lichtinger,M., Ingram,R., Hannah,R., Muller,D., Clarke,D., Assi,S.A. and others (2012) RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J.*, **31**, 4318-4333.

Liddiard,K., Hills,R., Burnett,A.K., Darley,R.L., & Tonks,A. (2010) OGG1 is a novel prognostic indicator in acute myeloid leukaemia. *Oncogene*, **29**, 2005-2012.

Liu,J., Xiao,Y., Xiong,H.M., Li,J., Huang,B., Zhang,H.B. and others (2012) Alternative splicing of apoptosis-related genes in imatinib-treated K562 cells identified by exon array analysis. *Int.J.Mol.Med.*, **29**, 690-698.

Liu,X., Mazanek,P., Dam,V., Wang,Q., Zhao,H., Guo,R. and others (2008) Deregulated Wnt/beta-catenin program in high-risk neuroblastomas without MYCN amplification. *Oncogene*, **27**, 1478-1488.

Lo,Y.H., Wu,C.C., Shih,H.M., & Lai,M.Z. (2008) Selective activation of NFAT by promyelocytic leukemia protein. *Oncogene*, **27**, 3821-3830.

Lowenberg,B. & Terpstra,W. (1998) Maturation hierarchy of leukemic stem cells. *Stem Cells*, **16 Suppl 1**, 85-88.

Lozzio,C.B. & Lozzio,B.B. (1975) Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood*, **45**, 321-334.

Lu,D., Zhao,Y., Tawatao,R., Cottam,H.B., Sen,M., Leoni,L.M. and others (2004) Activation of the Wnt signaling pathway in chronic lymphocytic leukemia. *Proc.Natl.Acad.Sci.U.S.A*, **101**, 3118-3123.

Luis,T.C., Ichii,M., Brugman,M.H., Kincade,P., & Staal,F.J. (2012) Wnt signaling strength regulates normal hematopoiesis and its deregulation is involved in leukemia development. *Leukemia*, **26**, 414-421.

Luo,J., Zhao,L., Chen,A.Y., Zhang,X., Zhu,J., Zhao,J. and others (2013) TCF7L2 variation and proliferative diabetic retinopathy. *Diabetes*, **62**, 2613-2617.

Lutz,C., Hoang,V.T., Buss,E., & Ho,A.D. (2013) Identifying leukemia stem cells - Is it feasible and does it matter? *Cancer Lett.*, **338**, 10-14.

MacDonald,B.T., Tamai,K., & He,X. (2009) Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev.Cell*, **17**, 9-26.

Magrane,M. & Consortium,U. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database.(Oxford)*, **2011**, bar009.

Majeti,R., Becker,M.W., Tian,Q., Lee,T.L., Yan,X., Liu,R. and others (2009) Dysregulated gene expression networks in human acute myelogenous leukemia stem cells. *Proc.Natl.Acad.Sci.U.S.A*, **106**, 3396-3401.

Malhotra,S. & Kincade,P.W. (2009) Wnt-related molecules and signaling pathway equilibrium in hematopoiesis. *Cell Stem Cell*, **4**, 27-36.

Mardis,E.R., Ding,L., Dooling,D.J., Larson,D.E., McLellan,M.D., Chen,K. and others (2009) Recurring mutations found by sequencing an acute myeloid leukemia genome. *N.Engl.J.Med.*, **361**, 1058-1066.

Martin,V., Valencia,A., Agirre,X., Cervera,J., San Jose-Eneriz,E., Vilas-Zornoza,A. and others (2010) Epigenetic regulation of the non-canonical Wnt pathway in acute myeloid leukemia. *Cancer Sci.*, **101**, 425-432.

Masters,J.R. (2002) HeLa cells 50 years on: the good, the bad and the ugly. *Nat.Rev.Cancer*, **2**, 315-319.

Matagne,A., Joris,B., & Frere,J.M. (1991) Anomalous behaviour of a protein during SDS/PAGE corrected by chemical modification of carboxylic groups. *Biochem.J.*, **280 (Pt 2)**, 553-556.

McCulloch,E.A., Kelleher,C.A., Miyauchi,J., Wang,C., Cheng,G.Y., Minden,M.D. and others (1988) Heterogeneity in acute myeloblastic leukemia. *Leukemia*, **2**, 38S-49S.

McNiece,I., Andrews,R., Stewart,M., Clark,S., Boone,T., & Quesenberry,P. (1989) Action of interleukin-3, G-CSF, and GM-CSF on highly enriched human hematopoietic progenitor cells: synergistic interaction of GM-CSF plus G-CSF. *Blood*, **74**, 110-114.

Medvinsky,A.L., Samoylina,N.L., Muller,A.M., & Dzierzak,E.A. (1993) An early pre-liver intraembryonic source of CFU-S in the developing mouse. *Nature*, **364**, 64-67.

Merrill,B.J., Pasolli,H.A., Polak,L., Rendl,M., Garcia-Garcia,M.J., Anderson,K.V. and others (2004) Tcf3: a transcriptional regulator of axis induction in the early embryo. *Development*, **131**, 263-274.

Meshinchi,S. & Arceci,R.J. (2007) Prognostic factors and risk-based therapy in pediatric acute myeloid leukemia. *Oncologist.*, **12**, 341-355.

Metcalf,D. (1980) Clonal analysis of proliferation and differentiation of paired daughter cells: action of granulocyte-macrophage colony-stimulating factor on granulocyte-macrophage precursors. *Proc.Natl.Acad.Sci.U.S.A*, **77**, 5327-5330.

Metcalf,D. & Burgess,A.W. (1982) Clonal analysis of progenitor cell commitment of granulocyte or macrophage production. *J.Cell Physiol*, **111**, 275-283.

Mieczkowski,J., Tyburczy,M.E., Dabrowski,M., & Pokarowski,P. (2010) Probe set filtering increases correlation between Affymetrix GeneChip and qRT-PCR expression measurements. *BMC.Bioinformatics.*, **11**, 104.

Mikesch,J.H., Steffen,B., Berdel,W.E., Serve,H., & Muller-Tidow,C. (2007) The emerging role of Wnt signaling in the pathogenesis of acute myeloid leukemia. *Leukemia*, **21**, 1638-1647.

Mills,K.I., Kohlmann,A., Williams,P.M., Wieczorek,L., Liu,W.M., Li,R. and others (2009) Microarray-based classifiers and prognosis models identify subgroups with distinct clinical outcomes and high risk of AML transformation of myelodysplastic syndrome. *Blood*, **114**, 1063-1072.

Miranda-Saavedra,D. & Gottgens,B. (2008) Transcriptional regulatory networks in haematopoiesis. *Curr.Opin.Genet.Dev.*, **18**, 530-535.

Moore,A.C., Amann,J.M., Williams,C.S., Tahinci,E., Farmer,T.E., Martinez,J.A. and others (2008) Myeloid translocation gene family members associate with T-cell factors (TCFs) and influence TCF-dependent transcription. *Mol.Cell Biol.*, **28**, 977-987.

Morgan,R.G., Liddiard,K., Pearn,L., Pumford,S.L., Burnett,A.K., Darley,R.L. and others (2013) gamma-Catenin is expressed throughout normal human hematopoietic development and is required for normal PU.1-dependent monocyte differentiation. *Leukemia*.

Morin,P.J., Sparks,A.B., Korinek,V., Barker,N., Clevers,H., Vogelstein,B. and others (1997) Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science*, **275**, 1787-1790.

Morita,Y., Ema,H., & Nakauchi,H. (2010) Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment. *J.Exp.Med.*, **207**, 1173-1182.

Morrison,S.J. & Weissman,I.L. (1994) The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity.*, **1**, 661-673.

Morrison,S.J., Wandycz,A.M., Akashi,K., Globerson,A., & Weissman,I.L. (1996) The aging of hematopoietic stem cells. *Nat.Med.*, **2**, 1011-1016.

Muller,A.M., Medvinsky,A., Strouboulis,J., Grosveld,F., & Dzierzak,E. (1994) Development of hematopoietic stem cell activity in the mouse embryo. *Immunity.*, **1**, 291-301.

Muller-Sieburg,C.E., Cho,R.H., Thoman,M., Adkins,B., & Sieburg,H.B. (2002) Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. *Blood*, **100**, 1302-1309.

Muller-Tidow,C., Steffen,B., Cauvet,T., Tickenbrock,L., Ji,P., Diederichs,S. and others (2004) Translocation products in acute myeloid leukemia activate the Wnt signaling pathway in hematopoietic cells. *Mol.Cell Biol.*, **24**, 2890-2904.

Muncan,V., Faro,A., Haramis,A.P., Hurlstone,A.F., Wienholds,E., van,E.J. and others (2007) T-cell factor 4 (Tcf7l2) maintains proliferative compartments in zebrafish intestine. *EMBO Rep.*, **8**, 966-973.

Nagalski,A., Irimia,M., Szewczyk,L., Ferran,J.L., Misztal,K., Kuznicki,J. and others (2012) Postnatal isoform switch and protein localization of LEF1 and TCF7L2 transcription factors in cortical, thalamic, and mesencephalic regions of the adult mouse brain. *Brain Struct.Funct.*

Nagarajan,L. (2010) Chromosomal deletions in AML. *Cancer Treat.Res.*, **145**, 59-66.

Nagasawa,T. (2006) Microenvironmental niches in the bone marrow required for B-cell development. *Nat.Rev.Immunol.*, **6**, 107-116.

Nagasawa,T. (2007) The chemokine CXCL12 and regulation of HSC and B lymphocyte development in the bone marrow niche. *Adv.Exp.Med.Biol.*, **602**, 69-75.

Nakahata,T., Gross,A.J., & Ogawa,M. (1982) A stochastic model of self-renewal and commitment to differentiation of the primitive hemopoietic stem cells in culture. *J.Cell Physiol*, **113**, 455-458.

Nakamura,Y., Ando,K., Chargui,J., Kawada,H., Sato,T., Tsuji,T. and others (1999) Ex vivo generation of CD34(+) cells from CD34(-) hematopoietic cells. *Blood*, **94**, 4053-4059.

Nathke,I. (2006) Cytoskeleton out of the cupboard: colon cancer and cytoskeletal changes induced by loss of APC. *Nat.Rev.Cancer*, **6**, 967-974.

Nazwar,T.A., Glassmann,A., & Schilling,K. (2009) Expression and molecular diversity of Tcf7l2 in the developing murine cerebellum and brain. *J.Neurosci.Res.*, **87**, 1532-1546.

Newland,A. (2002) Progress in the treatment of acute myeloid leukaemia in adults. *Int.J.Hematol.*, **76 Suppl 1**, 253-258.

Nguyen,H., Merrill,B.J., Polak,L., Nikolova,M., Rendl,M., Shaver,T.M. and others (2009) Tcf3 and Tcf4 are essential for long-term homeostasis of skin epithelia. *Nat.Genet.*, **41**, 1068-1075.

Novak,A. & Dedhar,S. (1999) Signaling through beta-catenin and Lef/Tcf. *Cell Mol.Life Sci.*, **56**, 523-537.

Novershtern,N., Subramanian,A., Lawton,L.N., Mak,R.H., Haining,W.N., McConkey,M.E. and others (2011) Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*, **144**, 296-309.

Ogawa,M., Matsuzaki,Y., Nishikawa,S., Hayashi,S., Kunisada,T., Sudo,T. and others (1991) Expression and function of c-kit in hemopoietic progenitor cells. *J.Exp.Med.*, **174**, 63-71.

Ogawa,M. (1993) Differentiation and proliferation of hematopoietic stem cells. *Blood*, **81**, 2844-2853.

Oh,I.H., Lau,A., & Eaves,C.J. (2000) During ontogeny primitive (CD34(+)CD38(-)) hematopoietic cells show altered expression of a subset of genes associated with early cytokine and differentiation responses of their adult counterparts. *Blood*, **96**, 4160-4168.

Oken,M.M., Creech,R.H., Tormey,D.C., Horton,J., Davis,T.E., McFadden,E.T. and others (1982) Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am.J.Clin.Oncol.*, **5**, 649-655.

Orkin,S.H. (1995) Hematopoiesis: how does it happen? *Curr.Opin.Cell Biol.*, **7**, 870-877.

Osawa,M., Hanada,K., Hamada,H., & Nakauchi,H. (1996) Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science*, **273**, 242-245.

Overton,W.R. (1988) Modified histogram subtraction technique for analysis of flow cytometry data. *Cytometry*, **9**, 619-626.

Pabst,T. & Mueller,B.U. (2007) Transcriptional dysregulation during myeloid transformation in AML. *Oncogene*, **26**, 6829-6837.

Paez,D., Gerger,A., Zhang,W., Yang,D., Labonte,M.J., Benhaim,L. and others (2013) Association of common gene variants in the WNT/beta-catenin pathway with colon cancer recurrence. *Pharmacogenomics.J.*

Park,C.Y., Kim,H.S., Jang,J., Lee,H., Lee,J.S., Yoo,J.E. and others (2013) ABCD2 is a direct target of beta-catenin and TCF-4: implications for X-linked adrenoleukodystrophy therapy. *PLoS.One.*, **8**, e56242.

Park,I.K., Qian,D., Kiel,M., Becker,M.W., Pihalja,M., Weissman,I.L. and others (2003) Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature*, **423**, 302-305.

Peng,S., Zhu,Y., Lu,B., Xu,F., Li,X., & Lai,M. (2013) TCF7L2 gene polymorphisms and type 2 diabetes risk: a comprehensive and updated meta-analysis involving 121,174 subjects. *Mutagenesis*, **28**, 25-37.

Penman,G.A., Leung,L., & Nathke,I.S. (2005) The adenomatous polyposis coli protein (APC) exists in two distinct soluble complexes with different functions. *J.Cell Sci.*, **118**, 4741-4750.

Petersen,C.P. & Reddien,P.W. (2009) Wnt signaling and the polarity of the primary body axis. *Cell*, **139**, 1056-1068.

Philippakis,A.A., Qureshi,A.M., Berger,M.F., & Bulyk,M.L. (2008) Design of compact, universal DNA microarrays for protein binding microarray experiments. *J.Comput.Biol.*, **15**, 655-665.

Philpott,N.J., Turner,A.J., Scopes,J., Westby,M., Marsh,J.C., Gordon-Smith,E.C. and others (1996) The use of 7-amino actinomycin D in identifying apoptosis: simplicity of use and broad spectrum of application compared with other techniques. *Blood*, **87**, 2244-2251.

Pinto,D. & Clevers,H. (2005) Wnt control of stem cells and differentiation in the intestinal epithelium. *Exp.Cell Res.*, **306**, 357-363.

Pise-Masison,C.A., Radonovich,M., Dohoney,K., Morris,J.C., O'Mahony,D., Lee,M.J. and others (2009) Gene expression profiling of ATL patients: compilation of disease-related genes and evidence for TCF4 involvement in BIRC5 gene expression and cell viability. *Blood*, **113**, 4016-4026.

Poy,F., Lepourcelet,M., Shivdasani,R.A., & Eck,M.J. (2001) Structure of a human Tcf4-beta-catenin complex. *Nat.Struct.Biol.*, **8**, 1053-1057.

Prokunina-Olsson,L., Welch,C., Hansson,O., Adhikari,N., Scott,L.J., Usher,N. and others (2009) Tissue-specific alternative splicing of TCF7L2. *Hum.Mol.Genet.*, **18**, 3795-3804.

Pukrop,T., Gradl,D., Henningfeld,K.A., Knochel,W., Wedlich,D., & Kuhl,M. (2001) Identification of two regulatory elements within the high mobility group box transcription factor XTCF-4. *J.Biol.Chem.*, **276**, 8968-8978.

Qiao,X., Zhang,H.H., Liu,Y., Todd,M.J., & Marron,J.S. (2010) Weighted Distance Weighted Discrimination and Its Asymptotic Properties. *J.Am.Stat.Assoc.*, **105**, 401-414.

Ramasamy,A., Mondry,A., Holmes,C.C., & Altman,D.G. (2008) Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS.Med.*, **5**, e184.

Rangatia,J., Vangala,R.K., Singh,S.M., Peer Zada,A.A., Elsasser,A., Kohlmann,A. and others (2003) Elevated c-Jun expression in acute myeloid leukemias inhibits C/EBPalpha DNA binding via leucine zipper domain interaction. *Oncogene*, **22**, 4760-4764.

Ravandi,F., Estey,E.H., Appelbaum,F.R., Lo-Coco,F., Schiffer,C.A., Larson,R.A. and others (2012) Gemtuzumab ozogamicin: time to resurrect? *J.Clin.Oncol.*, **30**, 3921-3923.

Reya,T. (2003) Regulation of hematopoietic stem cell self-renewal. *Recent Prog.Horm.Res.*, **58**, 283-295.

Reya,T. & Clevers,H. (2005) Wnt signalling in stem cells and cancer. *Nature*, **434**, 843-850.

Rhodes,D.R., Yu,J., Shanker,K., Deshpande,N., Varambally,R., Ghosh,D. and others (2004) Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc.Natl.Acad.Sci.U.S.A*, **101**, 9309-9314.

Ricciardi,M.R., Scerpa,M.C., Bergamo,P., Ciuffreda,L., Petrucci,M.T., Chiaretti,S. and others (2012) Therapeutic potential of MEK inhibition in acute myelogenous leukemia: rationale for "vertical" and "lateral" combination strategies. *J.Mol.Med.(Berl)*, **90**, 1133-1144.

Rocquain,J., Carbuccia,N., Trouplin,V., Raynaud,S., Murati,A., Nezri,M. and others (2010) Combined mutations of ASXL1, CBL, FLT3, IDH1, IDH2, JAK2, KRAS, NPM1, NRAS, RUNX1, TET2 and WT1 genes in myelodysplastic syndromes and acute myeloid leukemias. *BMC.Cancer*, **10**, 401.

Roman-Gomez,J., Jimenez-Velasco,A., Cordeu,L., Vilas-Zornoza,A., San Jose-Eneriz,E., Garate,L. and others (2007) WNT5A, a putative tumour suppressor of lymphoid malignancies, is inactivated by aberrant methylation in acute lymphoblastic leukaemia. *Eur.J.Cancer*, **43**, 2736-2746.

Roose,J., Huls,G., van,B.M., Moerer,P., van der Horn,K., Goldschmeding,R. and others (1999) Synergy between tumor suppressor APC and the beta-catenin-Tcf4 target Tcf1. *Science*, **285**, 1923-1926.

Rosin-Arbesfeld,R., Townsley,F., & Bienz,M. (2000) The APC tumour suppressor has a nuclear export function. *Nature*, **406**, 1009-1012.

Rowe,J.M. & Tallman,M.S. (2010) How I treat acute myeloid leukemia. *Blood*, **116**, 3147-3156.

Rozenveld-Geugien,M., Baas,I.O., van,G.D., Vellenga,E., & Schuringa,J.J. (2007) Expansion of normal and leukemic human hematopoietic stem/progenitor cells requires rac-mediated interaction with stromal cells. *Exp.Hematol.*, **35**, 782-792.

Rubinfeld,B., Albert,I., Porfiri,E., Fiol,C., Munemitsu,S., & Polakis,P. (1996) Binding of GSK3beta to the APC-beta-catenin complex and regulation of complex assembly. *Science*, **272**, 1023-1026.

Saneyoshi,T., Kume,S., Amasaki,Y., & Mikoshiba,K. (2002) The Wnt/calcium pathway activates NF-AT and promotes ventral cell fate in Xenopus embryos. *Nature*, **417**, 295-299.

Sarrazin,S. & Sieweke,M. (2011) Integration of cytokine and transcription factor signals in hematopoietic stem cell commitment. *Semin.Immunol.*, **23**, 326-334.

Sato,T., Yang,X., Knapper,S., White,P., Smith,B.D., Galkin,S. and others (2011) FLT3 ligand impedes the efficacy of FLT3 inhibitors in vitro and in vivo. *Blood*, **117**, 3286-3293.

Schaub,F.X., Looser,R., Li,S., Hao-Shen,H., Lehmann,T., Tichelli,A. and others (2010) Clonal analysis of TET2 and JAK2 mutations suggests that TET2 can be a late event in the progression of myeloproliferative neoplasms. *Blood*, **115**, 2003-2007.

Schmid,I., Krall,W.J., Uittenbogaart,C.H., Braun,J., & Giorgi,J.V. (1992) Dead cell discrimination with 7-amino-actinomycin D in combination with dual color immunofluorescence in single laser flow cytometry. *Cytometry*, **13**, 204-208.

Schneider,F., Hoster,E., Unterhalt,M., Schneider,S., Dufour,A., Benthaus,T. and others (2012) The FLT3ITD mRNA level has a high prognostic impact in NPM1 mutated, but not in NPM1 unmutated, AML with a normal karyotype. *Blood*, **119**, 4383-4386.

Schoch,C., Kohlmann,A., Schnittger,S., Brors,B., Dugas,M., Mergenthaler,S. and others (2002) Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. *Proc.Natl.Acad.Sci.U.S.A*, **99**, 10008-10013.

Schoch,C., Kern,W., Kohlmann,A., Hiddemann,W., Schnittger,S., & Haferlach,T. (2005) Acute myeloid leukemia with a complex aberrant karyotype is a distinct biological entity characterized by genomic imbalances and a specific gene expression profile. *Genes Chromosomes.Cancer*, **43**, 227-238.

Scott,E.W., Fisher,R.C., Olson,M.C., Kehrli,E.W., Simon,M.C., & Singh,H. (1997) PU.1 functions in a cell-autonomous manner to control the differentiation of multipotential lymphoid-myeloid progenitors. *Immunity.*, **6**, 437-447.

Seeling,J.M., Miller,J.R., Gil,R., Moon,R.T., White,R., & Virshup,D.M. (1999) Regulation of beta-catenin signaling by the B56 subunit of protein phosphatase 2A. *Science*, **283**, 2089-2091.

Sengupta,A., Banerjee,D., Chandra,S., Banerji,S.K., Ghosh,R., Roy,R. and others (2007) Deregulation and cross talk among Sonic hedgehog, Wnt, Hox and Notch signaling in chronic myeloid leukemia progression. *Leukemia*, **21**, 949-955.

Seo,J. & Hoffman,E.P. (2006) Probe set algorithms: is there a rational best bet? *BMC.Bioinformatics.*, **7**, 395.

Shah,A., Andersson,T.M., Rachet,B., Bjorkholm,M., & Lambert,P.C. (2013) Survival and cure of acute myeloid leukaemia in England, 1971-2006: a population-based study. *Br.J.Haematol.*, **162**, 509-516.

Sharma,S., Gurudutta,G.U., Satija,N.K., Pati,S., Afrin,F., Gupta,P. and others (2006) Stem cell c-KIT and HOXB4 genes: critical roles and mechanisms in self-renewal, proliferation, and differentiation. *Stem Cells Dev.*, **15**, 755-778.

Shiina,H., Igawa,M., Breault,J., Ribeiro-Filho,L., Pookot,D., Urakami,S. and others (2003) The human T-cell factor-4 gene splicing isoforms, Wnt signal pathway, and apoptosis in renal cell carcinoma. *Clin.Cancer Res.*, **9**, 2121-2132.

Shitashige,M., Naishiro,Y., Idogawa,M., Honda,K., Ono,M., Hirohashi,S. and others (2007) Involvement of splicing factor-1 in beta-catenin/T-cell factor-4-mediated gene transactivation and pre-mRNA splicing. *Gastroenterology*, **132**, 1039-1054.

Silva,M., Grillot,D., Benito,A., Richard,C., Nunez,G., & Fernandez-Luna,J.L. (1996) Erythropoietin can promote erythroid progenitor survival by repressing apoptosis through Bcl-XL and Bcl-2. *Blood*, **88**, 1576-1582.

Simon,M., Grandage,V.L., Linch,D.C., & Khwaja,A. (2005) Constitutive activation of the Wnt/beta-catenin signalling pathway in acute myeloid leukaemia. *Oncogene*, **24**, 2410-2420.

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

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

Sparks,A.B., Morin,P.J., Vogelstein,B., & Kinzler,K.W. (1998) Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. *Cancer Res.*, **58**, 1130-1134.

Spiekermann,K., Pau,M., Schwab,R., Schmieja,K., Franzrahe,S., & Hiddemann,W. (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. *Exp.Hematol.*, **30**, 262-271.

St Amand,A.L. & Klymkowsky,M.W. (2001) Cadherins and catenins, Wnts and SOXs: embryonic patterning in Xenopus. *Int.Rev.Cytol.*, **203**, 291-355.

Staal,F.J. & Luis,T.C. (2010) Wnt signaling in hematopoiesis: crucial factors for self-renewal, proliferation, and cell fate decisions. *J.Cell Biochem.*, **109**, 844-849.

Stone,R.M. & Mayer,R.J. (1993) Treatment of the newly diagnosed adult with de novo acute myeloid leukemia. *Hematol.Oncol.Clin.North Am.*, **7**, 47-64.

Struewing,I., Boyechko,T., Barnett,C., Beildeck,M., Byers,S.W., & Mao,C.D. (2010) The balance of TCF7L2 variants with differential activities in Wnt-signaling is regulated by lithium in a GSK3beta-independent manner. *Biochem.Biophys.Res.Commun.*, **399**, 245-250.

Suda,T., Suda,J., & Ogawa,M. (1983) Proliferative kinetics and differentiation of murine blast cell colonies in culture: evidence for variable G0 periods and constant doubling rates of early pluripotent hemopoietic progenitors. *J.Cell Physiol*, **117**, 308-318.

Sugiyama,T., Kohara,H., Noda,M., & Nagasawa,T. (2006) Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity.*, **25**, 977-988.

Sundstrom,C. & Nilsson,K. (1976) Establishment and characterization of a human histiocytic lymphoma cell line (U-937). *Int.J.Cancer*, **17**, 565-577.

Sutherland,H.J., Blair,A., & Zapf,R.W. (1996) Characterization of a hierarchy in human acute myeloid leukemia progenitor cells. *Blood*, **87**, 4754-4761.

Suzuki,R., Onizuka,M., Kojima,M., Shimada,M., Fukagawa,S., Tsuboi,K. and others (2007) Preferential hypermethylation of the Dickkopf-1 promoter in core-binding factor leukaemia. *Br.J.Haematol.*, **138**, 624-631.

Tabachnick,B.G. & Fidell,L.S. (2007) *Using multivariate statistics.,* 5 edn, Allyn and Bacon, Boston.

Tallman,M.S., Andersen,J.W., Schiffer,C.A., Appelbaum,F.R., Feusner,J.H., Ogden,A. and others (1997) All-trans-retinoic acid in acute promyelocytic leukemia. *N.Engl.J.Med.*, **337**, 1021-1028.

Tallman,M.S., Andersen,J.W., Schiffer,C.A., Appelbaum,F.R., Feusner,J.H., Woods,W.G. and others (2002) All-trans retinoic acid in acute promyelocytic leukemia: long-term outcome and prognostic factor analysis from the North American Intergroup protocol. *Blood*, **100**, 4298-4302.

Tallman,M.S., Gilliland,D.G., & Rowe,J.M. (2005) Drug therapy for acute myeloid leukemia. *Blood*, **106**, 1154-1163.

Tamhane,A.C. & Dunlop,D.D. (2000) Statistics and Data Analysis from Elementary to Intermediate., pp. 473-474. Prentice Hall.

Tang,W., Dodge,M., Gundapaneni,D., Michnoff,C., Roth,M., & Lum,L. (2008) A genome-wide RNAi screen for Wnt/beta-catenin pathway components identifies unexpected roles for TCF transcription factors in cancer. *Proc.Natl.Acad.Sci.U.S.A*, **105**, 9697-9702.

Tenen,D.G., Hromas,R., Licht,J.D., & Zhang,D.E. (1997) Transcription factors, normal myeloid development, and leukemia. *Blood*, **90**, 489-519.

Tenen,D.G. (2003) Disruption of differentiation in human cancer: AML shows the way. *Nat.Rev.Cancer*, **3**, 89-101.

Terpstra,W., Prins,A., Ploemacher,R.E., Wognum,B.W., Wagemaker,G., Lowenberg,B. and others (1996) Long-term leukemia-initiating capacity of a CD34-subpopulation of acute myeloid leukemia. *Blood*, **87**, 2187-2194.

Tetsu,O. & McCormick,F. (1999) Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature*, **398**, 422-426.

Theilgaard-Monch,K., Boultwood,J., Ferrari,S., Giannopoulos,K., Hernandez-Rivas,J.M., Kohlmann,A. and others (2011) Gene expression profiling in MDS and AML: potential and future avenues. *Leukemia*, **25**, 909-920.

Tian,W., Han,X., Yan,M., Xu,Y., Duggineni,S., Lin,N. and others (2012) Structure-based discovery of a novel inhibitor targeting the beta-catenin/Tcf4 interaction. *Biochemistry*, **51**, 724-731.

Tickenbrock,L., Schwable,J., Wiedehage,M., Steffen,B., Sargin,B., Choudhary,C. and others (2005) Flt3 tandem duplication mutations cooperate with Wnt signaling in leukemic signal transduction. *Blood*, **105**, 3699-3706.

Till,J.E. & McCulloch,E.A. (1961) A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat.Res.*, **14**, 213-222.

Tonks,A., Tonks,A.J., Pearn,L., Mohamad,Z., Burnett,A.K., & Darley,R.L. (2005) Optimized retroviral transduction protocol which preserves the primitive subpopulation of human hematopoietic cells. *Biotechnol.Prog.*, **21**, 953-958.

Tonks,A., Hills,R., White,P., Rosie,B., Mills,K.I., Burnett,A.K. and others (2007a) CD200 as a prognostic factor in acute myeloid leukaemia. *Leukemia*, **21**, 566-568.

Tonks,A., Pearn,L., Musson,M., Gilkes,A., Mills,K.I., Burnett,A.K. and others (2007b) Transcriptional dysregulation mediated by RUNX1-RUNX1T1 in normal human progenitor cells and in acute myeloid leukaemia. *Leukemia*, **21**, 2495-2505.

Trompouki,E., Bowman,T.V., Lawton,L.N., Fan,Z.P., Wu,D.C., DiBiase,A. and others (2011) Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell*, **147**, 577-589.

Tsai,M.S., Hwang,S.M., Chen,K.D., Lee,Y.S., Hsu,L.W., Chang,Y.J. and others (2007) Functional network analysis of the transcriptomes of mesenchymal stem cells derived from amniotic fluid, amniotic membrane, cord blood, and bone marrow. *Stem Cells*, **25**, 2511-2523.

Tsedensodnom,O., Koga,H., Rosenberg,S.A., Nambotin,S.B., Carroll,J.J., Wands,J.R. and others (2011) Identification of T-cell factor-4 isoforms that contribute to the malignant phenotype of hepatocellular carcinoma cells. *Exp.Cell Res.*, **317**, 920-931.

Tsuchiya,S., Kobayashi,Y., Goto,Y., Okumura,H., Nakae,S., Konno,T. and others (1982) Induction of maturation in cultured human monocytic leukemia cells by a phorbol diester. *Cancer Res.*, **42**, 1530-1536.

Tsuji,K. & Nakahata,T. (1989) Stochastic model for multipotent hemopoietic progenitor differentiation. *J.Cell Physiol*, **139**, 647-653.

Tsutsui,J., Moriyama,M., Arima,N., Ohtsubo,H., Tanaka,H., & Ozawa,M. (1996) Expression of cadherin-catenin complexes in human leukemia cell lines. *J.Biochem.*, **120**, 1034-1039.

Ungewickell,A. & Medeiros,B.C. (2012) Novel agents in acute myeloid leukemia. *Int.J.Hematol.*, **96**, 178-185.

Vacik,T., Stubbs,J.L., & Lemke,G. (2011) A novel mechanism for the transcriptional regulation of Wnt signaling in development. *Genes Dev.*, **25**, 1783-1795.

Valencia,A., Roman-Gomez,J., Cervera,J., Such,E., Barragan,E., Bolufer,P. and others (2009) Wnt signaling pathway is epigenetically regulated by methylation of Wnt antagonists in acute myeloid leukemia. *Leukemia*, **23**, 1658-1666.

Valenta,T., Lukas,J., & Korinek,V. (2003) HMG box transcription factor TCF-4's interaction with CtBP1 controls the expression of the Wnt target Axin2/Conductin in human embryonic kidney cells. *Nucleic Acids Res.*, **31**, 2369-2380.

Valk,P.J., Verhaak,R.G., Beijen,M.A., Erpelinck,C.A., Barjesteh van Waalwijk van Doorn-Khosrovani, Boer,J.M. and others (2004) Prognostically useful gene-expression profiles in acute myeloid leukemia. *N.Engl.J.Med.*, **350**, 1617-1628.

van de Wetering,M., Sancho,E., Verweij,C., de,L.W., Oving,I., Hurlstone,A. and others (2002) The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell*, **111**, 241-250.

Van Den Berg,D.J., Sharma,A.K., Bruno,E., & Hoffman,R. (1998) Role of members of the Wnt gene family in human hematopoiesis. *Blood*, **92**, 3189-3202.

van der Flier,L.G., Sabates-Bellver,J., Oving,I., Haegebarth,A., De,P.M., Anti,M. and others (2007) The Intestinal Wnt/TCF Signature. *Gastroenterology*, **132**, 628-632.

van Es,J.H., Giles,R.H., & Clevers,H.C. (2001) The many faces of the tumor suppressor gene APC. *Exp.Cell Res.*, **264**, 126-134.

van Es,J.H., Haegebarth,A., Kujala,P., Itzkovitz,S., Koo,B.K., Boj,S.F. and others (2012) A critical role for the Wnt effector Tcf4 in adult intestinal homeostatic self-renewal. *Mol.Cell Biol.*, **32**, 1918-1927.

van,N.M. & Clevers,H. (2002) TCF transcription factors, mediators of Wnt-signaling in development and cancer. *Dev.Biol.*, **244**, 1-8.

Vardiman,J.W., Harris,N.L., & Brunning,R.D. (2002) The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood*, **100**, 2292-2302.

Verbeek,S., Izon,D., Hofhuis,F., Robanus-Maandag,E., te,R.H., van de Wetering,M. and others (1995) An HMG-box-containing T-cell factor required for thymocyte differentiation. *Nature*, **374**, 70-74.

Verhaak,R.G., Wouters,B.J., Erpelinck,C.A., Abbas,S., Beverloo,H.B., Lugthart,S. and others (2009) Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica*, **94**, 131-134.

Wallmen,B., Schrempp,M., & Hecht,A. (2012) Intrinsic properties of Tcf1 and Tcf4 splice variants determine cell-type-specific Wnt/beta-catenin target gene expression. *Nucleic Acids Res.*, **40**, 9455-9469.

Waltzer,L. & Bienz,M. (1998) Drosophila CBP represses the transcription factor TCF to antagonize Wingless signalling. *Nature*, **395**, 521-525.

Wang,S.J., Wong,G., de Heer,A.M., Xia,W., & Bourguignon,L.Y. (2009) CD44 variant isoforms in head and neck squamous cell carcinoma progression. *Laryngoscope*, **119**, 1518-1530.

Wang,Y., Krivtsov,A.V., Sinha,A.U., North,T.E., Goessling,W., Feng,Z. and others (2010) The Wnt/beta-catenin pathway is required for the development of leukemia stem cells in AML. *Science*, **327**, 1650-1653.

Warnat,P., Eils,R., & Brors,B. (2005) Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes. *BMC.Bioinformatics.*, **6**, 265.

Waterman,M.L. (2002) Expression of lymphoid enhancer factor/T-cell factor proteins in colon cancer. *Curr.Opin.Gastroenterol.*, **18**, 53-59.

Weise,A., Bruser,K., Elfert,S., Wallmen,B., Wittel,Y., Wohrle,S. and others (2010) Alternative splicing of Tcf7l2 transcripts generates protein variants with differential promoter-binding and transcriptional activation properties at Wnt/beta-catenin targets. *Nucleic Acids Res.*, **38**, 1964-1981.

Welch,J.S., Ley,T.J., Link,D.C., Miller,C.A., Larson,D.E., Koboldt,D.C. and others (2012) The origin and evolution of mutations in acute myeloid leukemia. *Cell*, **150**, 264-278.

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

Willert,K. & Jones,K.A. (2006) Wnt signaling: is the party in the nucleus? *Genes Dev.*, **20**, 1394-1404.

Wilson,A., Laurenti,E., Oser,G., van der Wath,R.C., Blanco-Bose,W., Jaworski,M. and others (2008) Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell*, **135**, 1118-1129.

Wilson,A. & Trumpp,A. (2006) Bone-marrow haematopoietic-stem-cell niches. *Nat.Rev.Immunol.*, **6**, 93-106.

Wouters,B.J., Jorda,M.A., Keeshan,K., Louwers,I., Erpelinck-Verschueren,C.A., Tielemans,D. and others (2007) Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood*, **110**, 3706-3714.

Wright,M., Aikawa,M., Szeto,W., & Papkoff,J. (1999) Identification of a Wnt-responsive signal transduction pathway in primary endothelial cells. *Biochem.Biophys.Res.Commun.*, **263**, 384-388.

Wu,X., Tu,X., Joeng,K.S., Hilton,M.J., Williams,D.A., & Long,F. (2008) Rac1 activation controls nuclear localization of beta-catenin during canonical Wnt signaling. *Cell*, **133**, 340-353.

Xie,J., Xiang,D.B., Wang,H., Zhao,C., Chen,J., Xiong,F. and others (2012) Inhibition of Tcf-4 induces apoptosis and enhances chemosensitivity of colon cancer cells. *PLoS.One.*, **7**, e45617.

Xu,J., Suzuki,M., Niwa,Y., Hiraga,J., Nagasaka,T., Ito,M. and others (2008) Clinical significance of nuclear non-phosphorylated beta-catenin in acute myeloid leukaemia and myelodysplastic syndrome. *Br.J.Haematol.*, **140**, 394-401.

Yamamoto,H., Ihara,M., Matsuura,Y., & Kikuchi,A. (2003) Sumoylation is involved in beta-catenin-dependent activation of Tcf-4. *EMBO J.*, **22**, 2047-2059.

Yan,M., Ahn,E.Y., Hiebert,S.W., & Zhang,D.E. (2009) RUNX1/AML1 DNA-binding domain and ETO/MTG8 NHR2-dimerization domain are critical to AML1-ETO9a leukemogenesis. *Blood*, **113**, 883-886.

Yeoh,E.J., Ross,M.E., Shurtleff,S.A., Williams,W.K., Patel,D., Mahfouz,R. and others (2002) Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*, **1**, 133-143.

Young,R.M., Reyes,A.E., & Allende,M.L. (2002) Expression and splice variant analysis of the zebrafish tcf4 transcription factor. *Mech.Dev.*, **117**, 269-273.

Ysebaert,L., Chicanne,G., Demur,C., De,T.F., Prade-Houdellier,N., Ruidavets,J.B. and others (2006) Expression of beta-catenin by acute myeloid leukemia cells predicts enhanced clonogenic capacities and poor prognosis. *Leukemia*, **20**, 1211-1216.

Zeilstra,J., Joosten,S.P., van,A.H., Tolg,C., Berns,A., Snoek,M. and others (2013) Stem cell CD44v isoforms promote intestinal cancer formation in Apc(min) mice downstream of Wnt signaling. *Oncogene*.

Zhang,C., Bao,W., Rong,Y., Yang,H., Bowers,K., Yeung,E. and others (2013) Genetic variants and the risk of gestational diabetes mellitus: a systematic review. *Hum.Reprod.Update.*, **19**, 376-390.

Zhu,J. & Emerson,S.G. (2002) Hematopoietic cytokines, transcription factors and lineage commitment. *Oncogene*, **21**, 3295-3313.

Zufferey,R., Dull,T., Mandel,R.J., Bukovsky,A., Quiroz,D., Naldini,L. and others (1998) Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J.Virol.*, **72**, 9873-9880.

8 – Supplementary

Trial	Regime	Treatment
AML10		
	C1 DAT 3+10	Daunorubicin 50 mg/m^2 d1,3,5; Cytarabine 100 mg/m^2 d1-10 every 12h Thioguanine 100 mg/m^2 d-10 every 12h
	C2 DAT 3+8	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-8 every 12h Thioguanine 100 mg/m^2 d-8 every 12h
	C1 ADE 10+3+5	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-10 every 12h Etoposide 100 mg/m^2 d1-5
	C2 ADE 8+3+5	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-8 every 12h Etoposide 100 mg/m^2 d1-5
AML11		
	C1 DAT 3+10	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-10 every 12h Thioguanine 100 mg/m^2 d1-10 every 12h
	C2 DAT 2+5	Daunorubicin 50 mg/m^2 d1,3 Cytarabine 100 mg/m^2 d1-5 every 12h Thioguanine 100 mg/m^2 d-5 every 12h
	C1 ADE 10+3+5	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-10 every 12h Etoposide 100 mg/m^2 d1-5
	C2 ADE 5+2+5	Daunorubicin 50 mg/m^2 d1,3 Cytarabine 100 mg/m^2 d1-5 every 12h Etoposide 100 mg/m^2 d1-5
	C1 MAC 3+5	Mitozantrone 12mg/m^2 d1-3 Cytarabine 100 mg/m^2 bd, d1-5
	C2 MAC 2+5	Mitozantrone 12mg/m^2 d1,3 Cytarabine 100 mg/m^2 bd, d1-5

AML12:
Pre-amendment

C1 ADE 10+3+5	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-10 every 12h Etoposide 100 mg/m^2 d1-5
C2 ADE 8+3+5	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-8 every 12h Etoposide 100 mg/m^2 d1-5
C1 MAE 10+3+5	Mitoxantrone 12 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-10 every 12h Etoposide 100 mg/m^2 d1-5
C2 MAE 8+3+5	Mitoxantrone 12 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-8 every 12h Etoposide 100 mg/m^2 d1-5

Post-amendment

C1 H-DAT 3+10	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 200 mg/m^2 d1-10 every 12h Thioguanine 100 mg/m^2 d-10 every 12h
C2 H-DAT 3+8	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 200 mg/m^2 d1-8 every 12h Thioguanine 100 mg/m^2 d-8 every 12h
C1 S-DAT 3+10	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-10 every 12h Thioguanine 100 mg/m^2 d-10 every 12h
C2 S-DAT 3+8	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-8 every 12h Thioguanine 100 mg/m^2 d-8 every 12h

AML14:

C1 S-DAT 3+10	Daunorubicin 35/50 mg/m^2 d1-3 Cytarabine 100/200 mg/m^2 d1-10 every 12h Thioguanine 100 mg/m^2 d-10 every 12h
C1 S-DAT 3+8	Daunorubicin 35/50 mg/m^2 d1-3 Cytarabine 100/200 mg/m^2 d1-8 every 12h Thioguanine 100 mg/m^2 d-8 every 12h

AML15:		
	C1 DA 3+10	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-10 every 12h
	C2 DA 3+8	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-8 every 12h
	C1ADE 10+3+5	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-10 every 12h Etoposide 100 mg/m^2 d1-5
	C2 ADE 8+3+5	Daunorubicin 50 mg/m^2 d1,3,5 Cytarabine 100 mg/m^2 d1-8 every 12h Etoposide 100 mg/m^2 d1-5
	C1,2 FLAG-Ida	Fludarabine 30mg/m^2 iv days 2-6 inclusive, Cytosine Arabinoside 2g/m^2 over 4 hours starting 4 hours after Fludarabine on days 2-6, G-CSF [Lenograstim 263µg (1 vial)] SC daily days 1-7]

Supplementary 3.1: The specific intensive chemotherapy regimens used in UK MRC-NCRI AML Trials 10-15 relevant to patient samples used in this study are detailed.

Risk Group	Abnormality	Comment
Favorable	t(8;21)	Whether alone or in conjunction with other abnormalities.
	t(15;17)	
	inv(16)	
Intermediate	Normal	ie, Cytogenetic abnormalities not classified as favorable or adverse.
	8	
	21	
	22	Lack of additional favorable or adverse cytogenetic changes.
	del(7q)	
	del(9q)	
	Abnormal 11q23	
	All other structural/numerical abnormalities	
Adverse	5	Whether alone or in conjunction with intermediate-risk or other adverse-risk abnormalities.
	7	
	del(5q)	
	Abnormal 3q	
	Complex	

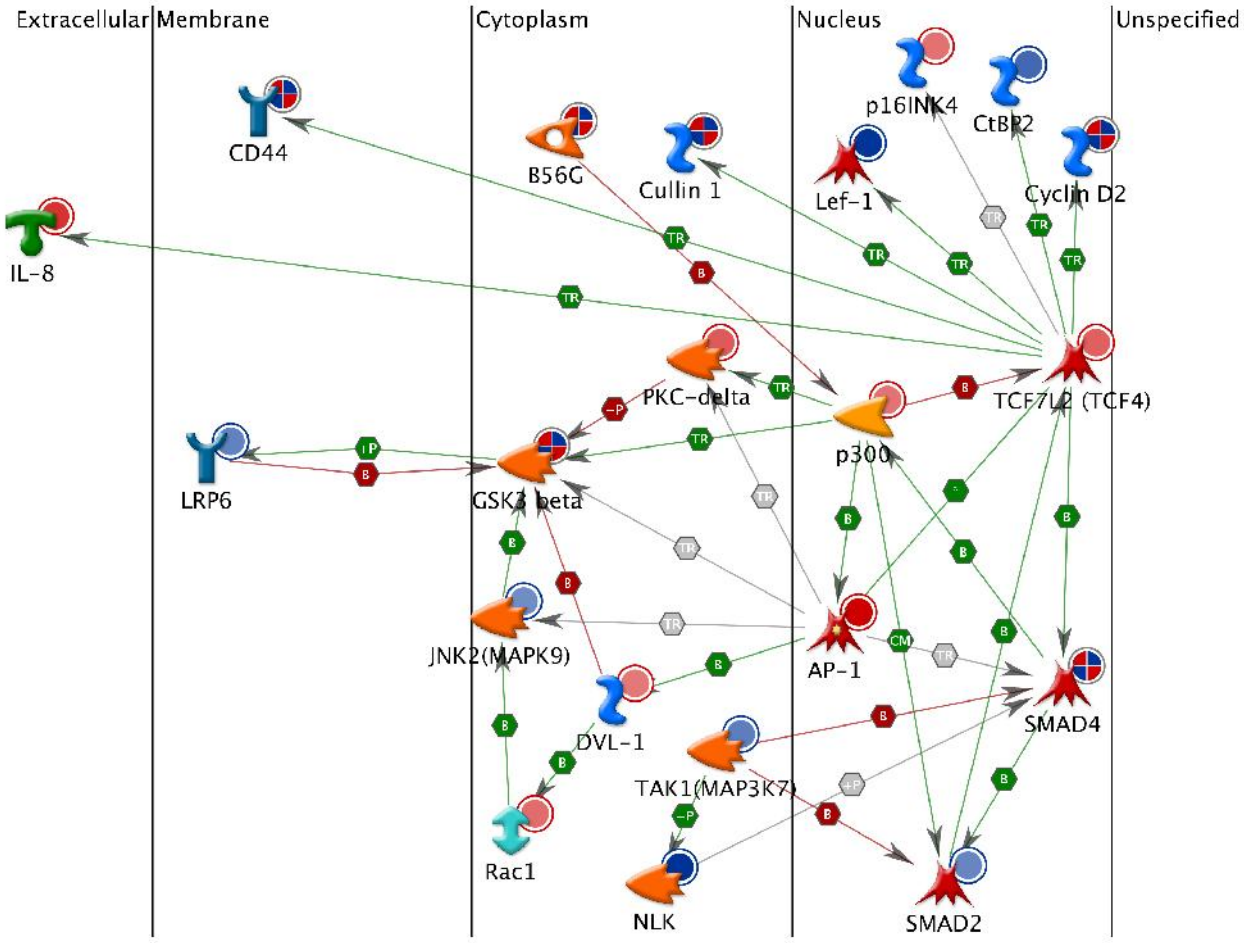
Supplementary 3.2: Cytogenetic prognostic classification based on established guidelines published by (Grimwade *et al*, 1998).

(Please refer electronic document on CD)

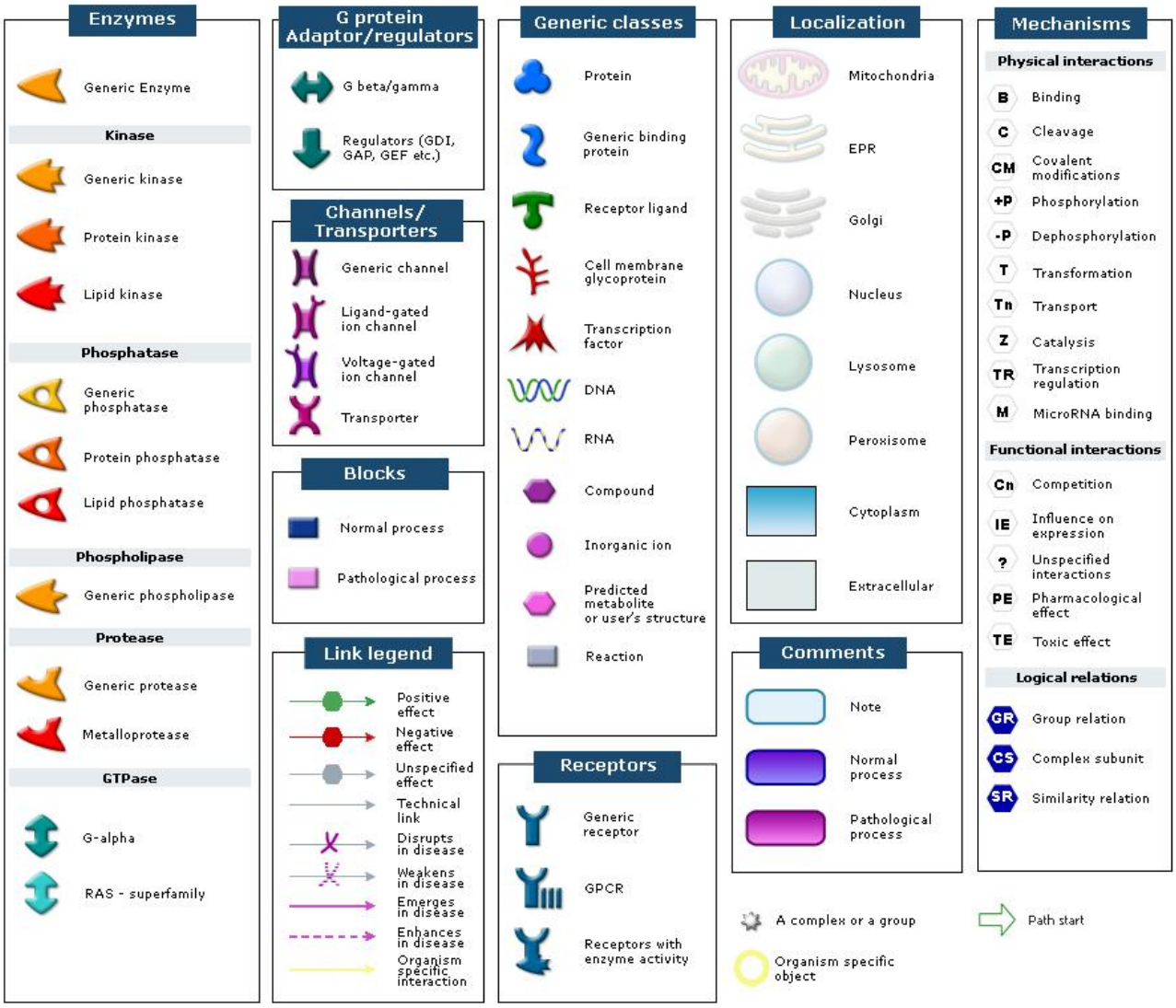
Supplementary 3.3: List of 651 WNT related probe-sets for HG-U133 Plus2.0

Score	WHO /ECOG Performance Status
0	Fully active, able to carry on all pre-disease performance without
	restriction
1	Restricted in physically strenuous activity but ambulatory and able to
	carry out work of a light or sedentary nature, e.g. light house work,
	office work
2	Ambulatory and capable of all selfcare but unable to carry out any work
	activities. Up and about more than 50% of waking hours
3	Capable of only limited selfcare, confined to bed or chair more than
	50% of waking hours
4	Completely disabled. Cannot carry on any selfcare. Totally confined to
	bed or chair

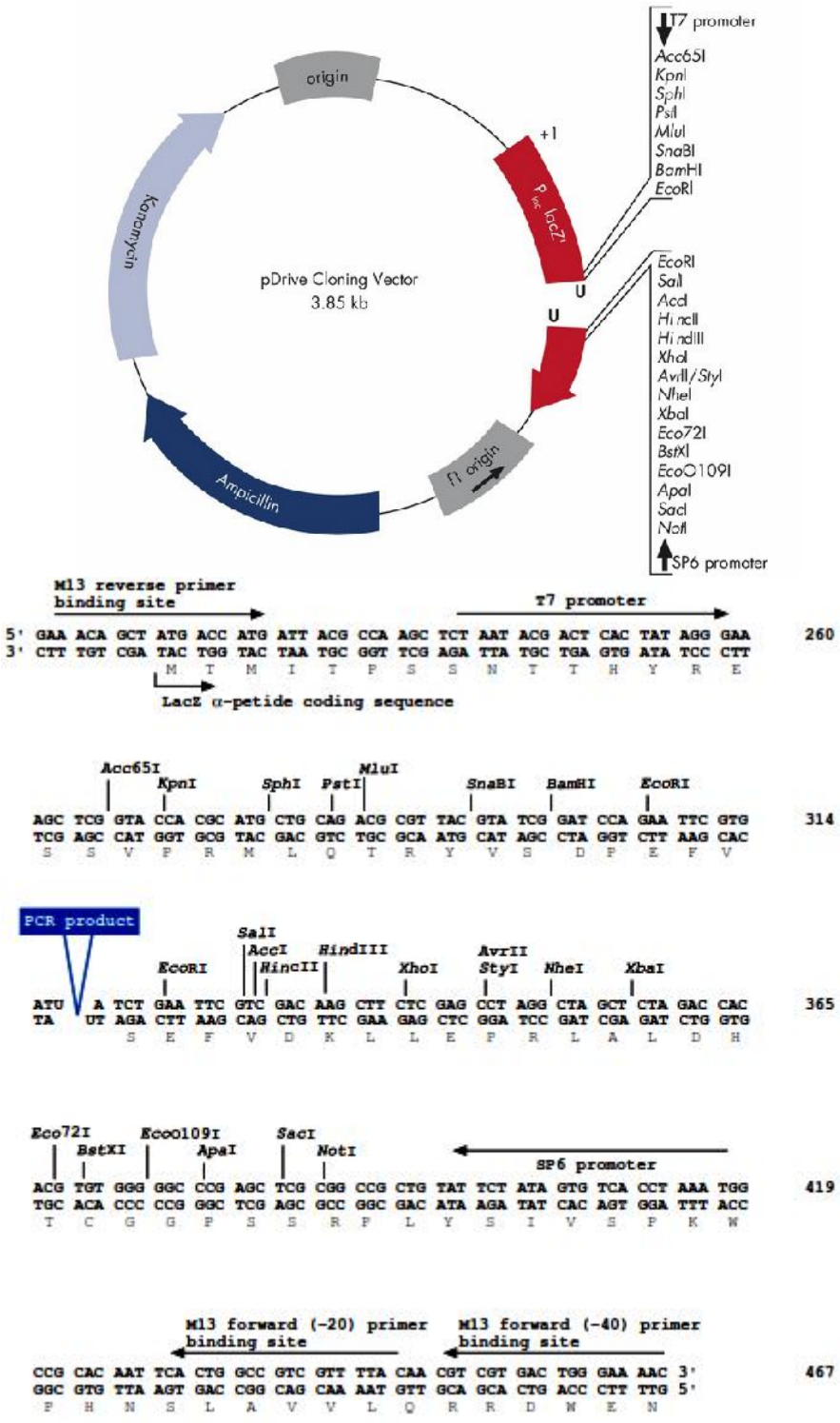
Supplementary 3.4: WHO or the Eastern Cooperative Oncology Group (ECOG) performance score. Published by Oken *et al*,1982.



Supplementary 3.5: Representative diagram showing Direct Interaction Network built for WNT Signalling genes taken from genelist F. All proteins shown are differentially expressed in AML (M0-M6) *vs* Normal Controls. Blue for underexpression and red for overexpression of corresponding gene. Localisation of each gene/protein and the hubs of its interconnection to other genes/proteins are shown. The full legend explaining identity of each protein and objects are listed in Supplementary 3.6 .



Supplementary 3.6: Legend showing identity of each objects mapped in Direct Interaction Network. Provided by GeneGo Metcore™.

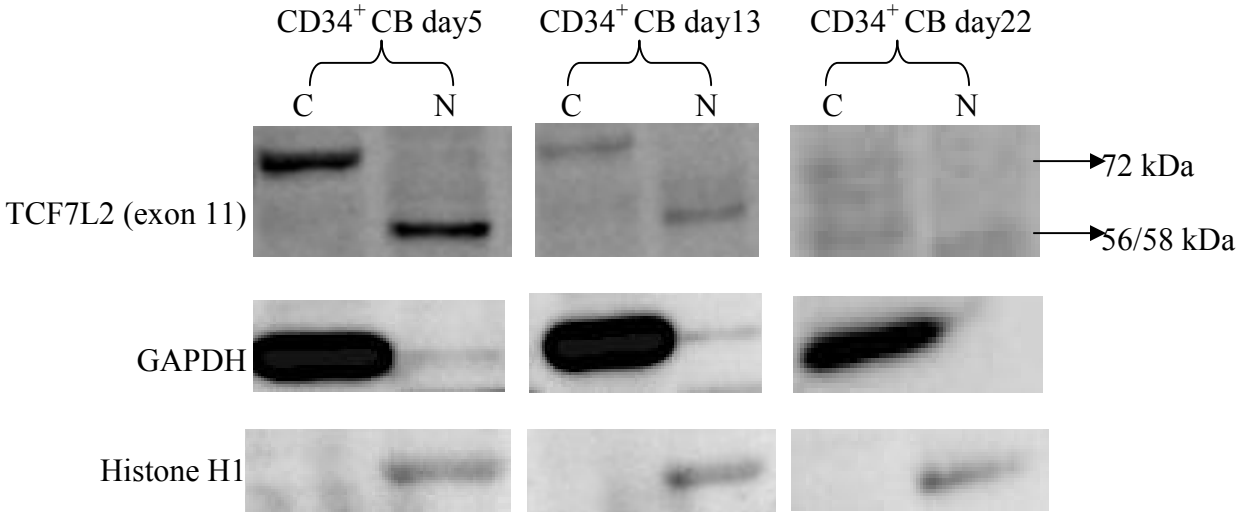


Supplementary 4.1: pDrive cloning vector map. The unique restriction endonuclease recognition sites on either side of the cloning site are listed. The positions of the T7 and SP6 promoter sites and the M13 forward and reverse sequencing primer binding sites are provided. Adapted from QIAGEN PCR Cloning Handbook.

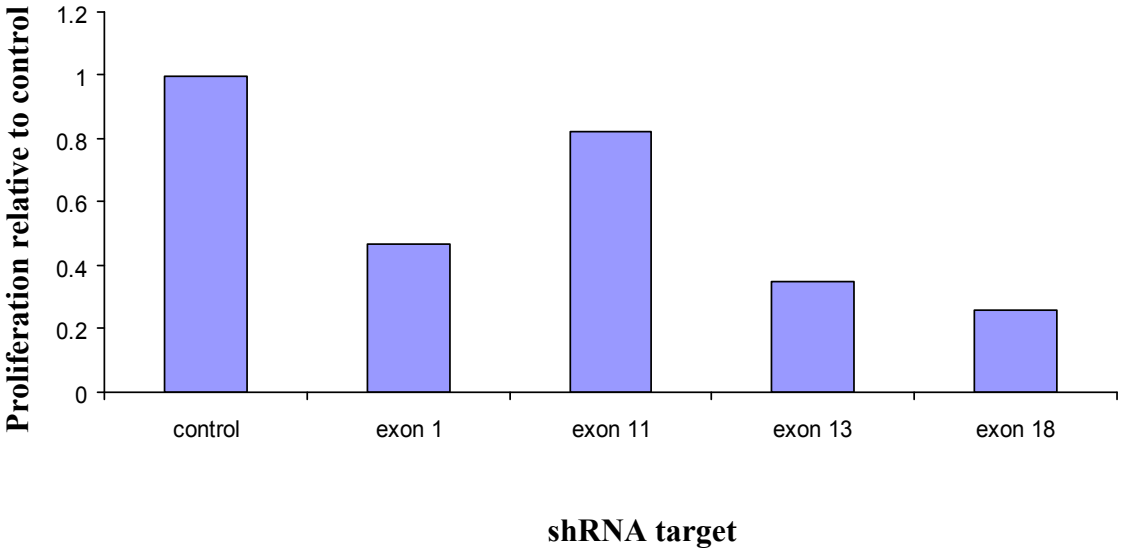
Exon number	Sequences 5’ to 3’
Exon 1	AATAATCTCCGCTCCCAGACTACTCCGTTCCTCCGGATTTCGATCCCCCTTTTTCTATCTGTCAATCAGC GCCGCCTTTGAACTGAAAAGCTCTCAGTCTAACTTCAACTCACTCAAATCCGAGCGGCACGAGCACCTCC TGTATCTTCGGCTTCCCCCCCCCTTTGCTCTTTATATCTGACTTCTTGTTGTTGTTGGTGTTTTTTTTTT TTTTACCCCCCTTTTTTATTTATTATTTTTTTGCACATTGATCGGATCCTTGGGAACGAGAGAAAAAAGA AACCCAAACTCACGCGTGCAGAAGATCTCCCCCCCCTTCCCCTCCCCTCCTCCCTCTTTTCCCCTCCCCA GGAGAAAAAGACCCCCAAGCAGAAAAAAGTTCACCTTGGACTCGTCTTTTTCTTGCAATATTTTTTGGGGG GGGCAAAACTTTTTGGGGGTGATTTTTTTTGGCTTTTCTTCCTCCTTCATTTTTCTTCCAAAATTGCTGC TGGTGGGTGAAAAAAAAATGCCGCAGCTGAACGGCGGTGGAGGGGATGACCTAGGCGCCAACGACGAACT GATTTCCTTCAAAGACGAGGGCGAACAGGAGGAGAAGAGCTCCGAAAACTCCTCGGCAGAGAGGGATTTA GCTGATGTCAAATCGTCTCTAGTCAATGAATCAGAAACGAATCAAAACAGCTCCTCCGATTCCGAG
Exon 2	GCGGAAAGACGGCCTCCGCCTCGCTCCGAAAGTTTCCGAGACAAATCCCGGGAAAGTTTGGAAGAAG
Exon 3	CGGCCAAGAGGCAAGATGGAGGGCTCTTTAAGGGGCCACCGTATCCCGGCTACCCCTTCATCATGATCCC CGACCTGACGAGCCCCTACCTCCCCAACGGATCGCTCTCGCCCACCGCCCGAACC
Exon 4	CTCCATTTTCAGTCCGGCAGCACACATTACTCTGCGTACAAAACGATTGAACACCAGATTGCAGTTCAG
Exon 5	TATCTCCAGATGAAATGGCCACTGCTTGATGTCCAGGCAGGGAGCCTCCAGAGTAGACAAGCCCTCAAGG ATGCCCGGTCCCCATCACCGGCACACATTGTC
Exon 6	AGCCCCCTCCCTTGCTGCACTCAGGGACATGACTGTCAGCACTTCTACCCCCCCTCAGACTTCACTGTCA GCACTCAAGTCTTCAGGGACATGAAAAGGAGCCACTCCTTACAAAAAGTTGGGGAGCCCTGGTGTATTGAG
Exon 7	TCTAACAAAGTGCCAGTGGTGCAGCACCCTCACCATGTCCACCCCCTCACGCCTCTTATCACGTACAGCA ATGAACACTTCACGCCGGGAAACCCACCTCCACACTTACCAGCCGACGTAGACCCCAAAACAG
Exon 8 b	GAATCCCACGGCCTCCGCACCCTCCAGATATATCCCCGTATTACCCACTATCGCCTGGCACCGTAGGACA AATCCCCCATCCGCTAGGATGGTTAGTACCACA
Exon 8 a	GAATCCCACGGCCTCCGCACCCTCCAGATATATCCCCGTATTACCCACTATCGCCTGGCACCGTAGGACA AATCCCCCATCCGCTAGGATG
Exon 9	GCAAGGTCAACCAGTGTACCCAATCACGACAGGAGGATTCAGACACCCCTACCCCACAGCTCTGACCGTC AATGCTTCCATGTCCAG
Exon 10 a	CTTTCTGTCTTCTAGGTTCCCTCCCCATATGGTCCCACCACATCATACGCTACACACGACGGGCATTCCG CATCCGGCCATAGTCACACCAACAGTCAAACAGGAATCGTCCCAGAGTGATGTCGGCTCACTCCATAGTT C
Exon 10 b	GTTCCCTCCCCATATGGTCCCACCACATCATACGCTACACACGACGGGCATTCCGCATCCGGCCATAGTC ACACCAACAGTCAAACAGGAATCGTCCCAGAGTGATGTCGGCTCACTCCATAGTTC
Exon 11	AAAGCATCAGGACTCCAAAAAGGAAGAAGAAAAGAAGAAGCCCCACATAAAGAAACCTCTTAATGCATTC ATGTTGTATATGAAGGAAATGAGAGCAAAGGTCGTAGCTGAGTGCACGTTGAAAGAAAGCGCGGCCATCA ACCAGATCCTTGGGCGGAGG
Exon 12	TGGCATGCACTGTCCAGAGAAGAGCAAGCGAAATACTACGAGCTGGCCCGGAAGGAGCGACAGCTTCATA TGCAACTGTACCCCGGCTGGTCCGCGCGGGATAACTAT

Exon 13	GGAAAGAAGAAGAAGAGGAAAAGGGACAAGCAGCCGGGAGAGACCAATG
Exon 14	AACACAGCGAATGTTTCCTAAATCCTTGCCTTTCACTTCCTCCGATTACAG
Exon 15	ACCTGAGCGCTCCTAAGAAATGCCGAGCGCGCTTTGGCCTTGATCAACAGAATAACTGGTGCGGCCCTTG CAG
Exon 16	ATGCAAATACTCCAAAGAAGTGTCGGGCACTGTTCGGGCTTGACCGACAGACTTTATGGTGCAAACCGTG CAG
Exon 17	TCTTTGAATTTGGAATATTACAATG
Exon 18	GAGAAAAAAAAAAGTGCGTTCGCTACATACAAGGTGAAGGCAGCTGCCTCAGCCCACCCTCTTCAGATGGA AGCTTACTAGATTCGCCTCCCCCCTCCCCGAACCTGCTAGGCTCCCCTCCCCGAGACGCCAAGTCACAGA CTGAGCAGACCCAGCCTCTGTCGCTGTCCCTGAAGCCCGACCCCCTGGCCCACCTGTCCATGATGCCTCC GCCACCCGCCCTCCTGCTCGCTGAGGCCACCCACAAGGCCTCCGCCCTCTGTCCCAACGGGGCCCTGGAC CTGCCCCCAGCCGCTTTGCAGCCTGCCGCCCCCTCCTCATCAATTGCACAGCCGTCGACTTCTTCCTTAC ATTCCCACAGCTCCCTGGCCGGGACCCAGCCCCAGCCGCTGTCGCTCGTCACCAAGTCTTTAGAATAG

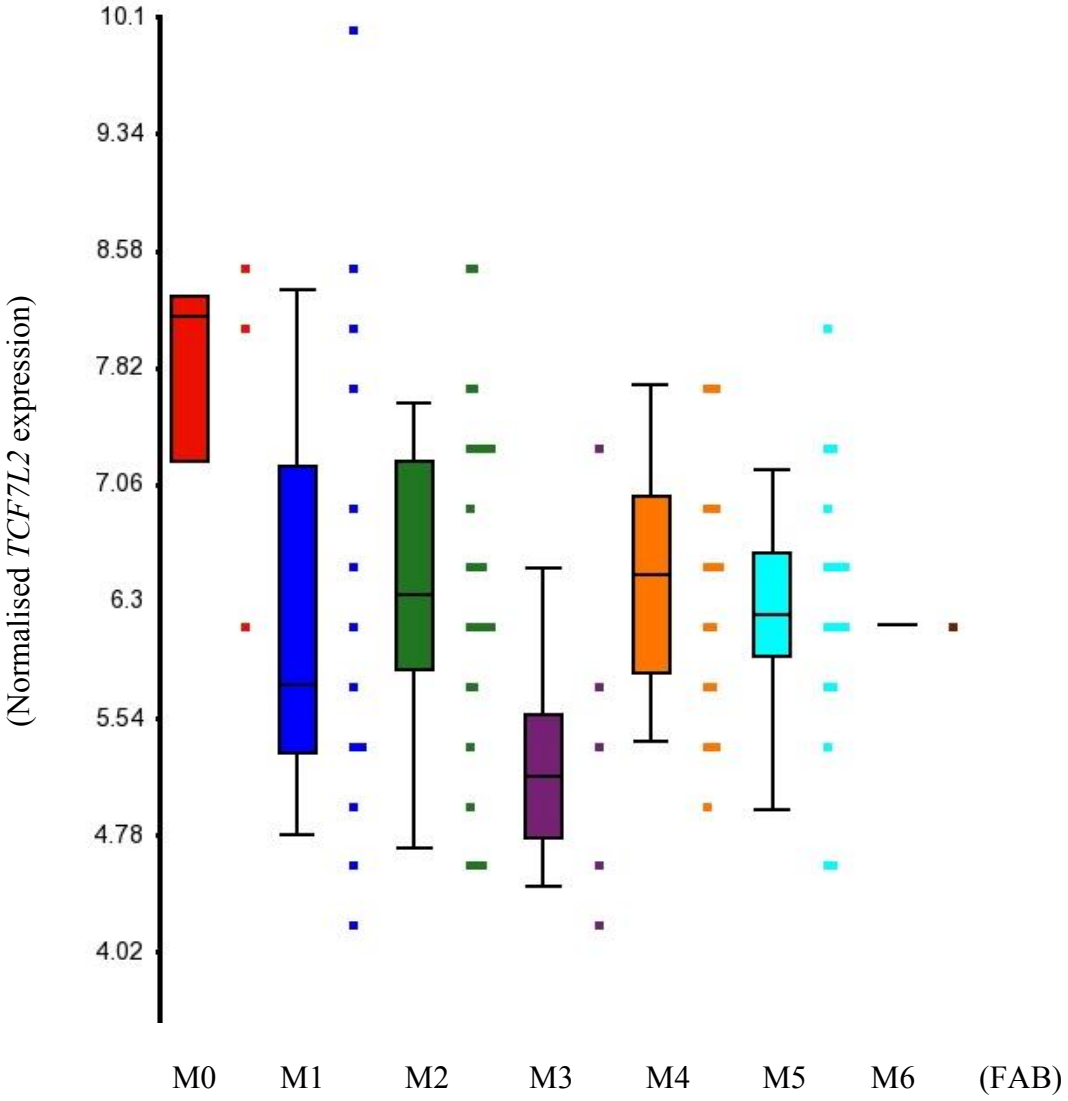
Supplementary 4.2: TCF7L2 exon sequences published for NM_001146274.1. Blue font indicates 5' UTR sequences.



Supplementary 4.3: Distribution of nuclear and cytosolic TCF7L2 protein isoforms detected by western blotting in normal $CD34^+$ HPC from CB at different stage of culture. Protein extracts were prepared from indicated samples and the blots were probed with monoclonal TCF7L2 antibody (exon11/Leu330 epitope). Loading controls used to show purity of each fraction were probed with GAPDH and Histone H1.



Supplementary 5.1: Proliferative activity of shRNA transduced U937-pBAR-VubiR reporter cell lines. Proliferation was counted as cell number /ml ($x10^5$). Relative proliferation shown for each shRNA target is relative to the control.



Supplementary 6.1: *TCF7L2* gene expression in AML patients (n = 98). Data was obtained from GC-RMA normalised expression generated by HG-U133 Plus2.0 Affymetrix probe set 212759_s_at.

Probe-set ID number	Gene name	Gene symbol	RefSeq ID (NCBI)	Probe -set ID HG-U133 Plus2.0
1	ATP-binding cassette, subfamily B, member 4 isoform A	*ABCB1*	NM_000443	209994_s_at
2		*ABCB1*	NM_000443	209993_at
3	adenomatous polyposis coli	*APC*	NM_000038	216933_x_at
4		*APC*	NM_000038	203526_s_at
5		*APC*	NM_000038	203525_s_at
6		*APC*	NM_000038	203527_s_at
7		*APC*	NM_000038	215310_at
8	adenomatosis polyposis coli 2	*APC2*	NM_005883	217174_s_at
9		*APC2*	NM_005883	205320_at
10	atonal homolog 1	*ATOH1*	NM_005172	221336_at
11	axin 1 isoform a	*AXIN1*	NM_003502	212849_at
12	axin 2	*AXIN2*	NM_004655	224498_x_at
13		*AXIN2*	NM_004655	222695_s_at
14		*AXIN2*	NM_004655	224176_s_at
15		*AXIN2*	NM_004655	222696_at
16	osteocalcin preproprotein	*BGLAP*	NM_199173	206956_at
17	baculoviral IAP repeat-containing protein 5 isoform 2	*BIRC5*	NM_001012270	202095_s_at
18		*BIRC5*	NM_001012270	210334_x_at
19		*BIRC5*	NM_001012270	202094_at
20	bone morphogenetic protein 4 preproprotein	*BMP4*	NM_001202	211518_s_at
21	beta-transducin repeat containing protein isoform 2	*BTRC*	NM_003939	1563620_at
22		*BTRC*	NM_003939	216091_s_at
23		*BTRC*	NM_003939	222374_at
24		*BTRC*	NM_003939	224471_s_at
25		*BTRC*	NM_003939	204901_at
26	calcyclin binding protein isoform 2	*CACYBP*	NM_001007214	210691_s_at
27		*CACYBP*	NM_001007214	201381_x_at
28		*CACYBP*	NM_001007214	211761_s_at
29		*CACYBP*	NM_001007214	201382_at
30	calcium/calmodulin-dependent protein kinase II alpha isoform 1	*CAMK2A*	NM_015981	207613_s_at
31		*CAMK2A*	NM_015981	213108_at
32	calcium/calmodulin-dependent protein kinase II beta isoform 1	*CAMK2B*	NM_001220	209956_s_at
33		*CAMK2B*	NM_001220	210404_x_at
34		*CAMK2B*	NM_001220	211483_x_at
35		*CAMK2B*	NM_001220	213276_at
36		*CAMK2B*	NM_001220	34846_at

37	calcium/calmodulin-dependent protein kinase II delta isoform 3	*CAMK2D*	NM_001221	231793_s_at
38		*CAMK2D*	NM_001221	224994_at
39		*CAMK2D*	NM_001221	225019_at
40		*CAMK2D*	NM_001221	228555_at
41		*CAMK2D*	NM_001221	230749_s_at
42	calcium/calmodulin-dependent protein kinase II gamma isoform 4	*CAMK2G*	NM_001222	212669_at
43		*CAMK2G*	NM_001222	214322_at
44		*CAMK2G*	NM_001222	212757_s_at
45	cyclin D1	*CCND1*	NM_053056	208711_s_at
46		*CCND1*	NM_053056	208712_at
47	cyclin D2	*CCND2*	NM_001759	200951_s_at
48		*CCND2*	NM_001759	200952_s_at
49		*CCND2*	NM_001759	200953_s_at
50	cyclin D3 isoform 1	*CCND3*	NM_001136017	201700_at
51		*CCND3*	NM_001136017	1562028_at
52	CD44 antigen isoform 1 precursor	*CD44*	NM_000610	210916_s_at
53		*CD44*	NM_000610	1557905_s_at
54		*CD44*	NM_000610	204490_s_at
55		*CD44*	NM_000610	209835_x_at
56		*CD44*	NM_000610	204489_s_at
57		*CD44*	NM_000610	229221_at
58		*CD44*	NM_000610	212063_at
59		*CD44*	NM_000610	212014_x_at
60		*CD44*	NM_000610	1565868_at
61		*CD44*	NM_000610	234411_x_at
62		*CD44*	NM_000610	216056_at
63		*CD44*	NM_000610	217523_at
64		*CD44*	NM_000610	234418_x_at
65	cyclin-dependent kinase inhibitor 2A isoform 1	*CDKN2A*	NM_000077	211156_at
66		*CDKN2A*	NM_000077	209644_x_at
67		*CDKN2A*	NM_000077	207039_at
68	caudal type homeobox 1	*CDX1*	NM_001804	206430_at
69	caudal type homeobox 4	*CDX4*	NM_005193	221340_at
70	cerberus 1	*CER1*	NM_005454	221378_at
71	chromodomain helicase DNA binding protein 8	*CHD8*	NM_020920	212571_at
72	calcium binding protein P22	*CHP*	NM_007236	207993_s_at
73		*CHP*	NM_007236	214665_s_at
74	hepatocellular carcinoma antigen gene 520	*CHP2*	NM_022097	206149_at
75	claudin 1	*CLDN1*	NM_021101	218182_s_at

76		*CLDN1*	NM_021101	222549_at
77	cAMP Response Element-Binding Protein (CREB-binding protein)	*CREBBP*	NM_001079846	211808_s_at
78		*CREBBP*	NM_001079846	202160_at
79		*CREBBP*	NM_001079846	228177_at
80	casein kinase 1, alpha 1 isoform 1	*CSNK1A1*	NM_001025105	206562_s_at
81		*CSNK1A1*	NM_001025105	213086_s_at
82		*CSNK1A1*	NM_001025105	213860_x_at
83		*CSNK1A1*	NM_001025105	208867_s_at
84		*CSNK1A1*	NM_001025105	208865_at
85		*CSNK1A1*	NM_001025105	1556006_s_at
86		*CSNK1A1*	NM_001025105	208866_at
87		*CSNK1A1*	NM_001025105	226920_at
88		*CSNK1A1*	NM_001025105	240221_at
89	casein kinase I epsilon E	*CSNK1E*	NM_001894	202332_at
90		*CSNK1E*	NM_001894	226858_at
91		*CSNK1E*	NM_001894	222015_at
92		*CSNK1E*	NM_001894	234943_at
93		*CSNK1E*	NM_001894	225756_at
94	casein kinase I epsilon G3	*CSNK1G3*	NM_001031812	220768_s_at
95		*CSNK1G3*	NM_001031812	229702_at
96		*CSNK1G3*	NM_001031812	227767_at
97	casein kinase II alpha 1 subunit isoform a	*CSNK2A1*	NM_001895	212073_at
98		*CSNK2A1*	NM_001895	212075_s_at
99		*CSNK2A1*	NM_001895	206075_s_at
100		*CSNK2A1*	NM_001895	212072_s_at
101		*CSNK2A1*	NM_001895	229216_s_at
102	casein kinase 2, alpha prime polypeptide	*CSNK2A2*	NM_001896	203575_at
103		*CSNK2A2*	NM_001896	224922_at
104	casein kinase 2, beta polypeptide	*CSNK2B*	NM_001320	201390_s_at
105		*CSNK2B*	NM_001320	231777_at
106	C-terminal binding protein 1 isoform 2	*CTBP1*	NM_001012614	203392_s_at
107		*CTBP1*	NM_001012614	212863_x_at
108		*CTBP1*	NM_001012614	213980_s_at
109		*CTBP1*	NM_001012614	1557714_at
110	C-terminal binding protein 2 isoform 1	*CTBP2*	NM_001083914	210554_s_at
111		*CTBP2*	NM_001083914	201220_x_at
112		*CTBP2*	NM_001083914	210835_s_at
113		*CTBP2*	NM_001083914	201218_at
114		*CTBP2*	NM_001083914	201219_at
115		*CTBP2*	NM_001083914	215377_at

116	cytotoxic T-lymphocyte-associated protein 4 isoform b precursor	*CTLA4*	NM_001037631	234362_s_at
117		*CTLA4*	NM_001037631	221331_x_at
118		*CTLA4*	NM_001037631	231794_at
119		*CTLA4*	NM_001037631	236341_at
120		*CTLA4*	NM_001037631	234895_at
121	catenin (cadherin-associated protein), beta 1, 88kDa	*CTNNB1*	NM_001098209	201533_at
122		*CTNNB1*	NM_001098209	1554411_at
123		*CTNNB1*	NM_001098209	223679_at
124	cadherin; E-cadherin;cadherin-associated protein), delta 1; Cadherin-associated Src substrate; ctenin delta 1 isoform1ABC	*CTNND1*	NM_001085458	211240_x_at
125		*CTNND1*	NM_001085458	208862_s_at
126		*CTNND1*	NM_001085458	208407_s_at
127		*CTNND1*	NM_001085458	1557944_s_at
128	cullin 1	*CUL1*	NM_003592	207614_s_at
129		*CUL1*	NM_003592	238509_at
130	CXXC finger 4	*CXXC4*	NM_025212	220277_at
131		*CXXC4*	NM_025212	229774_at
132	cysteine-rich, angiogenic inducer, 61	*CYR61*	NM_001554	210764_s_at
133		*CYR61*	NM_001554	201289_at
134	dishevelled associated activator of morphogenesis 1	*DAAM1*	NM_014992	226666_at
135		*DAAM1*	NM_014992	216060_s_at
136		*DAAM1*	NM_014992	244062_at
137		*DAAM1*	NM_014992	232552_at
138	dishevelled associated activator of morphogenesis 2	*DAAM2*	NM_015345	212793_at
139	dickkopf homolog 1 precursor	*DKK1*	NM_012242	204602_at
140	dickkopf homolog 2 precursor	*DKK2*	NM_014421	219908_at
141		*DKK2*	NM_014421	224199_at
142	dickkopf homolog 3 precursor	*DKK3*	NM_001018057	221127_s_at
143		*DKK3*	NM_001018057	230508_at
144		*DKK3*	NM_001018057	221126_at
145		*DKK3*	NM_001018057	214247_s_at
146		*DKK3*	NM_001018057	202196_s_at
147	dickkopf homolog 4 precursor	*DKK4*	NM_014420	206619_at
148	delta-like 1	*DLL1*	NM_005618	227938_s_at
149		*DLL1*	NM_005618	224215_s_at
150	dishevelled 1	*DVL-1*	NM_004421	203230_at
151	dishevelled 2	*DVL-2*	NM_004422	218759_at
152		*DVL-2*	NM_004422	57532_at
153	dishevelled, dsh homolog 3	*DVL-3*	NM_004423	201907_x_at

154		*DVL-3*	NM_004423	201908_at
155	ectodysplasin A isoform EDA-A2	*EDA*	NM_001005609	211130_x_at
156		*EDA*	NM_001005609	206217_at
157		*EDA*	NM_001005609	211131_s_at
158		*EDA*	NM_001005609	211128_at
159		*EDA*	NM_001005609	211129_x_at
160		*EDA*	NM_001005609	211127_x_at
161	ectodysplasin A receptor precursor	*EDAR*	NM_022336	220048_at
162	endothelin 1	*EDN1*	NM_001955	218995_s_at
163		*EDN1*	NM_001955	222802_at
164		*EDN1*	NM_001955	1564630_at
165	ephrin-B1 precursor	*EFNB1*	NM_004429	202711_at
166	ephrin B2	*EFNB2*	NM_004093	202669_s_at
167		*EFNB2*	NM_004093	202668_at
168	epidermal growth factor receptor isoform a precursor	*EGFR*	NM_005228	211607_x_at
169		*EGFR*	NM_005228	210984_x_at
170		*EGFR*	NM_005228	201983_s_at
171		*EGFR*	NM_005228	211550_at
172		*EGFR*	NM_005228	1565483_at
173		*EGFR*	NM_005228	1565484_x_at
174		*EGFR*	NM_005228	201984_s_at
175		*EGFR*	NM_005228	211551_at
176	E1A binding protein p300	*EP300*	NM_001429	213579_s_at
177		*EP300*	NM_001429	202221_s_at
178	F-box and WD repeat domain containing 2	*FBXW2*	NM_012164	218941_at
179		*FBXW2*	NM_012164	209630_s_at
180		*FBXW2*	NM_012164	235195_at
181		*FBXW2*	NM_012164	241736_at
182		*FBXW2*	NM_012164	1560752_at
183	fibroblast growth factor 18 precursor	*FGF18*	NM_003862	211485_s_at
184		*FGF18*	NM_003862	206987_x_at
185		*FGF18*	NM_003862	211029_x_at
186		*FGF18*	NM_003862	214284_s_at
187		*FGF18*	NM_003862	231382_at
188		*FGF18*	NM_003862	206986_at
189	fibroblast growth factor 20	*FGF20*	NM_019851	220394_at
190	fibroblast growth factor 4 precursor	*FGF4*	NM_002007	1552982_a_at
191		*FGF4*	NM_002007	206783_at
192	fibroblast growth factor 9 precursor	*FGF9*	NM_002010	239178_at

193		*FGF9*	NM_002010	206404_at
194	fibronectin 1 isoform 3 preproprotein	*FN1*	NM_002026	214701_s_at
195		*FN1*	NM_002026	210495_x_at
196		*FN1*	NM_002026	211719_x_at
197		*FN1*	NM_002026	216442_x_at
198		*FN1*	NM_002026	212464_s_at
199		*FN1*	NM_002026	214702_at
200		*FN1*	NM_002026	1558199_at
201	FOS-like antigen 1	*FOSL1*	NM_005438	204420_at
202	forkhead box N1	*FOXN1*	NM_003593	207683_at
203	frequently rearranged in advanced T-cell lymphoma 1	*FRAT-1*	NM_005479	219889_at
204	frequently rearranged in advanced T-cell lymphomas 2	*FRAT-2*	NM_012083	209864_at
205	frizzled-related protein	*FRZB*	NM_001463	203698_s_at
206		*FRZB*	NM_001463	203697_at
207	follistatin isoform FST317 precursor	*FST*	NM_006350	204948_s_at
208		*FST*	NM_006350	226847_at
209		*FST*	NM_006350	207345_at
210	frizzled homolog1	*FZD1*	NM_003505	204452_s_at
211		*FZD1*	NM_003505	204451_at
212	frizzled homolog 10	*FZD10*	NM_007197	219764_at
213	frizzled homolog 2	*FZD2*	NM_001466	210220_at
214	frizzled homolog 3	*FZD3*	NM_017412	219683_at
215	frizzled homolog 4	*FZD4*	NM_012193	224337_s_at
216		*FZD4*	NM_012193	218665_at
217	frizzled homolog 5	*FZD5*	NM_003468	206136_at
218		*FZD5*	NM_003468	221245_s_at
219	frizzled homolog 6	*FZD6*	NM_003506	203987_at
220	frizzled homolog 7	*FZD7*	NM_003507	203705_s_at
221		*FZD7*	NM_003507	203706_s_at
222	frizzled homolog 8	*FZD8*	NM_031866	227405_s_at
223		*FZD8*	NM_031866	216587_s_at
224		*FZD8*	NM_031866	224325_at
225	frizzled homolog 9	*FZD9*	NM_003508	207639_at
226	Gastrin	*GAST*	NM_000805	208138_at
227	gastrulation brain homeo box 2	*GBX2*	NM_001485	210560_at
228	connexin 43	*GJA1*	NM_000165	201667_at
229	gap junction protein, beta 6	*GJB6*	NM_001110219	231771_at
230	glucagon-like peptide 1 receptor	*GLP1R*	NM_002062	208391_s_at
231		*GLP1R*	NM_002062	211232_x_at

232		*GLP1R*	NM_002062	208390_s_at
233		*GLP1R*	NM_002062	208400_at
234		*GLP1R*	NM_002062	208401_s_at
235	gremlin 2 precursor	*GREM2*	NM_022469	220794_at
236		*GREM2*	NM_022469	240509_s_at
237		*GREM2*	NM_022469	235504_at
238	glycogen synthase kinase 3 beta isoform 2	*GSK3B*	NM_001146156	209945_s_at
239		*GSK3B*	NM_001146156	242336_at
240		*GSK3B*	NM_001146156	226183_at
241		*GSK3B*	NM_001146156	226191_at
242	inhibitor of DNA binding 2	*ID2*	NM_002166	213931_at
243		*ID2*	NM_002166	201565_s_at
244		*ID2*	NM_002166	201566_x_at
245	insulin-like growth factor 1 isoform 4 preproprotein	*IGF1*	NM_000618	211577_s_at
246		*IGF1*	NM_000618	209542_x_at
247		*IGF1*	NM_000618	209540_at
248		*IGF1*	NM_000618	209541_at
249	insulin-like growth factor 2 isoform 1 precursor	*IGF2*	NM_000612	210881_s_at
250		*IGF2*	NM_000612	202410_x_at
251		*IGF2*	NM_000612	202409_at
252	interleukin 6 precursor	*IL6*	NM_000600	205207_at
253	interleukin 8 precursor	*IL8*	NM_000584	211506_s_at
254		*IL8*	NM_000584	202859_x_at
255	integrin-linked kinase	*ILK*	NM_001014794	201234_at
256	iroquois homeobox 3	*IRX3*	NM_024336	229638_at
257	islet-1	*ISL1*	NM_002202	206104_at
258	immunoglobulin superfamily containing leucine-rich repeat	*ISLR*	NM_005545	207191_s_at
259	jagged 1 precursor	*JAG1*	NM_000214	209098_s_at
260		*JAG1*	NM_000214	209097_s_at
261		*JAG1*	NM_000214	216268_s_at
262		*JAG1*	NM_000214	209099_x_at
263		*JAG1*	NM_000214	231183_s_at
264	jun oncogene	*JUN*	NM_002228	201465_s_at
265		*JUN*	NM_002228	201466_s_at
266		*JUN*	NM_002228	201464_x_at
267	junction plakoglobin	*JUP*	NM_002230	201015_s_at
268	kringle-containing transmembrane protein 2 isoform b precursor	*KREMEN2*	NM_024507	219692_at
269	keratin 14	*KRT14*	NM_000526	209351_at

270	keratin 5	*KRT5*	NM_000424	201820_at
271	L1 cell adhesion molecule isoform 1 precursor	*L1CAM*	NM_000425	204585_s_at
272		*L1CAM*	NM_000425	204584_at
273	low density lipoprotein receptor precursor	*LDLR*	NM_000527	217173_s_at
274		*LDLR*	NM_000527	202067_s_at
275		*LDLR*	NM_000527	202068_s_at
276		*LDLR*	NM_000527	217103_at
277		*LDLR*	NM_000527	217005_at
278		*LDLR*	NM_000527	217183_at
279	lymphoid enhancer- binding factor - Lef	*LEF1*	NM_001130713	210948_s_at
280		*LEF1*	NM_001130713	221558_s_at
281		*LEF1*	NM_001130713	221557_s_at
282	leucine-rich repeat-containing G protein-coupled receptor 5	*LGR5*	NM_003667	210393_at
283		*LGR5*	NM_003667	213880_at
284	low density lipoprotein receptor-related protein 5	*LRP5*	NM_002335	229591_at
285		*LRP5*	NM_002335	209468_at
286	low density lipoprotein receptor-related protein 6	*LRP6*	NM_002336	34697_at
287		*LRP6*	NM_002336	205606_at
288		*LRP6*	NM_002336	225745_at
289	mitogen-activated protein kinase kinase kinase 7 isoform A	*MAP3K7*	NM_003188	211536_x_at
290		*MAP3K7*	NM_003188	211537_x_at
291		*MAP3K7*	NM_003188	206854_s_at
292		*MAP3K7*	NM_003188	206853_s_at
293	mitogen-activated protein kinase 10 isoform 1	*MAPK10*	NM_002753	204813_at
294	mitogen-activated protein kinase 8 isoform JNK1 alpha1	*MAPK8*	NM_002750	210477_x_at
295		*MAPK8*	NM_002750	210671_x_at
296		*MAPK8*	NM_002750	226048_at
297		*MAPK8*	NM_002750	226046_at
298		*MAPK8*	NM_002750	229664_at
299	mitogen-activated protein kinase 9 isoform JNK2 gamma	*MAPK9*	NM_001135044	210570_x_at
300		*MAPK9*	NM_001135044	203218_at
301		*MAPK9*	NM_001135044	225781_at
302	microphthalmia-associated transcription factor isoform 4	*MITF*	NM_000248	207233_s_at
303		*MITF*	NM_000248	226066_at
304		*MITF*	NM_000248	1554874_at
305	matrix metalloproteinase 2 isoform b	*MMP2*	NM_001127891	201069_at

306		*MMP2*	NM_001127891	1566678_at
307		*MMP2*	NM_001127891	1566677_at
308	matrix metalloproteinase 26 preproprotein	*MMP26*	NM_021801	220541_at
309	matrix metalloproteinase 7 preproprotein	*MMP7*	NM_002423	204259_at
310	matrix metalloproteinase 9 preproprotein	*MMP9*	NM_004994	203936_s_at
311	myc proto-oncogene protein, Transcription factor p64 , v-myc avian myelocytomatosis viral oncogene homolog	*MYC*	NM_002467	202431_s_at
312	c-myc binding protein	*MYCBP*	NM_012333	203360_s_at
313		*MYCBP*	NM_012333	203359_s_at
314		*MYCBP*	NM_012333	203361_s_at
315	Nanog homeobox	*NANOG*	NM_024865	220184_at
316	norrin precursor	*NDP*	NM_000266	206022_at
317	neurogenic differentiation 1	*NEUROD1*	NM_002500	206282_at
318		*NEUROD1*	NM_002500	1556057_s_at
319	neurogenin 1	*NEUROG1*	NM_006161	208497_x_at
320	nuclear factor of activated T-cells 5 isoform	*NFAT5*	NM_001113178	215092_s_at
321		*NFAT5*	NM_001113178	208003_s_at
322		*NFAT5*	NM_001113178	224984_at
323	nuclear factor of activated T-cells, cytosolic component 1 isoform B	*NFATC1*	NM_006162	209664_x_at
324		*NFATC1*	NM_006162	208196_x_at
325		*NFATC1*	NM_006162	211105_s_at
326		*NFATC1*	NM_006162	210162_s_at
327	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 isoform	*NFATC2*	NM_001136021	224542_s_at
328		*NFATC2*	NM_001136021	231801_at
329		*NFATC2*	NM_001136021	226991_at
330	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 3 isoform	*NFATC3*	NM_004555	210555_s_at
331		*NFATC3*	NM_004555	207416_s_at
332		*NFATC3*	NM_004555	210556_at
333	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 4 isoform	*NFATC4*	NM_001136022	205897_at
334		*NFATC4*	NM_001136022	236270_at
335		*NFATC4*	NM_001136022	213345_at
336	naked cuticle homolog 1	*NKD1*	NM_033119	1553115_at
337	naked cuticle homolog 2	*NKD2*	NM_033120	232201_at
338	NK2 transcription factor related, locus 2	*NKX2-2*	NM_002509	206915_at
339	nemo like kinase	*NLK*	NM_016231	222590_s_at
340		*NLK*	NM_016231	218318_s_at
341		*NLK*	NM_016231	222589_at
342	nitric oxide synthase 2A	*NOS2*	NM_000625	210037_s_at
343	neuronal cell adhesion molecule isoform A precursor	*NRCAM*	NM_001037132	204105_s_at

344		*NRCAM*	NM_001037132	216959_x_at
345	platelet-activating factor acetylhydrolase, isoform Ib, subunit 1 (45kDa)	*PAFAH1B1*	NM_000430	200815_s_at
346		*PAFAH1B1*	NM_000430	200813_s_at
347		*PAFAH1B1*	NM_000430	200816_s_at
348		*PAFAH1B1*	NM_000430	211547_s_at
349	paired-like homeodomain transcription factor 2 isoform c	*PITX2*	NM_000325	207558_s_at
350	plasminogen activator, urokinase isoform 2	*PLAU*	NM_001145031	211668_s_at
351		*PLAU*	NM_001145031	205479_s_at
352	phosphoinositide-specific phospholipase C beta 1 isoform a	*PLCB1*	NM_015192	211925_s_at
353		*PLCB1*	NM_015192	213222_at
354		*PLCB1*	NM_015192	215687_x_at
355	phospholipase C, beta 2	*PLCB2*	NM_004573	204046_at
356		*PLCB2*	NM_004573	210388_at
357	phospholipase C beta 3	*PLCB3*	NM_000932	213384_x_at
358	phospholipase C beta 4 isoform a	*PLCB4*	NM_000933	203895_at
359		*PLCB4*	NM_000933	203896_s_at
360		*PLCB4*	NM_000933	240728_at
361	porcupine isoform A	*PORCN*	NM_022825	219483_s_at
362	periostin, osteoblast specific factor isoform 2	*POSTN*	NM_001135934	1555778_a_at
363		*POSTN*	NM_001135934	1555777_at
364		*POSTN*	NM_001135934	210809_s_at
365		*POSTN*	NM_001135934	214981_at
366	POU class 5 homeobox 1B	*POU5F1*	NM_001159542	208286_x_at
367		*POU5F1B*	NM_001159542	214532_x_at
368	peroxisome proliferative activated receptor, delta isoform 1	*PPARD*	NM_006238	242218_at
369		*PPARD*	NM_006238	210636_at
370		*PPARD*	NM_006238	37152_at
371		*PPARD*	NM_006238	208044_s_at
372	protein phosphatase 2, catalytic subunit, alpha isoform	*PPP2CA*	NM_002715	208652_at
373	protein phosphatase 2, catalytic subunit, beta isoform	*PPP2CB*	NM_001009552	201375_s_at
374		*PPP2CB*	NM_001009552	201374_x_at
375	alpha isoform of regulatory subunit A, protein phosphatase 2	*PPP2R1A*	NM_014225	200695_at
376	beta isoform of regulatory subunit A, protein phosphatase 2 isoform a	*PPP2R1B*	NM_002716	202884_s_at
377		*PPP2R1B*	NM_002716	222351_at
378		*PPP2R1B*	NM_002716	202886_s_at
379		*PPP2R1B*	NM_002716	202883_s_at
380		*PPP2R1B*	NM_002716	202885_s_at
381	protein phosphatase 2, regulatory subunit B (B56), alpha isoform	*PPP2R5A*	NM_006243	202187_s_at

382		*PPP2R5A*	NM_006243	202186_x_at
383	beta isoform of regulatory subunit B56, protein phosphatase 2A	*PPP2R5B*	NM_006244	204611_s_at
384		*PPP2R5B*	NM_006244	635_s_at
385	gamma isoform of regulatory subunit B56, protein phosphatase 2A	*PPP2R5C*	NM_001161725	1554365_a_at
386		*PPP2R5C*	NM_001161725	213305_s_at
387		*PPP2R5C*	NM_001161725	1554364_at
388		*PPP2R5C*	NM_001161725	201877_s_at
389		*PPP2R5C*	NM_001161725	214083_at
390		*PPP2R5C*	NM_001161725	1557718_at
391	delta isoform of regulatory subunit B56, protein phosphatase 2A isoform 1	*PPP2R5D*	NM_006245	211159_s_at
392		*PPP2R5D*	NM_006245	202513_s_at
393	epsilon isoform of regulatory subunit B56, protein phosphatase 2A	*PPP2R5E*	NM_006246	203338_at
394		*PPP2R5E*	NM_006246	229322_at
395	protein phosphatase 3, catalytic subunit, alpha isoform isoform	*PPP3CA*	NM_000944	202425_x_at
396		*PPP3CA*	NM_000944	202457_s_at
397		*PPP3CA*	NM_000944	202429_s_at
398	protein phosphatase 3, catalytic subunit, beta isoform	*PPP3CB*	NM_001142353	202432_at
399		*PPP3CB*	NM_001142353	209817_at
400	protein phosphatase 3, catalytic subunit, gamma	*PPP3CC*	NM_005605	207000_s_at
401		*PPP3CC*	NM_005605	32541_at
402		*PPP3CC*	NM_005605	213950_s_at
403		*PPP3CC*	NM_005605	32540_at
404	protein phosphatase 3, regulatory subunit B, alpha isoform	*PPP3R1*	NM_000945	204507_s_at
405		*PPP3R1*	NM_000945	204506_at
406	protein phosphatase 3 regulatory subunit B, beta isoform	*PPP3R2*	NM_147180	231304_at
407		*PPP3R2*	NM_147180	244782_at
408	prickle homolog 1	*PRICKLE1*	NM_001144881	226065_at
409		*PRICKLE1*	NM_001144881	230708_at
410		*PRICKLE1*	NM_001144881	226069_at
411		*PRICKLE1*	NM_001144881	232811_x_at
412	prickle-like 2	*PRICKLE2*	NM_198859	225968_at
413	cAMP-dependent protein kinase catalytic subunit alpha isoform 1	*PRKACA*	NM_002730	216234_s_at
414		*PRKACA*	NM_002730	202801_at
415	cAMP-dependent protein kinase catalytic subunit beta isoform 2	*PRKACB*	NM_002731	202742_s_at
416		*PRKACB*	NM_002731	202741_at
417		*PRKACB*	NM_002731	235780_at
418	protein kinase, cAMP-dependent, catalytic, gamma	*PRKACG*	NM_002732	207228_at

419	protein kinase C, alpha	*PRKCA*	NM_002737	206923_at
420		*PRKCA*	NM_002737	213093_at
421		*PRKCA*	NM_002737	215195_at
422		*PRKCA*	NM_002737	1560074_at
423		*PRKCA*	NM_002737	215194_at
424	protein kinase C, beta isoform 2	*PRKCB*	NM_002738	209685_s_at
425		*PRKCB*	NM_002738	207957_s_at
426		*PRKCB*	NM_002738	228795_at
427		*PRKCB*	NM_002738	227824_at
428		*PRKCB*	NM_002738	227817_at
429		*PRKCB*	NM_002738	230437_s_at
430	protein kinase C, delta	*PRKCD*	NM_006254	202545_at
431	protein kinase C, epsilon	*PRKCE*	NM_005400	239011_at
432		*PRKCE*	NM_005400	236459_at
433		*PRKCE*	NM_005400	226101_at
434		*PRKCE*	NM_005400	206248_at
435	protein kinase C, gamma	*PRKCG*	NM_002739	206270_at
436		*PRKCG*	NM_002739	236195_x_at
437	protein kinase C, eta	*PRKCH*	NM_006255	206099_at
438		*PRKCH*	NM_006255	218764_at
439	protein kinase C, iota	*PRKCI*	NM_002740	209677_at
440		*PRKCI*	NM_002740	209678_s_at
441		*PRKCI*	NM_002740	213518_at
442	protein kinase C, theta	*PRKCQ*	NM_006257	210039_s_at
443		*PRKCQ*	NM_006257	210038_at
444	protein kinase C, zeta isoform 2	*PRKCZ*	NM_001033581	202178_at
445		*PRKCZ*	NM_001033581	1569748_at
446	Protein kinase D1	*PRKD1*	NM_002742	205880_at
447		*PRKD1*	NM_002742	217705_at
448	protein kinase, X-linked	*PRKX*	NM_005044	204060_s_at
449		*PRKX*	NM_005044	204061_at
450	protein kinase, Y-linked	*PRKY*	NM_002760	206279_at
451	presenilin 1 isoform I-467	*PSEN1*	NM_000021	203460_s_at
452		*PSEN1*	NM_000021	207782_s_at
453		*PSEN1*	NM_000021	226577_at
454		*PSEN1*	NM_000021	238816_at
455		*PSEN1*	NM_000021	1559206_at
456		*PSEN1*	NM_000021	1567443_x_at
457	presenilin 2 isoform 1	*PSEN2*	NM_000447	204261_s_at
458		*PSEN2*	NM_000447	204262_s_at

459		*PSEN2*	NM_000447	211373_s_at
460	prostaglandin-endoperoxide synthase 2 precursor	*PTGS2*	NM_000963	1554997_a_at
461		*PTGS2*	NM_000963	204748_at
462	pituitary tumor-transforming protein 1	*PTTG1*	NM_004219	203554_x_at
463	pituitary tumor-transforming_2	*PTTG2*	NM_006607	214557_at
464	ras-related C3 botulinum toxin substrate 1 isoform Rac1	*RAC1*	NM_006908	208641_s_at
465		*RAC1*	NM_006908	208640_at
466		*RAC1*	NM_006908	1567458_s_at
467		*RAC1*	NM_006908	1567457_at
468	ras-related C3 botulinum toxin substrate 2	*RAC2*	NM_002872	207419_s_at
469		*RAC2*	NM_002872	213603_s_at
470	ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3)	*RAC3*	NM_005052	206103_at
471	retinoic acid receptor, gamma isoform 1	*RARG*	NM_000966	204188_s_at
472		*RARG*	NM_000966	204189_at
473		*RARG*	NM_000966	217178_at
474	ring-box 1	*RBX1*	NM_014248	218117_at
475	ret proto-oncogene isoform c	*RET*	NM_020630	205879_x_at
476		*RET*	NM_020630	215771_x_at
477		*RET*	NM_020630	211421_s_at
478	ras homolog gene family, member A	*RHOA*	NM_001664	200059_s_at
479		*RHOA*	NM_001664	1555814_a_at
480		*RHOA*	NM_001664	240337_at
481	ras homolog gene family, member U	*RHOU*	NM_021205	223169_s_at
482	Rho-associated, coiled-coil containing protein kinase 1	*ROCK1*	NM_005406	213044_at
483		*ROCK1*	NM_005406	235854_x_at
484		*ROCK1*	NM_005406	214578_s_at
485		*ROCK1*	NM_005406	230239_at
486	Rho-associated, coiled-coil containing protein kinase 2	*ROCK2*	NM_004850	211504_x_at
487		*ROCK2*	NM_004850	202762_at
488	Receptor tyrosine kinase-like orphan receptor 2 precurso	*ROR2*	NM_004560	205578_at
489	R-spondin1	*RSPO1*	NM_001038633	241450_at
490	runt-related transcription factor 2 isoform b	*RUNX2*	NM_001015051	216994_s_at
491		*RUNX2*	NM_001015051	221282_x_at
492		*RUNX2*	NM_001015051	221283_at
493		*RUNX2*	NM_001015051	232231_at
494		*RUNX2*	NM_001015051	236858_s_at
495		*RUNX2*	NM_001015051	236859_at
496	RuvB-like 1	*RUVBL1*	NM_003707	201614_s_at

497	sal-like 4	*SALL4*	NM_020436	229661_at
498		*SALL4*	NM_020436	236501_at
499	SUMO1/sentrin/SMT3 specific protease 2	*SENP2*	NM_021627	218122_s_at
500		*SENP2*	NM_021627	222523_at
501	secreted frizzled-related protein 1	*SFRP1*	NM_003012	202036_s_at
502		*SFRP1*	NM_003012	202035_s_at
503		*SFRP1*	NM_003012	202037_s_at
504	secreted frizzled-related protein 2	*SFRP2*	NM_003013	223121_s_at
505		*SFRP2*	NM_003013	223122_s_at
506	secreted frizzled-related protein 4	*SFRP4*	NM_003014	204051_s_at
507		*SFRP4*	NM_003014	204052_s_at
508	secreted frizzled-related protein 5	*SFRP5*	NM_003015	207468_s_at
509	splicing factor, arginine/serine-rich 2, interacting protein	*SFRS2IP*	NM_004719	1570507_at
510	seven in absentia homolog 1 isoform b	*SIAH1*	NM_001006610	202980_s_at
511		*SIAH1*	NM_001006610	202981_x_at
512		*SIAH1*	NM_001006610	232365_at
513	sine oculis homeobox homolog 3	*SIX3*	NM_005413	206634_at
514	S-phase kinase-associated protein 1 isoform a	*SKP1*	NM_006930	207974_s_at
515		*SKP1*	NM_006930	200718_s_at
516		*SKP1*	NM_006930	200711_s_at
517		*SKP1*	NM_006930	200719_at
518	Sma- and Mad-related protein 2 isoform 1 ,SMAD family member 2	*SMAD2*	NM_001003652	203076_s_at
519		*SMAD2*	NM_001003652	203077_s_at
520		*SMAD2*	NM_001003652	226563_at
521		*SMAD2*	NM_001003652	203075_at
522		*SMAD2*	NM_001003652	235598_at
523		*SMAD2*	NM_001003652	239271_at
524	mothers against decapentaplegic homolog 3 isoform 2	*SMAD3*	NM_001145102	205397_x_at
525		*SMAD3*	NM_001145102	205398_s_at
526		*SMAD3*	NM_001145102	205396_at
527		*SMAD3*	NM_001145102	218284_at
528	mothers against decapentaplegic homolog 4	*SMAD4*	NM_005359	1565703_at
529		*SMAD4*	NM_005359	202527_s_at
530		*SMAD4*	NM_005359	235725_at
531		*SMAD4*	NM_005359	202526_at
532		*SMAD4*	NM_005359	1565702_at
533	snail 1 homolog	*SNAI1*	NM_005985	219480_at
534	snail 2	*SNAI2*	NM_003068	213139_at
535	snail homolog 3	*SNAI3*	NM_178310	1560228_at

536	SRY-box 17	*SOX17*	NM_022454	230943_at
537		*SOX17*	NM_022454	219993_at
538	transcription factor SOX9	*SOX9*	NM_000346	202935_s_at
539		*SOX9*	NM_000346	202936_s_at
540	Sp5 transcription factor	*SP5*	NM_001003845	235845_at
541	stimulated by retinoic acid gene 6 homolog isoform a	*STRA6*	NM_001142617	221701_s_at
542		*STRA6*	NM_001142617	1569334_at
543		*STRA6*	NM_001142617	1569335_a_at
544	transcription factor T	*T*	NM_003181	206524_at
545	transducin beta-like 1X isoform a	*TBL1X*	NM_001139466	201867_s_at
546		*TBL1X*	NM_001139466	201869_s_at
547		*TBL1X*	NM_001139466	201868_s_at
548		*TBL1X*	NM_001139466	213400_s_at
549		*TBL1X*	NM_001139466	1570293_at
550		*TBL1X*	NM_001139466	213401_s_at
551	transducin (beta)-like 1 X-linked receptor 1	*TBL1XR1*	NM_024665	222634_s_at
552		*TBL1XR1*	NM_024665	221428_s_at
553		*TBL1XR1*	NM_024665	223013_at
554		*TBL1XR1*	NM_024665	222633_at
555		*TBL1XR1*	NM_024665	235890_at
556		*TBL1XR1*	NM_024665	233633_at
557	transducin beta-like 1Y	*TBL1Y*	NM_033284	211462_s_at
558	transcription factor 12 isoform b	*TCF12*	NM_003205	208986_at
559		*TCF12*	NM_003205	215611_at
560		*TCF12*	NM_003205	235925_at
561	transcription factor 4 isoform a	*TCF4*	NM_001083962	222146_s_at
562		*TCF4*	NM_001083962	203753_at
563		*TCF4*	NM_001083962	212385_at
564		*TCF4*	NM_001083962	212386_at
565		*TCF4*	NM_001083962	212387_at
566		*TCF4*	NM_001083962	228837_at
567		*TCF4*	NM_001083962	212382_at
568		*TCF4*	NM_001083962	213891_s_at
569	T cell factor ,TCF1	*TCF7*	NM_001134851	205255_x_at
570		*TCF7*	NM_001134851	205254_x_at
571	HMG-box transcription factor ,TCF-3	*TCF7L1*	NM_031283	221016_s_at
572	transcription factor 7-like 2 isoform 1,TCF4	*TCF7L2*	NM_001146274	212759_s_at
573		*TCF7L2*	NM_001146274	212762_s_at
574		*TCF7L2*	NM_001146274	216511_s_at
575		*TCF7L2*	NM_001146274	216035_x_at

576		*TCF7L2*	NM_001146274	216037_x_at
577		*TCF7L2*	NM_001146274	212761_at
578		*TCF7L2*	NM_001146274	236094_at
579	T-cell lymphoma invasion and metastasis 1	*TIAM1*	NM_003253	206409_at
580		*TIAM1*	NM_003253	213135_at
581	transducin-like enhancer protein 1*groucho	*TLE1*	NM_005077	203220_s_at
582		*TLE1*	NM_005077	203222_s_at
583		*TLE1*	NM_005077	203221_at
584	transducin-like enhancer protein 3 isoform b *groucho	*TLE3*	NM_001105192	206472_s_at
585	tumor necrosis factor receptor superfamily, member 19 isoform 1 precursor	*TNFRSF19*	NM_018647	224090_s_at
586		*TNFRSF19*	NM_018647	223827_at
587		*TNFRSF19*	NM_018647	227812_at
588	tumor necrosis factor ligand superfamily, member 11 isoform 1	*TNFSF11*	NM_003701	211153_s_at
589		*TNFSF11*	NM_003701	210643_at
590	tumor necrosis factor (ligand) superfamily, member 9	*TNFSF9*	NM_003811	206907_at
591	tumor protein p53 isoform a	*TP53*	NM_000546	211300_s_at
592		*TP53*	NM_000546	201746_at
593	twist	*TWIST1*	NM_000474	213943_at
594	twist homolog 2	*TWIST2*	NM_057179	229404_at
595		*TWIST2*	NM_057179	1554163_at
596	vang-like 1	*VANGL1*	NM_138959	229997_at
597		*VANGL1*	NM_138959	219330_at
598		*VANGL1*	NM_138959	229134_at
599		*VANGL1*	NM_138959	229492_at
600	vang-like 2	*VANGL2*	NM_020335	226029_at
601	versican isoform 2 precursor	*VCAN*	NM_001126336	215646_s_at
602		*VCAN*	NM_001126336	211571_s_at
603		*VCAN*	NM_001126336	204620_s_at
604		*VCAN*	NM_001126336	221731_x_at
605		*VCAN*	NM_001126336	204619_s_at
606	vascular endothelial growth factor A isoform a precursor	*VEGFA*	NM_001025366	210513_s_at
607		*VEGFA*	NM_001025366	210512_s_at
608		*VEGFA*	NM_001025366	212171_x_at
609		*VEGFA*	NM_001025366	211527_x_at
610	vascular endothelial growth factor B	*VEGFB*	NM_003377	203683_s_at
611	vascular endothelial growth factor C preproprotein	*VEGFC*	NM_005429	209946_at
612	WNT inhibitory factor 1 precursor	*WIF1*	NM_007191	204712_at

613	WNT1 inducible signaling pathway protein 1 isoform 1 precursor	*WISP1*	NM_003882	211312_s_at
614		*WISP1*	NM_003882	206796_at
615	WNT1 inducible signaling pathway protein 2 precursor	*WISP2*	NM_003881	205792_at
616	wingless-type MMTV integration site family, member 1	*WNT1*	NM_005430	208570_at
617	wingless-type MMTV integration site family, member 10A	*WNT10A*	NM_025216	223709_s_at
618	wingless-type MMTV integration site family, member 10B	*WNT10B*	NM_003394	206213_at
619	wingless-type MMTV integration site family, member 11	*WNT11*	NM_004626	206737_at
620	wingless-type MMTV integration site family, member 16	*WNT16*	NM_016087	221113_s_at
621		*WNT16*	NM_016087	224022_x_at
622	wingless-type MMTV integration site family, member 2	*WNT2*	NM_003391	205648_at
623	wingless-type MMTV integration site family, member 2B	*WNT2B*	NM_004185	206459_s_at
624		*WNT2B*	NM_004185	206458_s_at
625	wingless-type MMTV integration site family, member 3	*WNT3*	NM_030753	221455_s_at
626		*WNT3*	NM_030753	231743_at
627		*WNT3*	NM_030753	229103_at
628	wingless-type MMTV integration site family, member 4	*WNT4*	NM_030761	208606_s_at
629		*WNT4*	NM_030761	230751_at
630		*WNT4*	NM_030761	1556689_a_at
631	Wingless-type MMTV integration site family, member 5A	*WNT5A*	NM_003392	205990_s_at
632		*WNT5A*	NM_003392	213425_at
633	Wingless-type MMTV integration site family, member 5B	*WNT5B*	NM_030775	223537_s_at
634		*WNT5B*	NM_030775	221029_s_at
635	wingless-type MMTV integration site family, member 6	*WNT6*	NM_006522	71933_at
636		*WNT6*	NM_006522	221609_s_at
637		*WNT6*	NM_006522	222086_s_at
638		*WNT6*	NM_006522	221608_at
639	wingless-type MMTV integration site family, member 7A	*WNT7A*	NM_004625	210248_at
640	wingless-type MMTV integration site family, member 7B	*WNT7B*	NM_058238	217681_at
641		*WNT7B*	NM_058238	238105_x_at
642	wingless-type MMTV integration site family, member 8A	*WNT8A*	NM_058244	224259_at
643	wingless-type MMTV integration site family, member 8B	*WNT8B*	NM_003393	207612_at
644	wingless-type MMTV integration site family, member 9A	*WNT9A*	NM_003395	1553045_at
645		*WNT9A*	NM_003395	230643_at
646	wingless-type MMTV integration site family, member 9B	*WNT9B*	NM_003396	1552973_at
647	Wilms tumor 1 isoform A	*WT1*	NM_000378	216953_s_at
648		*WT1*	NM_000378	206067_s_at

649	zinc finger, RAN-binding domain containing 1 protein	*ZRANB1*	NM_017580	225131_at
650		*ZRANB1*	NM_017580	225138_at
651		*ZRANB1*	NM_017580	225130_at

Supplementary 3.3: List of 651 WNT related probe-sets for HG-U133 Plus2.0