Investigating *HLXB9* as a biomarker in cancer

A thesis submitted for the Degree of Doctor of Philosophy

By Temitayo Olajumoke Owoka

Department of Life Sciences

College of Health and Life Sciences

Brunel University

July 2016

Declaration

I hereby declare that the research presented in this thesis is my own work, except where otherwise specified, and has not been submitted for any other degree.

Temitayo O Owoka

Abstract

A biomarker is a measurable biological characteristic that can be evaluated as an indicator of normal (physiological) or abnormal (pathogenic) processes. In cancer research, there remains a need for the identification of new biomarkers that can be used to close the gaps in the current understanding of cancer development, progression and treatment.

HLXB9 is a homeobox gene located at 7q36. It encodes a transcription factor important in embryonic development. Its accurate regulation is significant in the organogenesis of the endodermal germ layer particularly in the development of the pancreas. After development, its expression is downregulated in the majority of adult tissues. Recently, aberrant expression of *HLXB9* has been found in certain cancers such as hepatocellular carcinoma, testicular cancer, pancreatic cancer and leukaemia.

The location of genes and chromosomes in the nuclei of healthy human cells has been shown to be non-random, therefore understanding the mechanisms that regulate nuclear genome organisation is important in understanding of cancer biology in cases where genes are relocated. The nuclear localisation of genes as a biomarker of tumour development in cancer is a relatively new but promising field in cancer research. Previous research by our group found overexpression of *HLXB9* corresponded to an altered positioning of this gene in the nucleus of paediatric leukaemia patients harbouring the translocation t(7;12)(q36;p13).

In this project, *HLXB9* was evaluated as a biomarker in cancer development. In the first study, a new dual colour probe for the detection of the t(7;12)(q36; p13) was validated by fluorescence *in situ* hybridisation (FISH) in leukaemia patient samples previously described as harbouring the translocation.

The expression of *HLXB9* was then analysed by RT-PCR in 48 patients diagnosed with various haematological disorders. 25% of patients analysed expressed *HLXB9*. Additionally, *HLXB9* expression in leukaemia was found in patients with normal copies of chromosome 7 suggesting *HLXB9* expression can occur independently of chromosome 7 abnormalities. An attempt was

made to evaluate the link between *HLXB9* expression and its nuclear localisation in these patients.

Four online databases were interrogated to identify cancer types and subtypes that exhibit differential *HLXB9* expression. *HLXB9* expression was not altered in the majority of cancer cases investigated. However, aberrant *HLXB9* expression was found in cancer types not previously reported as showing differential *HLXB9* expression such as kidney cancer, lung cancer and endometrial cancer. The identification of aberrant expression of *HLXB9* in these cancer types provides a new avenue for research into understanding cancer development and progression in these tumour types.

Finally, the expression of *HLXB9* was analysed in four breast cancer cell lines by quantitative RT-PCR and immunofluorescence. Additionally, the prognostic significance of *HLXB9* expression was evaluated in publicly available breast cancer survival databases (Kaplan Meier plotter and BreastMark).

Altogether, the findings emerging from this thesis work show that, although the potential for *HLXB9* to be used as biomarker is appealing, further work is required to confirm the value of this biological parameter in the diagnosis, prognosis and progression of cancer.

Acknowledgement

I would like to express a special gratitude to my Principal Supervisor, Dr Sabrina Tosi for her enduring support, encouragement and guidance. Your support was a constant source of motivation to me. I also owe a debt of gratitude to my Secondary Supervisor, Dr Amanda Harvey, your help and guidance was incredibly valuable.

I would also like to thank Dr Chris Eskiw, Dr Helen Foster, Dr Joanna Bridger, and Dr Ines Castro. Special thanks to my fellow PhD students (past and present) especially Dr Abdulbasit Naiel and Chiara Papucci; no one else could really fully understand the process and thanks for the laughs.

To my family and friends, it could not have happened without you. I dedicate this work to my father, Thompson, my mother Mojisola and my brothers-Doyin, Femi, Siji and Sanmi. Your constant support was invaluable, your advice made sure I persevered to the end, your companionship made the journey immensely enjoyable. You give me a tremendous amount of hope and I thank you for your unconditional love.

Table of contents

List of figures

List of tables

List of abbreviations

-7	Monosomy of chromosome 7
aa	Amino acid
add	Addition
ALL	Acute lymphoid leukaemia
AML	Acute myeloid leukaemia
ANGPT1	Angiopoietin 1
bp	Base pair
BRCA 1/2	Breast cancer susceptibility gene 1 or 2
BSA	Bovine Serum Albumin
cDNA	Complementary DNA
CEBPA	CCAAT enhancer binding protein alpha
CLL	Chronic lymphoid leukaemia
CLP	Common myeloid progenitor
CML	Chronic myeloid leukaemia
CMML	Chronic myelomonocytic leukaemia
CMP	Common lymphoid progenitor
CO_2	Carbon dioxide
CS	Currarino syndrome
C_t	Cycle threshold
CUX1	Cut-Like Homeobox 1
DAPI	4',6-Diamidino-2-Phenylindole
del	Deletion
der	Derivative
DIG	Digoxigenin
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
dUTP	2'-deoxyuridine 5'-triphosphate
EBV	Epstein Barr Virus
EDTA	Ethylenediaminetetraacetic acid
ER	Estrogen receptor
ETV6	Ets Variant 6
EZH2	Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit
FAB	French- American- British classification of AML
FISH	Fluorescent in situ hybridisation
FITC	Fluorescein isothiocyanate
FLT3	Fms Related Tyrosine Kinase 3
GATA	GATA Binding Protein
HER2	Human epidermal growth factor receptor
HLXB9	Homeobox gene B9 also known as *HB9* or *MNX1*
HMEC	Human mammary epithelial cells
IF	Immunofluorescence
ins	Insertion
inv	Inversion
Ki 67	Ki- 67 protein
KIT	Kit Oncogene
LB	Luria Bertani

LT-HSC	Long term hematopoietic stem cell
MDS	Myelodysplatic syndrome
MEN1	Menin 1
MLL3	Myeloid/Lymphoid or Mixed-Lineage Leukaemia 3
MNX1	Motor Neuron And Pancreas Homeobox 1
MPD	Myeloproliferative neoplasms
MPP	Multipotent progenitor blood cells
mRNA	Messenger RNA
NPM1	Nucleophosmin 1
nsd	No significant difference
p-arm	Short arm of chromosome
PAC	P1- derived artificial chromosome
PBS	Phosphate Buffer Saline
PFA	Paraformaldehyde
PGTER2	Prostaglandin E2 receptor
PHA	Phytohaemagglutinin
PR	Progesterone receptor
PSA	Prostate-specific antigen
q-arm	Long arm of chromosome
qRT-PCR	Quantitative PCR
RARA	Retinoic Acid Receptor, Alpha
RNA	Ribonucleic acid
RPM	Revolutions per minute
RT-PCR	Reverse transcriptase-polymerase chain reaction
RUNX1	Runt-related Transcription Factor1
SDS	Sodium Dodecyl Sulfate
siRNA	Small interfering RNA
SSC	Saline-Sodium Citrate
ST-HSC	Short term hematopoietic stem cell
t	Translocation
t-AML	Therapy related AML
TBE	Tris-Borate-EDTA
TE	TRIS-EDTA
TNM	Tumour size, node involvement and metastasis system
WCP	Whole chromosome paint

Chapter 1: General Introduction

A section of this chapter was published as part of an article in Biomarker Research ISSN: 2050-7771

Tosi S, Kamel YM, Owoka T, Federico C, Truong T, Saccone S (2015). Paediatric acute myeloid leukaemia with t(7;12)(q36;p13) rearrangement: a review of the biological and clinical management aspects. Biomark Res 1–11. doi: 10.1186/s40364-015-0041-4

Introduction

1.1. Cancer biomarkers

A biomarker is a measurable biological characteristic that can be evaluated as an indicator of normal (physiological) or abnormal (pathogenic) processes (Strimbu and Tavel 2011; Henry and Hayes 2012). Therefore, biomarkers are used to objectively differentiate unaffected individuals from affected individuals (patients). Apart from the diagnosis of cancer, biomarkers can also be used to identify risk factors that predispose to cancer development, judge patients' prognosis, monitor disease progression (relapse or metastasis) and evaluate response to therapy (Henry and Hayes 2012; Burke 2016). There are various types of biomarkers. They include cellular/anatomical markers such as histological changes and molecular markers such as nucleic acids, genes, proteins, antibodies and gene expression profiles (Henry and Hayes 2012; Burke 2016).

In cancer, the common ways in which genetic biomarkers are routinely used in clinics include: the identification of germ line *BRCA1/2* mutations as a risk factor in the development of breast and ovarian cancer; the *BCR/ABL* fusion gene to confirm diagnosis and monitor disease status in chronic myeloid leukaemia or acute lymphoblastic leukaemia; HER2 amplification to predict response to trastuzumab in the treatment of breast or gastric tumours; elevated PSA protein level after treatment for prostate cancer indicating disease relapse (Henry and Hayes 2012; Burke 2016). An important aim of cancer research is the identification of biomarkers, primarily the detection of biomarkers that can be targeted for therapy.

1.2. The *HLXB9* gene

Homeobox gene HB9 (*HLXB9*) also known as *MNX1* (Motor Neuron and Pancreas Homeobox gene 1) encodes a 401 amino acid (aa) protein and maps to the long arm of chromosome 7 at q36.3 (Harrison et al. 1994; Ross et al. 1998; Hagan et al. 2000). *HLXB9* was first identified and subsequently cloned in human B lymphocytes by Harrison and colleagues in 1994 (Harrison et al. 1994). They also identified *HLXB9* mRNA expression in a variety of adult tissues/organs including pancreas, small intestine, tonsils, colon and CD34 positive bone marrow cells (Harrison et al. 1994).

HLXB9 is encoded by three exons. Exon one encodes the translation start site, two glycine stretches (amino acid (aa) 42-29 and aa100-105 respectively) and a polyadenine repeat of variable length (Harrison et al. 1994; Hagan et al. 2000; Crétolle et al. 2008; Garcia-Barceló et al. 2009; Wildenhain et al. 2012). Exons 1 and 2 together encode an interspecies conserved 82 amino acid sequence unique to *HLXB9* (aa 161-242) while exon 2 and 3 encode the 60 aa homeodomain (aa 243-302) (Harrison et al. 1994; Hagan et al. 2000; Crétolle et al. 2008; Garcia-Barceló et al. 2009; Wildenhain et al. 2012). The homeodomain in *HLXB9* is folded into three alpha helices in a helix turn helix motif, which facilitates both its DNA and protein binding function. (Hagan et al. 2000; Nagel et al. 2005; Garcia-Fernàndez 2005; Wildenhain et al. 2012)

HLXB9 belongs to the Antennapedia (ANTP) superclass of homeobox (hbx) genes. The identification of another *MNX1* orthologue called *MNR2* in chicken has led to a sub classification of *HLXB9* into its own gene family (Holland et al. 2007; Wendick et al. 2004; Odden et al. 2002). The *MNX* gene family includes one gene in mammals: *MNX1/HB9;* two genes in chicken: *MNX1/HB9* and *MNX2/MNR2;* and three genes in zebra fish: *MNX1/HB9*, *MNX2a/MNR2a* and *MNX2a/MNR2b* (Holland et al. 2007; Wendick et al. 2004; Odden et al. 2002). Orthologues of both *HB9* and *MNR2* genes have also been found in drosophila (Holland et al. 2007; Wendick et al. 2004; Odden et al. 2002).

Homeobox genes mainly function as transcription factors and regulate gene expression by binding to DNA and/or other proteins to either promote or repress transcription (Holland and Takahashi 2005; Holland 2013). They play an essential role in embryonic development by directing differentiation and the spatial positioning of cells of the germ layer into the mesoderm, endoderm and ectoderm (Garcia-Fernàndez 2005; Holland et al. 2007; Holland 2013). After embryonic development, they continue to be expressed in a tissue specific manner and play a pivotal role in regulating cell division, cell maturation and apoptosis (Garcia-Fernàndez 2005; Holland et al. 2007; Holland 2013). Abnormal expression of homeobox genes have been implicated in the development of different types of cancers such as leukaemia, breast cancer, renal cancer (Chen and Sukumar 2003; Svigen and Tonissen 2003; Nagel et al. 2005; Dormoy et al. 2012).

1.3. *HLXB9* in foetal development

Most of the research carried out to elucidate the function of *HLXB9* is primarily on its role in foetal development. *HLXB9* is known to play an essential role in the development of endoderm tissue derivatives, which forms the digestive and respiratory tracts (Harrison et al. 1999; Fonseca et al. 2008; Sherwood et al. 2009).

HLXB9 expression has been examined by in situ hybridisation in two studies on human embryo sections. The first study by Ross and colleagues found *HLXB9* was expressed in the sacral region of the embryo (Ross et al.1998). The second study by the same group found *HLXB9* expressed in the neural tube, the anterior horns (site of motor neuron development), the hindbrain, the pharynx, the oesophagus, the stomach, the pancreatic bud and the spinal cord at various stages of embryonic development (Hagan et al. 2000).

Other studies to elucidate the role of *HLXB9* in foetal development have been carried out on mouse embryos. Through these studies, *HLXB9* was found to be expressed in a dorsal-ventral gradient throughout the anterior-posterior axis of the endoderm (Li et al. 1999; Harrison et al. 1999; Sherwood et al. 2009). From these studies, *HLXB9* expression has been found to be crucial for the development of the pancreas and motor neurons (Li et al. 1999; Arber et al. 1999; Thaler et al. 1999; Harrison et al. 1999; Sherwood et al. 2009).

Two *HLXB9* mouse models have been developed to understand *HLXB9* role in foetal development (Li et al. 1999; Arber et al. 1999; Harrison et al. 1999). Heterozygous $HLXB9^{+/-}$ mice show no abnormal phenotype, however, their resulting homozygous offspring died at or soon after birth (Li et al. 1999; Harrison et al. 1999). The homozygous $HLXB9^{-/-}$ offspring were also smaller and showed abnormal spinal cord curvature compared to their littermates suggesting a role for *HLXB9* in spinal development (Li et al. 1999; Arber et al. 1999). Additionally, in homozygous $HLXB9^{-/-}$ mice, there were perturbations to pancreatic and motor neuron development. In these mice, the dorsal lobe of the pancreas failed to develop, and the pancreatic islets (islets of Langerhans) were abnormal (Harrison et al.1999; Li et al.1999; Li et al. 2001; Hagan et al. 2000). Additionally, defects in the migration, subtypes and projection of motor axons of the motor neurons are also a phenotype displayed in homozygous *HLXB9* mice (Arber et al. 1999; Thaler et al.1999).

Finally, the pattern of expression of *HLXB9* in mouse and human embryo has been investigated and compared (Hagan et al. 2000). In the study by Hagan and colleagues, they found similar to mouse *HLXB9*, human *HLXB9* is expressed in motor neuron progenitors. However, they also found a variety of differences in the pattern of *HLXB9* expression between the two species during the development of the hindgut and the pancreas. This difference in pattern of *HLXB9* expression might explain why heterozygous mouse show no phenotype while humans with loss of function mutation of *HLXB9* develop Currarino syndrome.

1.4. Currarino syndrome (CS)

In humans, loss of function mutation/(s) or micro deletions of *HLXB9* can be autosomal dominantly inherited or sporadically occur causing Currarino syndrome (CS) (Belloni et al. 2000; Hagan et al. 2000; Garcia-Barceló et al. 2009; Zu et al. 2011). Therefore, haploinsufficient expression of *HLXB9* can be described as the cause of most cases of CS. CS is a rare medical condition (Yoshida et al. 2010; Zu et al. 2011). A medical diagnosis of CS is characterised by three main phenotypes namely partial sacrum agenesis, presacral mass and anorectal malformations (Belloni et al. 2000; Hagan et al. 2000; Garcia-Barceló et al. 2009; Zu et al. 2011). However, not all CS cases present with all three phenotypes and not all individuals with mutations in *HLXB9* are symptomatic of CS. This suggests the possibility of a mechanism/(s) by which haploinsufficient *HLXB9* expression is compensated for by other genes during development. (Belloni et al. 2000; Hagan et al. 2000; Garcia-Barceló et al. 2009; Zu et al. 2011). However, the responsible compensatory mechanism/(s) is yet to be identified. Other phenotypes which may be associated with CS include neural tube defects, urological abnormalities such as duplication of the kidney and ureter, horseshoe kidney, neurogenic bladder and gynaecological abnormalities such as bicornuate uterus and bifid clitoris (Belloni et al. 2000; Zu et al. 2011). These findings suggest *HLXB9* has a role in urological development particularly in the organogenesis of the kidney.

There have been two large scale studies to identify *HLXB9* mutations in CS patients(Hagan et al. 2000; Crétolle et al. 2008). In 2000, Hagan and colleagues investigated mutations of *HLXB9* in 21 cases of familial CS and seven cases of sporadically occurring CS. They found 95% of familial CS patients and 29% of sporadically occurring CS patients had mutations in *HLXB9* (Hagan et al. 2000). The second study by

Crétolle and colleagues screened 20 cases of familial CS and 30 cases of sporadically occurring CS. They found the same incidence of *HLXB9* mutations in familial CS patients by Hagan and colleagues and a slight reduction in the incidence of *HLXB9* mutation (23%) in sporadic occurring CS (Crétolle et al. 2008).

Studies to characterise the type of *HLXB9* mutations found in CS patients found a majority of mutations in *HLXB9* occur in exon 1 or exon 2 (exon 2 contains the homeodomain) (Belloni et al. 2000; Hagan et al. 2000; Crétolle et al. 2008; Garcia-Barceló et al. 2009; Cuturilo et al. 2016). Types of *HLXB9* mutations observed in these patients include: nonsense, missense, frameshift (insertion or deletion), splice site, other point mutations (not yet defined) (Belloni et al. 2000; Hagan et al. 2000; Crétolle et al. 2008; Garcia-Barceló et al. 2009; Cuturilo et al. 2016). These mutations result in the loss of a functional HLXB9 protein.

Interestingly, deletions in the long arm of chromosome 7 (7q) is a cytogenetic abnormality found in some CS patients (Hagan et al. 2000; Crétolle et al. 2008; Cuturilo et al. 2016).

1.5. *HLXB9* constitutional mutations and diabetes

After foetal development, b-cells in the islets of Langerhans are the only pancreatic cell type that retain expression of *HLXB9* (Li et al. 1999; Harrison et al. 1999). Pancreatic b-cells are involved in the storage and release of insulin. Therefore, it is not surprising that loss of function mutations of *HLXB9* have now been reported in cases of neonatal diabetes, a disease of insulin deficiency (Flanagan et al. 2014; Pan et al. 2015). Additionally, HLXB9 has also been found to directly regulate genes that regulate cellular insulin such as *VSNL1* and *PLD1* (Shi et al. 2013).

1.6. *HLXB9* constitutional mutations and cancer

One of the observed phenotypes in CS is the presence of a presacral mass such as anterior meningocele, Lipomeningocele, teratoma, dermoid cyst and lipoma (Belloni et al. 2000; Hagan et al. 2000; Crétolle et al. 2008; Garcia-Barceló et al. 2009; Yoshida et al. 2010; Zu et al. 2011). Teratomas, a type of germ cell tumour accounts for 25-40% of presacral mass observed in CS (Dirix et al. 2015). In a study by Crétolle and colleagues, 88% of CS patients with intragenic *HLXB9* mutations developed benign teratomas compared to 50% of CS patients without *HLXB9* mutations (Crétolle et al. 2008). This indicates the

possibility of *HLXB9* functioning as a tumour suppressor. However, only a very small proportion of CS teratomas have been observed to undergo malignant transformation in clinic (nine cases) (Yoshida et al. 2010). Additionally, only two cases of metastasis to other sites have been observed and both patients showed brain metastasis. This finding suggests a possible association for *HLXB9* mutations as a single cause of cancer development. This also indicates that *HLXB9* acts as a modulator of cancer development in conjunction with other genetic factors.

1.7. *HLXB9* in pancreatic cancer

Given the role *HLXB9* plays in pancreatic development, it is not surprising that aberrant expression has been found in pancreatic cancer. *HLXB9* was found to downregulated in the squamous subtype of pancreatic ductal adenocarcinoma (Bailey et al. 2016). Downregulation of *HLXB9* was associated with a loss of differentiation in this pancreatic tumour subtype.

Conversely, upregulation of *HLXB9* has been found in insulinomas, a cancer of pancreatic b-cells (Shi et al. 2013; Desai et al. 2014; Desai et al. 2015). In insulinomas, *HLXB9* has been found to be crucial in the regulation of apoptosis of b-cells (Shi et al. 2013). In the study by Shi and colleagues, they found that *HLXB9* was able to regulate apoptosis by binding to *Menin (MEN1).* However, when *MEN1* was deficient either due to loss of function mutations or experimental knockdown, *HLXB9* becomes upregulated and becomes unable to induce apoptosis.

In a subsequent study by the same group to understand the mechanism by which *HLXB9* becomes deregulated in insulinomas, they found that serine/threonine kinase glycogen synthase kinase 3 (GSK-3B) phosphorylates overexpressed *HLXB9* (Desai et al. 2014). The phosphorylation of *HLXB9* was shown to result in a loss of its pro apoptotic function. They also showed that loss of *MEN1* results in an increase of phosphorylated *HLXB9*.

A final study was carried out by the group to identify the targets and binding partners of phospho-HLXB9 in insulinomas (Desai et al. 2015). They found that phosphorylated HLXB9 was able to bind to the promoter of Casitas B-lineage lymphoma b (Cblb), an E3 ubiquitin ligase to repress its expression (Desai et al. 2015). They also found the downregulation of CBIB results in an upregulation of CMET, a tyrosine kinase because CBIB is unavailable to target CMET for degradation. Additionally, they found knocking

down *HLXB9* expression directly decreased the expression of CMET. They proposed that CMET overexpression because of *HLXB9* overexpression could be a mechanism by which *HLXB9* contributes to oncogenesis in insulinomas. This series of studies was the first comprehensive investigation into the function of *HLXB9* in cancer development and demonstrated that accurate regulation of *HLXB9* is crucial in the pancreas.

1.8. *HLXB9* in leukaemia

Leukaemia is a progressive disorder of haematological cells characterised by uncontrolled proliferation and a block in differentiation (Chang and Weiss 2007; Bomken and Josef Vormoor 2009). This disorder results in the accumulation of immature 'blast' cells in the bone marrow and circulation of these 'blast' cells in the peripheral blood. There are two main types of leukaemia: acute leukaemia and chronic leukaemia. In acute leukaemia, a majority of stem or progenitor blood cells are unable to give rise to mature functional blood cells whereas in chronic leukaemia, a smaller proportion of stem or progenitor blood cells are unable to give rise to mature blood cells. Because of the difference in proportion of stem cell affected, chronic leukaemia has a better prognosis than acute leukaemia.

In the UK, leukaemia is the 11th most commonly diagnosed cancer and the 9th most common cause of cancer deaths (Cancer Research, UK). Acute leukaemia is classified by the lineage of the blood cells affected into acute myeloid leukaemia (AML) and acute lymphoid/lymphoblastic leukaemia (ALL). Chronic leukaemia is also classified by the linage of the affected blood cells into chronic myeloid leukaemia (CML) and chronic lymphoid/lymphoblastic leukaemia (CLL). According to Cancer Research UK, these four main types of leukaemia account for 85% of leukaemia cases diagnosed in the UK in 2013 (as shown in figure 1.1).

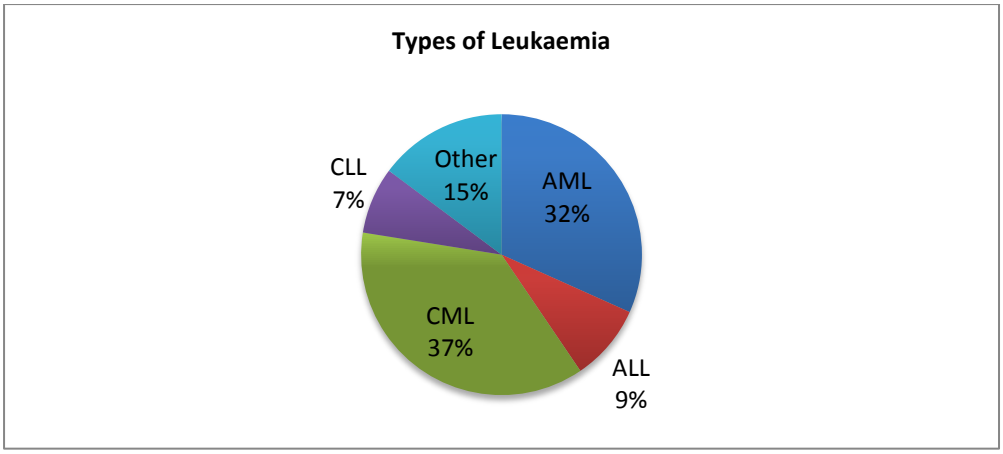

Figure 1.1: Leukaemia types by proportion of cases diagnosed in the UK in 2013 (Cancer Research, UK). AML, acute myeloid leukaemia, ALL: acute lymphoid leukaemia, CML, chronic myeloid leukaemia, CLL, chronic lymphoid leukaemia, other types of leukaemia diagnosed in the UK in 2013 include chronic myelomonocytic leukaemia (CMML) and hairy cell leukaemia.

1.8.1. Normal blood cell development

Mature blood cells have a limited life span so need to be produced throughout life (Chang and Weiss 2007; Mazo et al. 2011). The process by which blood cells are produced is called haematopoiesis (Chang and Weiss 2007; Mazo et al. 2011). In the early weeks of foetal development, haematopoiesis occurs in the embryo's yolk sac, however, by the third month of foetal development, the liver becomes the main site of haematopoiesis (Chang and Weiss 2007; Mazo et al. 2011; Jagannathan-Bogdan and Zon 2013). By the fourth month of embryonic development, the bone marrow is the primary site of haematopoiesis (Chang and Weiss 2007; Mazo et al. 2011; Jagannathan-Bogdan and Zon 2013).

The process of blood cell development involves the division and differentiation of pluripotent hematopoietic stem cells (HSCs) in the bone marrow (Kondo 2010; Kosan and Godmann 2016). Upon activation, long term quiescent HSCs (LT-HSC) with unlimited self-renewal potential give rise to short term HSCs (ST-HSCs) with limited self-renewal potential (Chang and Weiss 2007; Kondo 2010; Bruin et al. 2014; Kosan and Godmann 2016). These ST-HSCs give rise to multipotent progenitors blood cells (MPPs), which give rise to either common lymphoid or myeloid progenitor cells (CMP or CLP). CMP or CLP undergo subsequent cell division and differentiation to give rise to the mature blood cells of myeloid and lymphoid linages found in the peripheral blood (as shown in figure 1.2) (Chang and Weiss 2007; Kondo 2010; Bruin et al. 2014; Kosan and Godmann 2016).

For this process to occur, HSCs and subsequent progenitor cells respond preferentially to molecules such as growth factors, cytokines, transcription factors, signalling pathway proteins, colony stimulating factors

that control differentiation and proliferation to give rise to mature blood cells of myeloid and lymphoid cell linages found in the peripheral blood (Chang and Weiss 2007; Mazo et al. 2011).

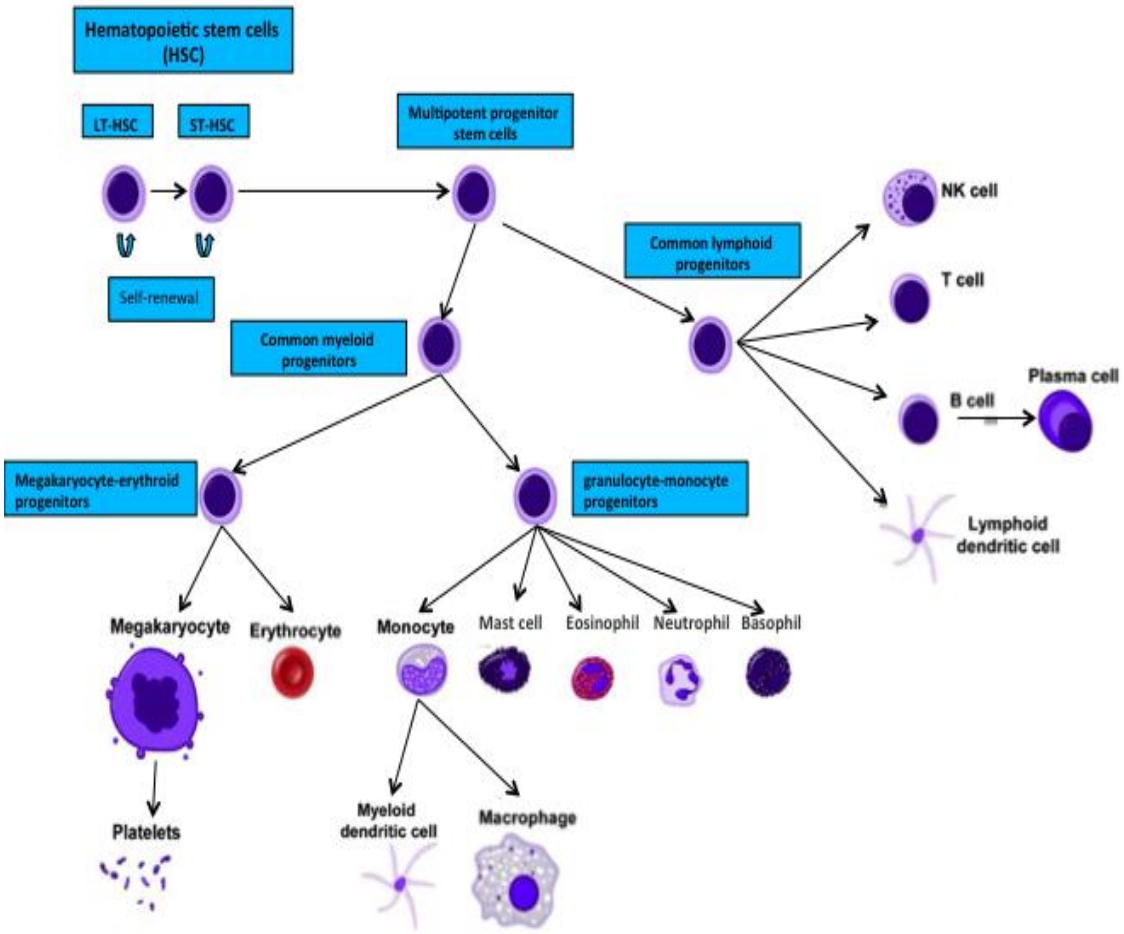

Figure 1.2: Schematic representation of haematopoiesis. A long term hematopoietic stem cell (LT-HSC) gives rise to short term hematopoietic stem cells (ST-HSCs), which undergo subsequent division and differentiation to give rise to lymphoid, and myeloid progenitor cells, which divides and differentiates into mature blood cells. Adapted from Bruin et al. 2014.

1.8.2. Commonly mutated genes in haematopoiesis

Mutation in crucial genes involved in haematopoiesis such as *FMS-related tyrosine kinase 3 (FLT-3), CCAAT enhancer binding protein alpha (CEBPA) and KIT* tyrosine receptor kinase are some of the commonly mutated genes found in leukaemia patients (Gregory et al. 2009; Gulley et al. 2010; Betz and Hess 2010; Ohgami and Arber 2015; Yohe 2015).

Both mutations in genes that confer a proliferative advantage such as *FLT, KIT, KRAS, NRAS* and mutations in genes that block differentiation such as *CEBPA, NPM1* have been observed in leukaemia patients (Gregory et al. 2009; Gulley et al. 2010; Betz and Hess 2010; de Rooij et al. 2015; Yohe 2015). The presence of mutations in these genes are used as biomarkers of patient prognosis (Gregory et al. 2009; Gulley et al. 2010; Yohe 2015).

1.8.3. Leukaemia development

The aetiology of leukaemia is unknown; however, risk factors associated with the development of leukaemia include: pre-existing genetic disorders such as Down's syndrome, Fanconi anaemia; exposure to environmental factors such as ionising radiation, chemical or cytotoxic agents; and genetic instability such as abnormal DNA repair or gene mutations (Greaves 1999; Langmuir et al. 2001; Bomken and Josef Vormoor 2009; Eden 2010). A hypothesis for leukaemia development is that it is a 'two hit hypothesis', where the first genetic mutation occurs in stem cell in utero termed the 'first hit 'and a subsequent genetic mutation occurs after birth due to exposure to one of the aforementioned risk factors termed the 'the second hit' (Langmuir et al. 2001; Estey and Döhner 2006; Bomken and Josef Vormoor 2009; Eden 2010). The two hit model has been validated in literature for most cases of Infant Acute Lymphoid Leukaemia (ALL) and in some cases of AML (Bomken and Josef Vormoor 2009; Eden 2010).

The two hit hypothesis is thought to contribute to cancer development via two mechanisms: the 'cancer stem cell model' or the 'stochastic model' (Nguyen et al. 2012). The cancer stem cell model (also termed the hierarchal model of cancer development) proposes the existence of an abnormal hematopoietic stem cell termed the 'cancer stem cell'. Cancer stem cells are able to undergo subsequent cell division to produce additional cancer stem cells thereby initiating and supporting cancer development (Nguyen et al. 2012; Terwijn et al. 2014). In fact the existence of cancer stem cells were first observed in AML; the $CD34^{+}/CD\ 38^{low/-}$ cells isolated from AML patients were found to be tumour initiating in immunocompromised mice (Nguyen et al. 2012; Terwijn et al. 2014). This model of cancer development is also thought to explain the frequency of relapse observed in AML patients because if any of the cancer stem cell remains after treatment they are able to support tumour growth causing relapse (Terwijn et al. 2014).

The stochastic (random) model of cancer development proposes that any of the cells in the population can acquire a mutation or epigenetic change to initiate tumour development (Nguyen et al. 2012). In ALL, cancer initiating cells have been observed at different differentiation stages suggesting that ALL is developed in a stochastic manner (Purizaca et al. 2012; Mcclellan and Majeti 2013). Although there have been numerous attempts to identify cancer stem cells in ALL, there remains no agreement on the existence of ALL cancer stem cells(Mcclellan and Majeti 2013).

1.8.4. Myeloid Malignancies

Myeloid malignancies are disorders that affect myeloid hematopoietic stem or progenitor blood cells. Myeloid malignancies can be chronic or acute. Types of chronic myeloid malignancy include myeloproliferative neoplasms (MPN), myelodysplastic syndromes (MDS) and chronic myeloid leukaemia (CML). Acute myeloid malignancies are classified as acute myeloid leukaemia. The incidence and five-year relative survival rate of myeloid malignancies diagnosed in Europe between 1995 and 2002 are described in table 1.1. From the table, it is evident that patients diagnosed with AML, MPN and MDS have very poor prognosis, therefore finding biomarkers that can be targeted in these malignancy types is crucial.

Myeloid Malignancies	Case per 100,000 diagnosed in Europe	5-year relative survival rate in Europe (%)
Acute Myeloid Leukaemia (AML) and related precursor neoplasms	3.7	19%
Acute Promyelocytic Leukaemia (AML with t(15;17)	0.1	66
AML, other and not otherwise defined (NOS)	3.4	17
Myeloproliferative Neoplasms (MPN)	3.1	62%
Myelodysplastic syndrome (MDS)	1.5	29%
MDS and MPD	1.8	27%
Chronic Myeloid Leukaemia (CML)	1.2	44%

Table 1.1: Incidence and 5-year survival rate of Myeloid Malignancies diagnosed in Europe between 1995 and 2002 (Visser et al. 2012).

Myeloid malignancies occur as a result of genetic and epigenetic changes that affect either or both the proliferation and differentiation of myeloid blood cells. In MDS, the differentiation of myeloid stem or progenitor cells into matured cells becomes impaired whilst in MPD, there is an abnormal increase in proliferation of myeloid stem or progenitor cells (Foucar 2009; Cazzola et al. 2011). A proportion of MDS/ MPD patients have been observed to undergo further malignant transformation and progress to AML (Foucar 2009; de Rooij et al. 2015).

In AML, both the differentiation and proliferation of myeloid stem/ progenitor blood cells are impaired (de Jonge et al. 2011). Myeloid disorders are characterised by the presence of chromosomal abnormalities, which have prognostic significance. Common chromosomal abnormality found in MDS or AML include del(5q), del(7q), monosomy 7, trisomy 8, 11q23 (Nimer 2008, Visser et al. 2012, Murati et al. 2012; Van Etten and Shannon 2004).

1.8.5. Acute Myeloid Leukaemia (AML)

AML is a heterogeneous disease of myeloid progenitor cells in the bone marrow. A diagnosis of AML has a poor prognosis with a 5 year survival rate of 30-40% (Langmuir et al. 2001; Estey and Döhner 2006). It is the most commonly diagnosed myeloid malignancy in western countries with an annual incidence of 2 cases per 100,000 (Estey and Döhner 2006; Chang and Weiss 2007). It is more commonly diagnosed in adults than children with a peak incidence age of 65 years; when diagnosed in children it usually occurs before the age of 2 (Zwaan and Kaspers 2004; Estey and Döhner 2006; Chang and Weiss 2007; Vardiman et al. 2009). Children diagnosed with AML have a higher incidence of chromosomal abnormalities than adults diagnosed with AML (Mrozek et al. 2004).

- **Classification of AML**

The most commonly used systems for categorising AML diagnosis are the French-American-British (FAB) classification and the World Health Organisation (WHO) classification (Chang and Weiss 2007). The classification of AML into subtypes with the same biological characteristics has been significant for understanding of the mechanism of oncogenesis and informing patients prognosis (Segeren and Veer 1996; Chang and Weiss 2007; Vardiman et al. 2009; Gulley et al. 2010).

- ***The French-American-British (FAB) Classification* of AML**

In the 1970s, the FAB classification of AML based on the morphology and cytochemistry of leukaemia cells was developed (Bennett et al. 1976; Segeren and Veer 1996). AML was classified into eight subtypes by the different degrees of differentiation and cell linage affected (table 1.2).

FAB subtype	Description (name)	Commonly associated cytogenetic finding
M0	Myeloblastic leukaemia with minimal differentiation	Trisomy 11(Alseraye et al. 2011)
M1	Myeloblastic leukaemia with little differentiation	Trisomy 11 (Alseraye et al. 2011)
M2	Myeloblastic leukaemia with differentiation	t(8;21) (Klaus et al. 2004)
M3	Acute promyelocytic leukaemia (APL)- two variants M3h APL, hypergranular or M3v APL, microgranular	t(15; 17)(Allford et al. 1999)
M4/M4 eo	Acute myelomonocytic leukaemia (AMML) or M4 eo- AMML with dysplastic marrow eosinophils	t(16;16), inv(16), del(16q)(Delaunay et al. 2003)
M5	Acute monoblastic leukaemia (AMoL)- two types M5A poorly differentiated and M5 B- differentiated	t(9;11) (Kuipers et al. 2011), del(11q), t(11;19)(Tallman et al. 2004)
M6	Erythroleukaemia- 2 variants M6a AML with erythroid dysplasia M6b Erythroleukaemia	t(3;5) (Kwong 1998)
M7	Acute megakaryoblastic leukaemia (AMkL)	t(1;22)(Mercher et al. 2002)

Table 1.2: The FAB classification of AML showing subtypes and the most common chromosomal abnormality (cytogenetic) finding for each subtype (Rubin and Hansen 2012).

- ***World Health Organisation (WHO) classification of AML***

In 2001, the WHO integrated the morphology and cytogenetic characteristics of blast cells (already used in the FAB classification) with additional information on the molecular genetics and immunological characteristics of blast cells (Tallman et al. 2004; Chang and Weiss 2007; Vardiman et al. 2009). This review of the classification of AML was carried out to improve treatment and prognostic decisions. The list was updated in 2008 and categorised into four groups, which are shown in table 1.3 (Vardiman et al. 2009).

***AML with recurrent cytogenetic abnormalities*:** This group of AML patients have a de novo cytogenetic abnormality. The most common are translocations, which results in a fusion protein that contributes to leukaemogenesis.(Falini et al. 2010). There are eight categories under this classification: I. t(8;21)(q22;q22) II. inv(16)(p13;q22) or t(16;16)(p13;q22), del (16)(q22) III. t(15;17)(q22;q12) IV. t(9;11)(p22;q23) V. t(6;9)(p23;q34) VI. inv(3)(q21q26.2) or t(3;3)(q21;q26.2) t(1;22)(p13;q13) VII. Provisional entity: AML with mutated Nucleophosmin-1 NPM1 VIII. Provisional entity: AML with mutated CEBPA Category I-III are classified as AML regardless of blast counts and the other categories are classified as AML if the blast count is greater than 20%.	***AML with Myelodysplasia (MD) related changes:*** This group of AML patients fit into one or more of the following criteria below. I. A history of MD/MD neoplasm II. MD related cytogenetic abnormalities e.g-7/del7q, -5/ del5q, t(5;7)(q33;q11.2), t(5;17)(q33;p13) III. Multilinage dysplasia in bone marrow or peripheral blood smears in 50% of two or more cell linage.	***AML and myelodysplastic syndrome, therapy related:*** This group of patients develop chromosomal abnormalities following chemotherapy treatment with alkylating agents and topoisomerase inhibitors, most patients in this group have a worse prognosis than patients with de novo chromosomal abnormalities.	***AML, not otherwise specified:*** This group of patients are categorised has having AML after using morphologically and cytochemistry techniques but do not fall into the other 3 groups.

Table 1.3: WHO classification of AML showing the four groups of AML subtypes and the criteria for classification into each group. (Tallman et al. 2004; Vardiman et al. 2009)

Although older than the WHO classification, the FAB classification is still very useful in diagnosis, treatment and prognosis as it takes less time to carry out thereby facilitating quick decision making by clinical staff (Tallman et al. 2004; Chang and Weiss 2007; Vardiman et al. 2009).

1.8.6. Prognostic indicators in leukaemia

The single most important prognostic indicator used in leukaemia diagnosis and treatment is the presence of specific chromosomal abnormalities.

- **Chromosomal abnormalities in leukaemia**

Chromosomal abnormalities, which are changes to the normal chromosomal composition, are a recurrent finding in leukaemia. (Greaves 1999; Langmuir et al. 2001; Bomken and Josef Vormoor 2009; Eden 2010; Ebert 2011; Wiemels 2012). They are used as biomarkers to classify patient response to therapy into favourable, intermediate or adverse categories (Bertz and Hess 2010; Ferrarra and Schiffer 2013; Heng et al. 2006). For instance t(15;17) which fuses *PML-RARA* is successfully treated with trans-retinoic acid whilst t(11;17) which fuses *ZBTB16/RARA* does not respond well to trans-retinoic acid treatment (Mrózek et al. 2004; Hasle et al. 2007; Betz and Hess 2010; Ferrara and Schiffer 2013; Estey et al. 2014; Ohgami and Arber 2015).

Although the majority of AML patients have no chromosomal abnormality (55%), there remains a significant proportion of AML patients that have recurrent abnormal chromosome rearrangements (45%) (Betz and Hess 2010). The most common chromosomal abnormality found in AML patients and their impact on patient's outcome are summarised in table 1.4.

Risk	Patient outcome	Chromosomal abnormality
Low	Favourable	inv(16)
		t(16;16)
		t(8;21)
		t(15;17)
Medium	Intermediate	Normal Karyotype
		Trisomy 8
		t(9;11)
		Other chromosomal abnormality
High	Unfavourable	Monosomy 5*
		del(5q)*
		Monosomy 7*
		del(7q)*
		11q23
		Inv(3), t(3;3)
		t(6;9)
		t(9;22)
		t(11;17)
		Complex karyotype (>3 chromosomal abnormalities)

Table 1.4: Risk classification of common chromosomal abnormality found in AML patients. Patients outcome are determined based on patient response to therapy, remission and survival rate. (*Conflicting reports – some studies have sometimes classified as being associated with medium risk). (Mrózek et al. 2004; Hasle et al. 2007; Manola 2009; Betz and Hess 2010; Grimwade et al. 2013; Ferrara and Schiffer 2013; Estey et al. 2014; Ohgami and Arber 2015)

1.8.7. Chromosomal abnormalities and leukaemia development

Most normal human cells have a diploid karyotype (2n) comprised of 46 chromosomes, however, due to genome instability arising as a result of several essential cellular processes such as DNA damage repair and mitosis, chromosomal abnormalities may arise (Agarwal et al. 2006; Thompson et al. 2010; Bakhoum and Compton 2012). Chromosomal abnormalities can either be changes in the number of chromosomes (numerical abnormalities) or changes in the structure of chromosomes (structural abnormalities). In cancer, chromosomal abnormalities are referred to as 'acquired' because they are confined to the cancer cells. Chromosomal abnormalities may contribute to cancer development by a number of mechanisms including: disruption of normal gene expression due to the juxtaposition of genes with regulatory or enhancer elements; the deletion or inactivation of tumour suppressor genes; the activation of oncogenes; increase in copy number (amplification) or the formation of fusion genes that have oncogenic potential (Stephens et al. 2011).

Once an initial progenitor cell develops a chromosomal abnormality that confers an oncogenic advantage, that cell and its progeny are able to divide more rapidly and may gain additional chromosomal abnormalities that aid cancer development and progression (Stephens et al. 2011; Greaves and Maley 2012; Potapova et al. 2013).

Chromosomal abnormalities are detected by cytogenetic methods such as chromosome banding, fluorescent in situ hybridisation (FISH) or molecular methods such as polymerase chain reaction (PCR), Microarray, next generation sequencing (NGS) (Bain 2001; Mrózek et al. 2004; Talkowski et al. 2011; Choi et al. 2011; Shaffer et al. 2012; Braoudaki and Tzortzatou-Stathopoulou 2012).

o Structural chromosomal abnormalities and leukaemia development

Chromosomal abnormalities that result in the rearrangement of the structure of the chromosome are termed structural abnormalities. Structural chromosomal abnormalities are referred to as balanced if there is no loss or addition of genetic material or as unbalanced if there is a loss or additional of genetic material. Structural abnormality occur due to incorrect DNA damage repair or abnormal mitosis (Charames and Bapat 2003; Lieber et al. 2003; Stephens et al. 2011; Ganem and Pellman 2012; Shaffer et al. 2012). Some of the most common types of structural abnormalities observed in AML and the mechanism by which they contribute to cancer development are discussed in detail below.

❖ **Chromosomal deletions and leukaemia development**

This structural rearrangement results in the loss of genetic material, which can be the loss of whole arms of chromosomes or segments of chromosomes. Deletions of genetic material might lead to cancer development due to haploinsufficiency or loss of heterozygosity of tumour suppressor genes (Ebert 2011). Chromosomal deletions are more common in myelodysplastic syndrome (MDS) than in AML; common deletions in both disorders include del(5q) and del(7q) (Zhang and Rowley 2006; Ebert 2011; Braoudaki and Tzortzatou-Stathopoulou 2012; Ferrara and Schiffer 2013).

❖ **Chromosomal translocations and leukaemia development**

This structural rearrangement occurs when two chromosome exchange segments of genetic material (Nambiar and Raghavan 2011; Shaffer et al. 2012; Braoudaki and Tzortzatou-Stathopoulou 2012). Translocations are the most commonly occurring structural chromosomal abnormality in leukaemia. Additionally, most translocations observed in cancer are reciprocal. The resulting translocations are referred to as 'balanced' if no loss or gain of genetic material occurs or 'unbalanced' indicating loss or gain of genetic material (Shaffer et al. 2012).

Translocations initiate leukaemogenesis by two main mechanisms. The first mechanism is due to the disruption of normal gene expression by incorrect juxtaposition with regulatory or enhancer elements of other genes. For example, in CML patients, t(8;14) the proto-oncogene c-*MYC* on chromosome 8 is overexpressed due to juxtaposition with the *IgH* gene locus on chromosome 14 (Agarwal et al. 2006; Nambiar and Raghavan 2011). The second mechanism is by the formation of an oncogenic fusion protein. For example in AML patients, the formation of PML-RARA fusion protein due to t(15;17); PML-RARA represses the expression of genes involved in differentiation and apoptosis thereby promoting oncogenesis (Rabbitts 1994; Martens and Stunnenberg 2010; Nambiar and Raghavan 2011).

Apart from these mechanisms, other hypothesis for translocation related cancer development include a position effect mechanism of aberrant gene expression or inactivation of tumour suppressor genes (Ballabio et al. 2009; Harewood et al. 2010; Nambiar and Raghavan 2011).

❖ **Chromosomal inversions (inv) and leukaemia development**

This chromosomal rearrangement ensues when two breaks occurring within a chromosome are incorrectly repaired. The repair results in a wrong orientation of the chromosomal segment thereby changing the order of genes (Strachan and Read 2011). They are two types of inversions: paracentric

inversions where the chromosomal breakpoints are on either side of the centromere and pericentric inversions where the breakpoints occur within one of the chromosome arms. Inversions of chromosomal material are thought to aid leukaemogenesis in a similar manner to translocations. Common inversion in leukaemia include inv(16) in AML, inv (12p) in ALL (Welborn et al. 2004; Braoudaki and Tzortzatou-Stathopoulou 2012).

- **Numerical chromosomal abnormality and leukaemia development**

Numerical chromosomal abnormalities are a common occurrence in leukaemia. Aneuploidy is a type of numerical chromosomal abnormality where there is loss or gain of one or more chromosomes resulting in a loss of the normal diploid karyotype. It is the most common type of chromosomal abnormality and it occurs as a result of chromosome missegregation during mitosis resulting in an abnormal distribution of chromosomes into daughter cells (Gordon et al. 2012; Braoudaki and Tzortzatou-Stathopoulou 2012; Potapova et al. 2013). Types of aneuploidy include trisomy (three copies of a particular chromosome) and monosomy (one copy of chromosome). Examples of commonly found aneuploid karyotypes in leukaemia include monosomy 7(-7), monosomy 5(-5), trisomy 8 (+8) (Estey and Döhner 2006; Hasle et al. 2007; Manola 2009; Braoudaki and Tzortzatou-Stathopoulou 2012).

Depending on the number of chromosomes lost or gained by cells, the karyotype can be described as being near-haploid, hypodiploid or hyperdiploid. Near-haploid cells have lost almost half of the chromosomes, hypodiploid cells have fewer chromosomes than a diploid karyotype (<46) and hyperdiploid cells have more chromosomes than a diploid karyotype (>46). Hyperdiploidy is associated with a more favourable outcome than hypodiploidy or near haploid karyotype, and the prognosis of an hypodiploid or near haploid karyotype is linked in severity with the number of chromosomes lost (Braoudaki and Tzortzatou-Stathopoulou 2012).

Another type of numerical abnormality is polyploidy; as most human cells are diploid (2n), polyploidy involves gains of whole sets of chromosomes e.g. triploidy (3n) and tetraploidy (4n). Polyploidy can occur in leukaemia and although this is very rare, it is associated with very poor prognosis (Kim et al. 2012; Braoudaki and Tzortzatou-Stathopoulou 2012; Li et al. 2012). The mechanisms by which numerical chromosomal abnormalities contribute to leukaemogenesis remains unknown due to large amount of genetic material lost or gained (whole chromosomes) (Gordon et al. 2012).

1.8.8. Chromosome 7 abnormalities in leukaemia

Monosomy 7 (-7) and deletions in the long arm of chromosome 7, del(7q) are commonly observed in myeloid malignancy (Tosi et al. 1996; Johnson and Cotter 1997; van der Straaten et al. 2005; Hasle et al. 2007). Del(7q) is particularly associated with secondary myelodysplastic syndrome (MDS) or AML, which arises as a result of chemotherapy treatment with alkylating agents and topoisomerase (Döhner et al. 2010; Hasle et al. 2007; Tosi et al. 1996). Monosomy 7 and Del(7q) are also found in De novo cases of MDS and AML although with significantly reduced incidence when compared to secondary AML/MDS (Hasle et al. 2007; Tosi et al. 1999; Döhner et al. 2010).

Monosomy 7 and deletion in 7q in myeloid disorders are more commonly found in elderly patients compared to younger patients (Hasle et al. 2007; Honda et al. 2015). Yet, these chromosomal abnormalities are also found in significant proportion of paediatric and adolescent patients. In a study of 124 paediatric and adolescent cases of AML, monosomy 7 was found in 10% of cases while del(7q) was found in 4 % of cases (Manola et al. 2013).

Reported cases of AML with chromosome 7 abnormalities include cases with del(7q) as the only abnormality, cases of del(7q) associated with complex karyotype, deletions of 7q due to unbalanced translocation, deletion in 7p due to unbalanced translocation and translocation of chromosome 7 without deletions in 7q/7p (Březinová et al. 2007).

Monosomy 7 and del(7q) are not associated with a particular FAB subtype implying that all myeloid linages are equally affected (Hasle et al. 2007; Manola et al. 2013). Clinically, 7/del (7q) is used as a prognostic biomarker as patients with -7/del(7q) respond poorly to therapy and have a low survival rate (three year event free survival of 34%) (Tosi et al. 1996; van der Straaten et al. 2005; von Bergh et al. 2006; Březinová et al. 2007).

Due to the poor prognosis of -7/del(7q) in AML, research was carried out to identify the commonly deleted regions on 7q. The common breakpoints on 7q are at q22, q31-q35, q36. Therefore, it was hypothesised that these regions might contain tumour suppressor gene(s), which when deleted or disrupted lead to leukaemogenesis (De Weer et al. 2010; Tosi et al. 2003; Tosi et al. 1999; Tosi et al. 1996; Park et al. 2009).

Genes identified in the commonly deleted region of chromosome 7 that are thought to contribute to leukaemogenesis include *CUX1, EZH2, MLL3* and *LUCL7* (Hosono et al. 2014; Honda et al. 2015). Using transcriptome sequencing and SNP array analysis, McNerney and colleagues identified a commonly deleted region at 7q 22.1 in monsomy 7/del 7q patients (McNerney et al. 2013). They subsequently identified the deleted region contained the *CUX1* tumour suppressor gene. They found patients with del(7q) showed haploinsufficient expression of *CUX1*, a transcription factor crucial in the regulation of the cell cycle, cell migration and cell differentiation. They also found transplanted *CUX1* haploinsufficient hematopoietic stem cells were better able to engraft immunocompromised mouse by 40% (McNerney et al. 2013). *CUX1* as a tumour suppressor has also been implicated in other cancers (McNerney et al. 2013; Hosono et al. 2014; Honda et al. 2015).

Another gene in the commonly deleted region at 7q36 is *EZH2* (De Weer et al. 2010). Ernst and colleagues found *EZH2* was commonly mutated in MDS/MPN patients. They found these patients had loss of function mutations of *EZH2* suggesting *EZH2* functions as a tumour suppressor in myeloid cells (Ernst et al. 2010; De Weer et al. 2010; Hosono et al. 2014; Honda et al. 2015).

The method of identifying and investigating the function of genes in the commonly deleted region in -7/7q patients is promising and has started to help in the understanding of leukaemogenesis in these patients. However, considering the loss of whole chromosome observed in patients harbouring monosomy 7 and that the deleted region in 7q (7q22-7q36) is heterogeneous, several genes that are lost or deregulated might play a role in contributing to leukaemogenesis in these patients (Ernst et al. 2010; De Weer et al. 2010; Boultwood 2013; Hosono et al. 2014; Honda et al. 2015).

1.8.9. Chromosomal abnormalities and *HLXB9* expression in leukaemia

Studies carried out on AML infant patients harbouring a translocation of chromosome 7 and 12 t(7;12)(q36;p13) show an overexpression of the *HLXB9* gene in all of the cases investigated (von Bergh et al. 2006; Ballabio et al. 2009; Park et al. 2009). The breakpoint in these cases were variable but have been observed not to disrupt the *HLXB9* gene (Tosi et al. 2000; Simmons et al. 2002; Tosi et al. 2003; Naiel et al. 2013). Patients with this translocation have a very poor prognosis, a three-year event free survival of 0% (von Bergh et al. 2006).

In t(7;12) patients, the translocation involves the *HLXB9* and *ETV6* genes. *ETV6* (TEL) belongs to a group of ETS transcription factor family; it is a ubiquitously expressed and is encoded at 12p13 (Tosi et al. 2000). *ETV6* plays a very important in haematopoiesis (Kralik et al. 2011). The chromosome band 12p13 has been shown to be involved in translocation with various partners in many haematological malignancies including AML (Bohlander 2005; Kralik et al. 2011; Tosi et al. 1999). The adverse prognosis of patients diagnosed with AML harbouring t(7;12) has naturally led to research to identify the molecular mechanism of leukaemogenesis in these patients.

As one of the mechanisms by which translocations contribute to leukaemogenesis is the production of a oncogenic fusion protein, a number of studies have been employed to identify a fusion transcript in t(7;12) patients. Reverse transcriptase-polymerase chain reaction (RT-PCR) used to identify the fusion transcript in t(7;12) patients showed a fusion of exon 1 of the *HLXB9* gene to either exons 2 or 3 of the *ETV6* gene (von Bergh et al. 2006; Taketani et al. 2008; Ballabio et al. 2009; Park et al. 2009). Two different *HLXB9-ETV6* fusion transcripts have been described, an out of frame longer variant (exon 1 of *HLXB9* to exon 2 of *ETV6*) and an in-frame shorter variant (exon 1 of *HLXB9* to exon 3 of *ETV6*). However, both fusion transcripts contains the regulatory (promoter) sequence, the first exon of the *HLXB9* gene, the ETS domain (DNA and protein binding domain) and the pointed N terminal domain (PNT) (protein-protein binding domain) of the *ETV6* gene (von Bergh et al. 2006; Taketani et al. 2008; Ballabio et al. 2009; Park et al. 2009). Although a fusion transcript has been observed for this translocation, it is only present in 50% of cases suggesting that the formation of an oncogenic fusion protein may not be the mechanism for oncogenesis in these patients (von Bergh et al. 2006; Taketani et al. 2008; Ballabio et al. 2009; Park et al. 2009).

A commonality in all AML patients harbouring t(7;12) is the overexpression of the *HLXB9* gene suggesting that the overexpression of *HLXB9* might be the driver of leukaemogenesis in these patients. Overexpression of *HLXB9* in t(7;12) patients was identified by von Bergh and colleagues when they quantified the expression of genes at 7q36 in these patients by Real Time Quantitative PCR (von Bergh et al. 2006). They found the expression of other genes localised on 7q36-*NOM1, LMBR1* and *RNF32* in t(7;12) AML patients was not altered compared to normal bone marrow. They also found the expression of the ETV6 gene (*HLXB9* translocation partner) was also unchanged compared to normal bone marrow. The

aberrant expression of *HLXB9* in t(7;12) AML patients has subsequently been observed in studies by other groups (Ballabio et al. 2009; Wildenhain et al. 2012).

Although all t(7;12) AML patients overexpress *HLXB9,* only 50% of patients show the *HLXB9-ETV6* fusion transcript. This raised a question on how ectopic expression of *HLXB9* occurs in these patients. A study by Ballabio et al. 2009 suggested that a change in the nuclear localisation of the translocated 7 from the nuclear periphery to the nuclear interior might result in the overexpression of *HLXB9* observed due to a position change effect. Another hypothesis for the potential driver of aberrant *HLXB9* expression observed is that the wild type chromosome 7 is responsible although this is yet to be confirmed experimentally.

Although these studies identified the aberrant overexpression of *HLXB9* in t(7;12) AML patients, the question of whether *HLXB9* is expressed by healthy blood cells (bone marrow or peripheral blood) is debated. A study by Deguchi and Kehrl in 1991 found low expression of *HLXB9* in whole population of bone marrow cells and increased expression in bone marrow fractions enriched for CD34+ cells (Deguchi and Kehrl 1991). However, subsequent studies have been unable to detect *HLXB9* expression in healthy CD34+ bone marrow cells and peripheral blood (Nagel et al. 2005; von Bergh et al. 2006; Wildenhain et al. 2012). Notwithstanding the question of whether *HLXB9* is expressed in healthy blood cells, it can be concluded from studies comparing *HLXB9* expression in t(7;12) AML patients to healthy progenitor blood cells and other AML patients without t(7;12) that the overexpression of *HLXB9* observed in t(7;12) AML patients is abnormal (von Bergh et al. 2006; Balgobind et al. 2011; Wildenhain et al. 2012).

- **Gene expression changes in t(7;12) AML patients**

The functional significance of the overexpression of *HLXB9* observed in t(7;12) AML patients is an important research question because of the extremely poor prognosis of these patients.

Wildenhain and colleagues compared the gene expression profile of t(7;12) AML patients to mixed linage leukaemia (*MLL*) rearranged AML patients, another group of patients with a poor prognosis (Wildenhain et al. 2010). They found compared to *MLL* rearranged AML patients, t(7;12) patients shown downregulation of *HOX* genes and up regulation of cell to cell interacting genes *EDIL3, CNTNAP5, ANGPT1, DSG2, ITGA9, ITGAV, KDR, SIGLEC6*. This finding suggests the mechanism of leukaemogenesis in *MLL* AML patients is different from t(7;12) AML patients. Out of the genes upregulated in t(7;12) patients, two

genes: *angiopoietin-1 (ANGPT1)* and *KDR* are involved in maintaining normal quiescent haematopoietic stem cells (HSC) in healthy bone marrow niche. Their findings suggest dysregulation of these crucial of these genes could contribute to leukaemogenesis.

Another gene expression profiling study by Balgobind and colleagues also identified a unique gene expression signature in t(7;12) AML patients (Balgobind et al. 2011). They identified 16 genes that are differential expressed in t(7;12) AML patients compared to healthy bone marrow cells and other non t(7;12) AML patients. This finding implies that the presence of the translocation t(7;12) or *HLXB9* overexpression results in a unique gene expression profile that may contributes to leukaemogenesis in these patients (Balgobind et al. 2011).

Wildenhain and colleagues also used ChIP on ChIP and gene expression profiling to identify *HLXB9* target genes in t(7;12) AML patients by transfecting an AML derived cell line (HL-60) with a pMC vector containing HLXB9 (Wildenhain et al. 2012). They found that HLXB9 was able to bind to *prostaglandin E receptor 2 (PGTER2)* to downregulate the expression of *PGTER2*, thereby reducing intracellular cAMP level. They suggested the downregulation of *PGTER2* might contribute to the disruption of bone marrow homeostasis, as *PGE2* the ligand for *PGTER2* has a crucial function in maintaining bone marrow homeostasis. They also found HLXB9 expression in HL60 cell line had a repressive effect on other genes, particularly downregulating the expression of genes involved in cell adhesion and cell to cell interaction such as *ZYX, ETS1* (Wildenhain et al. 2012).

In summary, these studies indicate the dysregulation of genes involved in cell to cell interaction occurs in t(7;12) patients. Therefore, *HLXB9* overexpression in these patients may disrupt the homeostasis of the bone marrow niche contributing to leukaemogenesis.

1.9. *HLXB9* in breast cancer

Breast cancer is the most commonly diagnosed cancer in the UK (Cancer Research UK). It is more commonly diagnosed in females than in males. In 2013, approximately 53,700 women were diagnosed with breast cancer compared to 340 men in the UK. Breast cancer is also the second leading cause of cancer deaths in women (Cancer Research UK). Breast cancer is a heterogeneous disease encompassing a variety of histological and biological subtypes.

Histologically, there are more than 20 distinct types of breast cancer (Weigelt et al. 2010). The most common type of breast tumours originates in the milk ducts. It is classified into ductal carcinoma *in situ,* DCIS (benign tumour) or invasive ductal carcinoma (malignant tumour). Approximately 50-80% of all breast tumours diagnosed are of the invasive ductal carcinomas (no specific subtypes) (Weigelt et al. 2010; Caldarella et al. 2013).

The second most commonly diagnosed type of breast tumours diagnosed originates from the breast lobules. They are classified into lobular carcinomas *in situ* (LCIS) and invasive lobular carcinomas. Approximately 5%-15% of breast tumours are invasive lobular carcinoma (Weigelt et al. 2010; Caldarella et al. 2013). There are other rare forms of breast carcinomas that originate from breast tissues such as the fatty tissue, connective tissue and nipples (Weigelt et al. 2010).

Risk factors

Factors that affect the risk of developing breast cancer in women are divided into factors that increase risk and those that reduce risk. Factors that increase the risk of developing breast cancer include age, lifestyle factors, exogenous hormones, reproductive factors and genetic factors (García-Closas et al. 2006; Murray and Davies 2013). The risk of a woman developing breast cancer increases with age; approximately half of all breast cancer cases are diagnosed in women aged 65 and over (Cancer Research, UK). The use of exogenous hormones such as oral contraceptives and hormone replacement therapy (HRT) has also been showed to increase the risk of breast cancer. Lifestyle factors such as obesity, alcohol and lack of exercise have also been linked to breast cancer development. Other factors that have been shown to increase breast cancer risk include reproductive factors such as early age at menarche, late age at first birth, late menopause, null parity, high breast density and mutations in tumour suppressor genes.

Germline mutations in tumour suppressor genes that predispose inheritants to developing breast cancer can be classified into high penetrant, moderate penetrant and low penetrant mutations (Buchholz and Wazer 2002; Mavaddat et al. 2010). High penetrant mutations in tumour suppressor genes such as breast cancer gene 1 and 2 (*BRCA1/2*), *TP53* (Li Fraumeni syndrome), *ATM* (ataxia telangiectasia) and PTEN (cowden's factor) are rare in the population but are responsible for a majority of familial breast cancer development (Buchholz and Wazer 2002; Liebens et al. 2007; Murray and Davies 2013). For instance, germline mutations of *BRCA1/2* account for about 20-25% of familial case of breast cancer but only 5% of

all cases of breast cancers (Buchholz and Wazer 2002; García-Closas et al. 2006; Liebens et al. 2007; Mavaddat et al. 2010; Davies 2012; Murray and Davies 2013). Women who carry germline mutations of high penetrant tumour suppressor genes have a significantly increased risk of breast cancer development. For example *BRCA1/2* mutations confers a 60-85% risk of breast cancer development in inheritants (Liebens et al. 2007; Mavaddat et al. 2010; Davies 2012).

Moderate penetrant tumour suppressor mutations account for less than 3% of breast cancer cases (Mavaddat et al. 2010). They include mutation in tumour suppressor genes such as *CHEK2, PALB, BRIP1*; genes important in DNA repair (Mavaddat et al. 2010; Apostolou and Fostira 2013). Low penetrant tumour suppressor mutations in genes such as *FGFR2*, *MAP3K1* have very weak associations with breast cancer development (Mavaddat et al. 2010; Apostolou and Fostira 2013). The mechanism by which low penetrant mutations confer familial risk is thought to be through interactions with other genetic, environmental or lifestyle factors.

Factors that reduce risk of developing breast cancer includes young age at first birth, high parity and an increased period of breast-feeding (García-Closas et al. 2006; Davies 2012).

Prognostic indicators in breast cancer

Prognostic indicators in breast cancer include tumour size, lymph node involvement, distant metastasis, tumour grade and tumour intrinsic molecular markers (Arnone et al. 2010; Davies 2012). These characteristics are used to judge a patient's treatment options and prognosis in clinic (Arnone et al. 2010; Davies 2012; Tang et al. 2016).

- **TNM prognostic staging of breast cancer**

The traditional way of assessing patients' prognosis is an assessment of tumour size, node involvement and metastasis (TNM system) described in table 1.5. Tumour size (T) is an indicator of cancer progression. Smaller tumours are usually associated with a better prognosis than larger tumours as it usually means the tumour was detected early and is unlikely to have metastasised (Narod 2012). Lymph node involvement (N) is also an indicator of cancer progression as the presence of local metastasis in the lymph node indicates the cancer has spread from the breast (metastasis) (Galea et al. 1992). Metastasis (M) is also an indicator of cancer progression; the presence of distant metastasis significantly reduces a

patient's survival and is the leading cause of breast cancer related death (Arnone et al. 2010; Davies 2012). Common sites of metastasis in breast cancer include lungs, bone, liver and brain (Davies 2012).

Tumour Size		
Tx	Primary tumour cannot be assessed	
T0	No evidence of primary tumour	
Tis	Carcinoma in situ	
T1	T1mi: Tumour ≤ 1 mm T1a Tumour: > 1 mm but ≤ 5 mm T1b Tumour: > 5 mm but ≤ 10 mm T1c Tumour: > 10 mm but ≤ 20 mm	
T2	> 20 mm but ≤ 50 mm	
T3	> 50 mm	
T4	Tumour of any size with direct extension to the chest wall and/or to the skin or inflammatory carcinoma	
Lymph node Involvement	**Clinical**	**Pathological**
N0	No regional lymph node metastases	Negative of isolated tumour clusters (ITC)*.
N1	Metastasis to mobile regional nodes	1–3 micro or macro metastasis
N2	Metastasis to Fixed regional or internal mammary nodes	4–9 nodes
N3	Metastasis to supraclavicular lymph node	>10 nodes, axillary and internal mammary or supraclavicular nodes
Distant Metastasis		
M0	No clinical or radiographic evidence of distant metastases	
cM0(i+)	No clinical or radiographic evidence of distant metastases, but deposits of molecularly or microscopically detected tumour cells in circulating blood, bone marrow, or other nonregional nodal tissue that are no larger than 0.2 mm in a patient without symptoms or signs of metastases	
M1	Distant detectable metastases as determined by classic clinical and radiographic means and/or histologically proven larger than 0.2 mm	

Table 1.5: TNM classification of breast tumours. 7th Edition of American Joint Committee of Cancer (AJCC) breast cancer staging. *Isolated tumour cell clusters (ITC) are defined as small clusters of cells not greater than 0.2 mm, single tumour cells, or a cluster of fewer than 200 cells in a single histologic cross-section (Davies 2012; Ozsaran and Alanyah 2013).

The TNM classification of breast tumours is used to assign the stage of breast cancer progression in patients (as shown in table 1.6). There are four main stages of breast cancer. Breast cancer stages are associated with patient's prognosis; the earlier in the staging classification the tumour falls into, the better the prognosis for patients. Stage IV breast cancer has the worst prognosis as metastasis has occurred.

Stage	TNM	Five year Survival Rate*
Stage 0	Tis, M0, NO	N/A
Stage I	Stage IA: T1 N0 M0 Stage IB: T0-T1 N1M0	99%
Stage II	Stage IIA: T0 N1 M0 T1 N1M0 T2 N0 M0 Stage IIB: T2 N1 M0 T3 N0 M0	90%
Stage III	Stage IIIA: T0 N2 M0 T1 N2 M0 T2 N2 M0 T3 N1 M0 T3 N2 M0 Stage IIIB: T4 N0 M0 T4 N1 M0 T4 N2 M0 Stage IIIC Any T N3 M0	60%
Stage IV	Any T, Any N, M1	15%

Table 1.6: American Joint Committee of Cancer Breast Cancer Staging (7th Edition) (Davies 2012; Ozsaran and Alanyah 2013). The five-year survival rates were obtained from the Cancer Research UK and are based on survival rates of breast cancer patients diagnosed between 2002 and 2006.

- **Tumour grade**

Histological tumour grade is an indicator of tumour aggressiveness. Tumour grade is used in conjunction with the TNM stage to determine the patient's prognosis (Rakha et al. 2010; Weigelt et al. 2010). It is incorporated into various prognostic indices, algorithms and treatment guidelines such as the Nottingham Prognostic Index, Adjuvant! Online and St. Gallen guidelines (Rakha et al. 2010; Weigelt et al. 2010). Tumour grade is a measure of the tumour cell differentiation and proliferation (Rakha et al. 2010; Weigelt et al. 2010). Tumour cell differentiation is a judgement of the tumour's cell ability to form functional glands and nuclear appearance. Progression in cancer development is characterised by a block in the ability to form functional glands and abnormal nuclear appearance termed 'nuclear pleomorphism', which is change to nuclear shape and size in tumour cells. Tumour proliferation or mitotic index is a measure of cellular growth pattern.

In the UK, tumour grade is measured using the Nottingham Grading System (NGS) (Rakha et al. 2010; Weigelt et al. 2010). The NGS classifies breast tumours into three grades (Grade 1 to 3). Grade 1 tumours are described as well differentiated tumour cells that are slowly growing. Grade 2 tumours have moderately differentiated cells that are starting to proliferate uncontrollably. Grade 3 tumours have poorly differentiated cells with a marked increase in proliferation. Similar to breast cancer stages, higher tumour grades are associated with significant reduction in patient's survival (Rakha et al. 2010; Weigelt

et al. 2010). Higher-grade tumours are also associated with increased metastatic potential and is usually an indication that metastasis had already occurred or is likely to occur (Harvey et al. 2009).

- **Genetic biomarkers in breast cancer**

The heterogeneity of breast tumours is also observed in its gene expression profile. Microarray gene expression profiling of breast tumours resulted in the classification of breast cancer into four main 'intrinsic' molecular subtypes (Perou et al. 2000; Sorlie et al. 2003; Parker et al. 2009; Curigliano 2012; Tang et al. 2016). These 'intrinsic' molecular subtypes include luminal A, luminal B, HER2-overexpressing and basal like tumours. These molecular subtypes are also used to inform patient's treatment decisions and can be used to predict patients' outcome as they have prognostic significance.

This classification is based on differential expression of genes associated with breast cancer proliferation such as oestrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 (HER 2) and cell proliferation measured by Ki 67 expression. Increased expression of oestrogen receptor (ER) and progesterone receptor (PR) were identified in breast tumours early on in breast cancer research as reviewed in Sommer and Fuqua 2001 and Murphy and Watson 2002. Activation of these receptors results in cellular signalling processes that induce uncontrolled cell proliferation (Murphy and Watson 2002; Sommer and Fuqua 2001).

ER positive tumours are termed luminal tumours because their gene expression profiles are similar to non-cancer breast luminal epithelium cells. ER positive tumours are classified into luminal A and luminal B tumours (Perou et al. 2000; Sorlie et al. 2003; Parker et al. 2009; Weigelt et al. 2010).

Luminal A subtype breast tumours are the most commonly diagnosed type of breast tumours, representing 40-60% of all breast cancer cases (Eroles et al. 2012; Tang et al. 2016). These tumours are characterised by high expression of ER related genes; low expression of HER2 related genes; low levels of proliferation as detected by Ki-67, and are usually low grade tumours (Perou et al. 2000; Sorlie et al. 2003; Parker et al. 2009; Weigelt et al. 2010; Eroles et al. 2012). These tumours are associated with a favourable prognosis as they can be treated successfully with hormonal therapies that target ER signalling such as tamoxifen or aromatase inhibitors (Curigliano 2012; Davies 2012; Tang et al. 2016).

Luminal B subtypes account for 10-20% of all breast cancers (Eroles et al. 2012; Tang et al. 2016). These tumours are characterised by a lower expression of ER related genes than luminal A tumours, usually show overexpression of HER2; have high levels of proliferation as detected by Ki-67; lymph node involvement; and are usually high grade tumours (Perou et al. 2000; Sorlie et al. 2003; Parker et al. 2009; Weigelt et al. 2010; Eroles et al. 2012; Inic et al. 2014). As this tumour molecular subtype is associated with poorer prognostic characteristics markers, it is usually associated with poor prognosis. These tumours can also be treated with ER targeted therapy, however, they are less responsive to ER targeted therapy than luminal A tumours and may respond better to treatment with chemotherapy agents (Curigliano 2012; Eroles et al. 2012; Tang et al. 2016).

ER negative molecular subtype tumours include HER2 overexpressing tumours and basal and normal like tumours. HER2 overexpressing breast tumour accounts for 10-20% of diagnosis (Eroles et al. 2012; Tang et al. 2016). These tumours are characterised by their overexpression of HER2. They usually lack the expression of ER and PR, are poorly differentiated, highly proliferative, show lymph node involvement and a high risk of metastasis (Perou et al. 2000; Sorlie et al. 2003; Parker et al. 2009; Weigelt et al. 2010; Eroles et al. 2012; Tang et al. 2016). Therefore, this tumour molecular subtype is associated with poor prognosis. HER2 belongs to the human epidermal growth factor receptor (HER) family of receptors (Davoli et al. 2010). Upon activation, these receptors are involved in signalling pathways that control cell growth, differentiation and survival (Davoli et al. 2010). Targeted antibody therapy to the HER2 receptor such as trastuzumab (Herceptin) and pertuzumab (Perjeta), has considerably improved the prognosis of patients diagnosed with this molecular subtype (Parker et al. 2009; Davoli et al. 2010; Eroles et al. 2012; Tang et al. 2016).

Basal tumours, also called triple negative tumours are so termed because they do not express either ER, PR or HER2 receptors (Perou et al. 2000; Sorlie et al. 2003; Parker et al. 2009; Weigelt et al. 2010; Bertucci et al. 2012; Eroles et al. 2012). They are characterised by high proliferation and are associated with a poor prognosis because of the lack of molecular targets for therapy (Perou et al. 2000; Sorlie et al. 2003; Parker et al. 2009; Weigelt et al. 2010; Bertucci et al. 2012; Eroles et al. 2012). Basal tumours account for 10-20% of breast cancer cases (Bertucci et al. 2012; Eroles et al. 2012; Tang et al. 2016). Their gene expression profiles are similar to basal epithelial cells (Bertucci et al. 2012; Eroles et al. 2012). This tumour molecular subtype is characterised by large tumour size, high levels of proliferation, poor

differentiation, high tumour grade, lymph node involvement, high metastatic potential, high relapse rate, TP53 mutations, and are thereby associated with a very poor prognosis (Bertucci et al. 2012; Eroles et al. 2012). Basal tumours also make up most breast cancer associated with *BRCA1* mutations (Bertucci et al. 2012; Eroles et al. 2012; Murray and Davies 2013). Currently, there are no options for targeted gene therapy in the treatment of triple negative tumours. Therefore, patients are treated with standard options such as surgery, radiotherapy and chemotherapy. However, patients with triple negative tumours are known to respond poorly or acquire resistance to chemotherapy. Consequently, there are on-going attempts to identify gene targets for therapy for this intrinsic subtype (Curigliano 2012; Bertucci et al. 2012; Tang et al. 2016).

Although the use of these prognostic factors to inform treatment decisions has considerably improved patients' outcome, several gaps remain in the understanding of the class of diseases termed 'breast cancer'. Some of these gaps include: the heterogeneity of tumours (both intra and inter tumour), identification of markers for early diagnosis, stratification of in situ breast tumours to prevent overtreatment of patients, understanding disease progression (relapse and metastasis), side effects of current treatments and therapy resistance. Therefore, the identification and validation of new genetic biomarkers, which can be used to answer some of these questions, will be very useful in improving patients' outcome.

- **Breast cancer tumour sequencing based classifications**

There have been a number of studies to identify other genes that can be targeted for therapy in breast cancer. The studies have arisen due to the development of new techniques such as DNA and RNA sequencing methods (Desmedt et al. 2013). A landmark study on 2,000 breast tumours by Curtis et al in 2012 (METABRIC) identified ten 'types' of breast cancer as well as a number of frequently deleted genes such as *PPP2R2A, MTAP,* and *MAP2K4.* A follow up study by the same group on 2,433 breast tumours samples identified 173 frequently mutated genes from which they were able to identify 40 potential cancer driver genes including *PIK3CA*, *AKT1, KMT2C* (Pereira et al. 2016). These studies are starting to provide potential targets for therapy. They have also started to advance our understanding of breast cancer, for instance, they have shown that although there are numerous gene mutations in breast cancer, these mutations are not frequently occurring but usually involve mutation in genes involved in similar signalling pathways or biological processes (Desmedt et al. 2013; Pereira et al. 2016). A frequent process

deregulated in breast cancer is apoptosis; therefore aberrant *HLXB9* expression, which is known to be anti-apoptotic in insulinomas, was examined in breast cancer.

HLXB9 expression in Breast Cancer

Aberrant expression of *HLXB9* has been found in breast cancer (Neufing et al. 2003; Tommasi et al. 2009; Lian et al. 2012; Siletz et al. 2013b; Siletz et al. 2013a). The earliest study on *HLXB9* expression in breast cancer was carried out by in 2003 (Neufing et al. 2003). In their study, they compared *HLXB9* expression in matched normal and cancer samples in breast cancer patients. They found *HLXB9* was expressed in both normal and cancer breast sections. Additionally, they found, although a greater proportion of cancer epithelial cells in breast cancer patients showed *HLXB9* expression, there was a significant reduction in the intensity of *HLXB9* staining in these cancer sections compared to normal sections. They also found a decrease in the staining intensity of *HLXB9* associated with increased tumour grade in patients. These findings imply a loss of *HLXB9* expression in breast cancer associated with cancer progression suggesting that *HLXB9* has a tumour suppressive role in breast cells.

Other studies have subsequently supported these findings by identifying a downregulation of *HLXB9* in breast cancer (Tommasi et al. 2009; Lian et al. 2012; Siletz et al. 2013b; Siletz et al. 2013a). These include two studies that found hypermethylation of *HLXB9* in breast cancer samples compared to control samples (Tommasi et al. 2009; Lian et al. 2012). In the first study by Tomassi and colleagues, they found early occurring breast cancer specific hypermethylation of *HLXB9* in breast cancer patients diagnosed with ductal carcinoma in situ tumours. The second study by Lian and colleagues also found hypermethylation of *HLXB9* in early stage invasive breast cancer samples compared to normal breast samples (Lian et al. 2012). In conjunction, results from both studies showed hypermethylation of *HLXB9* was an early occurrence in breast cancer development and downregulation of *HLXB9* may have potential as an early biomarker of breast cancer development.

Two other studies on transcription factor activity in breast cancer by Siletz and colleagues have also reported downregulation of *HLXB9* was associated with cancer progression (Siletz et al. 2013b; Siletz et al. 2013a). In their first study investigating the epithelial-mesenchymal transition (EMT) process in cancer, they identified down regulation of *HLXB9* in all models of EMT that they investigated (Siletz et al. 2013b). EMT is a process that facilitates cancer progression as epithelial cells take on mesenchymal

characteristics such as loss of polarity, loss of cell-to-cell adhesion that enables cells migration and invasion, and are critical processes for cancer progression. They also found downregulation of *HLXB9* in breast carcinoma associated fibroblasts (CAF) in their other study (Siletz et al. 2013a). CAFs are tumour activated stromal cells and are associated with cancer progression as they promote proliferation, invasion, angiogenesis and other cellular process that aid cancer progression.

Although downregulation of HLXB9 has been found in breast cancer, the function of HLXB9 in healthy breast is unknown. Consequently, its function in breast cancer is also unknown. However, as HLXB9 role in foetal development is in the differentiation of cells, it can be assumed that that HLXB9 may be implicated in the differentiation of breast cells. This hypothesis has also been proposed by other studies (Siletz et al. 2013b; Siletz et al. 2013a). Therefore, downregulation of *HLXB9* in breast cancer is not surprising as a block in differentiation is a hallmark of cancer development (Xu et al. 2010).

Finally, another study that implicated *HLXB9* deregulation to breast cancer development is the finding that patients with germline mutations of *MENIN (MEN1)* show susceptibility to developing breast cancer (Shi et al. 2013; Dreijerink et al. 2014). These patients are also susceptible to developing other tumours associated with deregulation in *HLXB9* expression such as pancreatic cancers. In the study by Shi and colleagues, *HLXB9* was shown to interact with *MEN1* to regulate cellular proliferation by promoting apoptosis in mouse pancreatic b-cells. *MEN1* is a known co-activator of oestrogen receptor in breast cancer (Imachi et al. 2010). Menin has also been observed to directly activate the oestrogen receptor (Dreijerink et al. 2006; Dreijerink et al. 2014). Therefore, these studies indicate the possibility of *HLXB9* interacting with *MEN1* in ER related breast cancer to aid development and/or progression.

1.10. *HLXB9* in other cancers

Aberrant *HLXB9* expression has also been observed in other cancers such as testicular cancer, hepatocellular carcinoma, colorectal and lymphomas. In testicular cancer, overexpression of *HLXB9* has been found in patients diagnosed with pre-invasive testicular carcinoma *in situ* and malignant seminoma (a malignant type of testicular cancer) (Almstrup et al. 2005; Almstrup et al. 2007; Novotny et al. 2007). These studies suggest *HLXB9* has a pro oncogenic role in testicular cancer.

Hepatocellular carcinoma (HCC), the most commonly diagnosed type of liver cancer, is also a cancer type associated with overexpression of *HLXB9* (Wilkens et al. 2011). In the study by Wilkens and colleagues,

HLXB9 overexpression was found in HCC cell lines by microarray analysis. Additionally, they found *HLXB9* was upregulated in patients diagnosed with high-grade HCC tumours. *HLXB9* overexpression in these patients was associated with abnormal development of glands indicating a dedifferentiation of the tumour cells. In HCC patients overexpressing *HLXB9*, they also found the upregulation of genes involved in cell-cell contacts, which are genes previous reported to be upregulated in leukaemia patients with *HLXB9* overexpression.

Aberrant overexpression of *HLXB9* has been found in lymphoma cell lines (Nagel et al. 2005; Nagel et al. 2014; Nagel et al. 2015)**.** In lymphoma, *HLXB9* overexpression has been implicated with increased immune response (a characteristic of lymphoma) by binding to *interleukin 6* (*IL6*) promoter to increase its expression (Nagel et al. 2005). Further studies into the role of homeobox genes in lymphoma identified the aberrant overexpression of *OTX2* and *MSX2*, which are homeobox genes involved in early b cell differentiation (Nagel et al. 2014; Nagel et al. 2015). Results from both studies also showed that overexpression of the homeobox genes (*OTX2 and MSX2*) directly contributes to the overexpression of *HLXB9* observed in lymphoma

Finally, *HLXB9* has been found to be downregulated in colorectal cancer (Hollington et al. 2004). In the study by Hollington and colleagues, *HLXB9* was found to be downregulated in moderately differentiated colorectal patient samples (Hollington et al. 2004).

These studies indicate the aberrant *HLXB9* expression may have potential as a biomarker in cancer as its abnormal expression was observed in a number of cancer types.

1.11. The localisation of *HLXB9* in the interphase nucleus

After mitosis, decondensed chromosomes are organised in the nucleus into distinct regions termed 'chromosome territories' (Meaburn and Misteli 2007). Gene rich chromosomes were observed to be localised in the nuclear interior whilst gene poor chromosomes were localised in the nuclear periphery (Boyle et al. 2001; Cremer and Cremer 2001; Cremer and Cremer 2010). Consequently, this nuclear organisation has been linked to gene expression as the nuclear periphery is characterised by the presence of closed chromatin 'heterochromatin' whilst the nuclear interior is characterised by open chromatin 'euchromatin', which is more readily accessible for transcription. This pattern of nuclear organisation has

also been observed for individual chromosomes. For instance, in the nuclear organisation of chromosome 7 in the nucleus of lymphocytes, gene poor chromosomal regions were localised in the nuclear periphery whilst gene rich chromosomal regions were localised in the nuclear interior (Federico et al. 2008). In their study, Federico and colleagues found that 7q36, a poor gene region containing the *HLXB9* gene, was localised in the nuclear periphery.

Apart from chromatin structure, regions or organelles in the nucleus have also been associated with transcription activation or repression. For instance, in yeast nuclear periphery, regions associated with gene silencing have been observed, in addition to regions associated with transcription such as the nuclear pore complex (Taddei et al. 2006). In yeast, tethering of genes to the nuclear pore complex has been observed to directly result in their expression. In mammalian cells, although there have been cases of genes associating with the nuclear pore complex and subsequent gene expression, there remains a lack of consensus on whether the association is the direct cause of the observed expression (Dieppois and Stutz 2010).

In support of the presence of transcriptionally active regions in the mammalian nuclear periphery are genes, which have been observed to transcribe in the nuclear periphery such as *β-globin* during murine b-cell differentiation (Ragoczy et al. 2006). However, the link between localisation in the nuclear periphery and gene repression in the mammalian nucleus is better understood. For instance, genes localised within heterochromatin blocks on the periphery have been found associated with the nuclear lamina in 'lamina associated domains' (LADs) which are characterised by the presence of gene repressive markers such as histone3 lysine 9 dimethylation (Pombo and Dillion 2015). Although the nuclear interior is associated with active transcription, factors such as the distribution of transcription factories and splicing factors, polycomb repressive complex (PCG) and epigenetic regulation can have a repressive effect on expression of gene localised in the region (Geyer et al. 2011).

A change in nuclear localisation of gene associated with subsequent expression has been observed in a number of studies. For example, during ex-vivo adipocyte differentiation, genes involved in differentiation were observed to move from the periphery to the interior to facilitate gene expression (Szczerbal et al. 2009). Additionally, during murine b-cell differentiation, the *β-globin* locus is observed to move progressively away from the nuclear periphery as b-cell maturation progresses (Ragoczy et al.

2006). Other examples of genes where a change in their nuclear localisation away from the nuclear periphery or heterochromatin to the nuclear interior results in increased gene expression include differentiation genes such as *GATA, IgH* and *IgB* (Skok et al. 2001; Zink et al. 2004; Szczerbal et al. 2009). This mechanism of gene regulation has also been observed in genes localised at 7q31, a region of chromosome 7 with the same gene density as 7q36 where *HLXB9* is located (Zink et al. 2004; Federico et al. 2008).

Further studies linking gene localisation to gene expression include a study into the monoallelic regulation of *GFAP* expression during neuronal differentiation (Takizawa et al. 2008). They found the active 'expressing' *GFAP* allele had a more interior nuclear position compared to the inactive 'non-expressing' *GFAP* allele. Additionally, a change in the nuclear localisation from the nuclear interior to the nuclear periphery associated with gene repression has also been observed for a number of genes. For example, the *SOX9* gene is downregulated during adipocyte differentiation, which is associated with a move of gene to the nuclear periphery (Szczerbal et al. 2009). Furthermore, deliberate repositioning of chromosomal regions to the nuclear periphery or to heterochromatin blocks has been linked to their subsequent downregulation (Finlan et al. 2008).

Similarly, gene activation due to change in the nuclear localisation such as looping away chromosome territories have also been observed for a number of genes. For example, *HOXB* genes looping out of their chromosome territories is associated with their gene expression during b-cell differentiation (Chambeyron and Bickmore 2004). These studies suggest a position effect mechanism as a means of regulating gene expression. A number of the studies described above have utilised P1 or bacterial artifical chromosome (PAC/BAC) as probes to localise the genes of interest (Szczerbal et al. 2009, Takizawa et al. 2008; Ragoczy et al. 2006). For instance, In the Takizawa study, BAC probes of 188kb and 183kb were used to localise *β-actin* and *IGF-2* in murine neuronal cells.

Cancer specific nuclear reorganisation of genes, which can be used to accurately differentiate cancer cells from normal cells, has also been observed in breast cancer and prostate cancer (Meaburn and Misteli 2008; Meaburn et al. 2009; Leshner et al. 2015; Meaburn et al. 2016). Additionally, the formation of chromosomal translocations (an hallmark of cancer genome instability) by incorrect repairs of double

strand breaks between chromosomes have been found to be related to the proximal localisation of genes in the 3D nucleus (Agarwal et al. 2006). 3D chromosomal capture experiments have shown chromosomal translocations occur more frequently between genes occupying close spatial proximity in the interphase nucleus (Meaburn et al. 2007; Engreitz et al. 2012; Corces and Corces 2016).

Apart from the nuclear reorganisation of genes itself being a biomarker of cancer, changes in the nuclear localisation of genes resulting in aberrant gene expression have also been observed in cancer (Wiech et al. 2009; Ballabio et al. 2009). For instance, a change in the nuclear localisation of the *BCL2* gene to the nuclear periphery in cervical cancer has been associated with its aberrant increased expression. Similarly, a change in the nuclear localisation of *HLXB9* associated with subsequent overexpression of *HLXB9* has been previously described in leukaemia patients (Ballabio et al.2009). In the study by our group, there was a change in the nuclear localisation of *HLXB9* from the nuclear periphery to the nuclear interior in leukaemia patients harbouring the translocation t(7;12)(q36;p13). In these patients, the translocated *HLXB9* gene occupies a more interior nuclear position to the normal non-translocated *HLXB9* gene. Another study has also described a relationship between *HLXB9* expression and its nuclear localisation in neuronal differentiation (Leotta et al. 2014). In their study, Leotta and colleagues showed a change in the nuclear localisation of *HLXB9* from the nuclear periphery to the nuclear interior was associated with its increased expression during neuronal differentiation.

These studies suggest that an altered nuclear localisation of *HLXB9* can result in its aberrant expression in cancer.

1.12. Aim and objectives

The primary aim of this project was to investigate *HLXB9* as a potential biomarker in cancer. The project covers the following areas and addresses the following questions:

1. The detection of the t(7;12)(q36;p13) translocation in leukaemia patients.

   Question: Can a dual colour FISH assay be used to accurately detect this translocation in leukaemia patients?

2. The investigation of *HLXB9* expression and its nuclear localisation in patients diagnosed with various haematological disorders.

   Questions:

   - Is *HLXB9* expression limited to patients harbouring chromosome 7 abnormalities?
   - Is the disruption of 7q due to a deletion event sufficient to activate aberrant *HLXB9* expression by a position effect mechanism?
   - Do interstitial deletions of 7q result in an altered nuclear localisation of *HLXB9*?
   - Is there a link between the deletion in 7q, *HLXB9* nuclear localisation and its expression?

3. The investigation of *HLXB9* expression in various cancers by datamining online transcriptomics database.

   Questions:

   - What is the incidence of differential expression of *HLXB9* in cancer?
   - What cancer types and subtypes are associated with differential expression of *HLXB9*?
   - Is there a link between *HLXB9* expression and aggressive cancer types?

4. The investigation of *HLXB9* expression and nuclear localisation as a biomarker in breast cancer.

Questions:

- Does differential expression of *HLXB9* have prognostic significance in breast cancer?
- Is *HLXB9* upregulated or downregulated in breast cancer?
- Is the nuclear localisation of *HLXB9* altered in breast cancer?
- Is there a link between *HLXB9* expression and its nuclear localisation in breast cancer?

The following objectives were pursued in order to answer to these questions:

1. The testing of a new dual colour probe produced by MetaSystems GmbH to detect t(7;12)(q36;p13) in leukemia patients.
2. The analysis of the expression of *HLXB9* in leukaemia patients with or without deletion in 7q by RT-PCR. The nuclear localisation of *HLXB9* will be examined by fluorescence in situ hybridisation

(FISH) and erosion analysis. The results will then be evaluated to determine if there is a link between *HLXB9* gene expression and nuclear localisation.

3. The investigation of online transcriptomics databases to identify cancer and cancer subtypes exhibiting differential expression of *HLXB9*.
4. The analysis of breast cancer survival databases to determine the prognostic significance of *HLXB9* expression. Moreover, the expression of *HLXB9* in breast cancer cell line will be analysed by RT-PCR, qRT-PCR and immunofluorescence. The nuclear localisation of *HLXB9* will be analysed by FISH and erosion analysis. The results will then be evaluated to determine if there is a link between gene expression and nuclear localisation in breast cancer.

Chapter 2: Materials and Methods

2.1. Description of leukaemia patient's samples

Archival bone marrow cell suspensions in 3:1 methanol: glacial acetic acid from patients diagnosed with haematological disorders was used in this study. Fixed chromosome and nuclei suspensions used in this project were leftover from previous studies (Tosi et al., 1996; Tosi et al., 1999; Tosi et al., 2000; Tosi et al., 2003; Ballabio 2009; Naiel et al., 2013). The samples used were obtained from the following cytogenetic laboratories in the UK and Europe: Oncocytogenetic laboratory, Children's Hospital, University of Giessen, Germany; Children's Cancer Research Institute, St. Anna Children's Hospital, Vienna, Austria; Department of Medical Genetics, Churchill Hospital, Oxford, UK; Wessex Regional Genetics Laboratory, Salisbury, UK; Clinica Pediatrica dell'Universitá di Milano, Ospedale S. Gerardo, Monza, Italy; King's College School of Medicine and Dentistry, London, UK; Paterson Institute of Cancer Research, Christie Hospital, Manchester; Institute of Pathology, Odense University Hospital, Odense, Denmark; Institute of Human Genetics, University of Marburg, Germany and St Bartholomew Hospital Medical College, London, UK.

2.2. Cell culture

Leukaemia study- Chapter 4

Lymphoblastoid cell lines; GM00893 and GM17208b donated by Dr Predrag Slijepcevic were grown in RPMI-1640 containing 10% foetal calf serum (v/v), 2mM L-glutamine and 100 Uml^{-1}/μgml^{-1} Penicillin/Streptomycin at 37^0C at 5% CO_2. Cells were subcultured twice a week.

Breast cancer study –Chapter 6

Three non-cancer cell lines HMEC 184D, HMEC 240L and MCF 10A were analysed in this study. HMEC 184D and HMEC 240L were donated by Dr Yassei, MCF10A was obtained from Dr Harvey and cultured in 1:1 HamsF12: DMEM supplemented with 10% foetal bovine serum, 1% PSQ and 1 µg/ml hydrocortisone, 5 ng/ml EGF and 5 µg/ml insulin.

Six breast cancer cell lines were analysed in this study. MCF 10A-T1K and MCF 10A-CA1a cell lines were donated by Dr Paola Vagnarelli. T47D, BT474 and MDA-MB-231 breast cancer cell lines were obtained from Dr Amanda Harvey and maintained in a RPMI-1640 medium supplemented with 10% foetal calf serum and 1% PSQ (Penicillin-Streptomycin-Glutamine) in T25 cell culture flasks. These cells were grown at 37°C with 5% CO_2 in a humidified incubator.

Cells/Cell Line	Description and Reference
HMEC 184D	Non-cancer cells derived from a reduction mammoplasty from patient aged 21 (Garbe et al. 2009, Garbe et al. 2014).
HMEC 240L	Non-cancer cells derived from a reduction mammoplasty from patient aged 19 (Garbe et al. 2009,Garbe et al. 2014).
MCF 10A	Non-tumorigenic epithelial breast cell line derived from patient with fibrocystic cysts (Dawson et al. 2009).
MCF 10A – T1K	A pre-malignant cell line obtained by transforming the MCF 10A cell line by HRAS (Kadota et al. 2010, Wendt and Schiemann 2009).
MCF 10A-CA1a	A well differentiated metastatic tumour cell line derived from MCF-10A-T1K (Kadota et al. 2010, Wendt and Schiemann 2009).
T47D	Cell line was derived from a patient with invasive ductal breast carcinoma and has been classified to be a Luminal A molecular breast cancer cell line as it was positive for estrogen and progesterone receptor (ATCC, Perou et al. 2000; Sorlie et al. 2003).
BT474	Cell line was derived from an invasive ductal carcinoma and has been classified to be a Luminal B subtype as it is positive for estrogen, progesterone and human epidermal growth factor receptors 2 (ATCC, Perou et al. 2000; Sorlie et al. 2003).
MDA-MB-231	Cell line was derived from an invasive ductal carcinoma and has been classified as a basal molecular subtype as it negative for estrogen, progesterone and human epidermal growth factor receptors (ATCC, Perou et al. 2000; Sorlie et al. 2003).

Table 2.1: Description of control and cancer cells/cell lines used in breast cancer study

2.3. RNA extraction

Breast cancer cell lines: medium from cells in T25 tissue culture flask was removed and cells were washed in 2mls of PBS. 2mls of TrypLE express, a dissociation reagent (Life Technologies, Paisley, UK) was added to cells flask. The flasks were incubated at 37°C for two minutes. 1 ml of fresh medium was added to neutralise TrypLE express. The content of the flask was transferred into 15ml tube and cells were counted using a haemocytometer. An aliquot of cell suspension containing $5X10^6$ cells was transferred to a sterile 2ml tube. The tube was centrifuged at 1200rpm for five minutes to collect cells and the supernatant completely removed.

Lymphoblastoid, Leukaemia and lymphoma cell lines: cells were grown in T25 tissue culture flask. The content of the flask was transferred into 15ml tube; cells were counted using a haemocytometer. An aliquot of cell suspension containing $5X10^6$ cells was transferred to a sterile 2ml tube. The tube was centrifuged at 1200rpm for five minutes to collect cells and the supernatant completely removed.

Archival patient samples: archival bone marrow samples fixed in 3:1 methanol: acetic acid were spun at 12,000g for 5 mins at room temperature. The supernatant was completely removed and 1ml of absolute ethanol or 96% ethanol was added without mixing or dislodging the cell pellet. The samples were spun at 12,000g for 3 mins at room temperature and supernatant completely removed.

RNA was extracted using Qiagen RNeasy mini Kit (Qiagen, Manchester, UK) by adding 350µl RLT buffer containing 1:10 ratio β-mercaptoethanol to lyse the cells. Cells in RLT buffer were homogenised by vortexing for 1 minute. 350µl of 70% ethanol was added to the homogenised lysate and mixed by pipetting. The mixture was transferred to the RNeasy spin column and centrifuged at 8000g for 15 seconds. 700µl of RW1 buffer was added to the RNeasy spin column and centrifuged at 8000g at 15 seconds to wash the spin column. 500µl of RPE buffer was then added to RNeasy spin column and centrifuged at 8000g for 15 seconds to wash the column. 500µl of RPE buffer was added to the spin column and centrifuged at 8000g for 2 minutes to dry the spin column membrane and to eliminate any residual ethanol. The spin column was transferred to a 1.5 ml collection tube and 29µl of RNase-free water was added directly to the spin column membrane, the sample was then centrifuged at 8000g for one minute to elute the RNA.RNA from the samples was immediately placed on ice and 1µl of 20U/µl Superase In © (Life technologies, Paisley, UK), a RNASE inhibitor was added then the sample was mixed by gentle pipetting.

The concentration of the sample was measured in ng/µl using a spectrophotometer (Nanodrop 2000; Thermoscientific, UK). Reading the A260/280 value assessed the quality of the sample. RNA samples with A260/280 ratios ≥1.8 were used immediately in experiments or stored at -80°C for future use.

2.4. cDNA synthesis by superscript III reverse transcriptase

cDNA was synthesised using Superscript III reverse transcriptase kit (Life technologies, Paisley, UK). Dependant on the RNA yield, approximately 1µg of RNA was used in a 20µl final reaction volume. 1µl of 10mM dNTP MIX (Roche, Burgess Hill, UK), 1µl of 100ng/µl Random Hexamers (Life technologies, Paisley, UK), 1µg of RNA and PCR grade water was added to a sterile 0.2ml PCR tube, mixed gently by pipetting and quickly centrifuged. The mixture was incubated at 65°C for 5 minutes and collected on ice to cool; the tubes were then quickly centrifuged to collect any condensation formed. 4µl of First Strand Buffer (250 mM Tris-HCl, pH 8.3, 375 mM KCl; 15 mM MgCl2), 2µl of 0.1M DTT, 1µl of 20U/µl Superase In © (Life technologies, Paisley, UK), was added to the tube and the contents of the tube was gently mixed by pipetting. The sample was then incubated at 25°C for 2 minutes, 1 µl of Superscript III reverse transcriptase was added and mixed gently by pipetting, the tubes were spun quickly to collect sample and

the tube was incubated at 25°C for 10 minutes followed by 42°C for 50 minutes. The cDNA synthesis reaction was stopped by heating at 70°C for 15 minutes then holding the samples at 4°C.

2.5. Reverse transcriptase- polymerase chain reaction (RT-PCR)

PCR experiments were carried out in 50µl volume reaction using Clonetech Advantage GC 2 PCR Kit (Takara bio Europe/ Clontech, Saint-Germain-en-Laye, France). The 50µl reaction contained: 1 X master mix using 2µl (10pmol/µl) of each primer, 10µl of 5X Advantage GC 2 Buffer, 5µl of GC-Melt, 1µl of 50X dNTP mix, 1µl of 50X Advantage GC2 Polymerase mix and PCR grade water to make up the final volume.

The amplification conditions were 1 cycle of denaturation at 94°C for 3 minutes, followed by 35 cycles of annealing at 60°C (β-actin) or 68°C (*HLXB9*) for 30 seconds, then extension at 68°C for 1 minute, final extension at 68°C for 3 minutes and holding the sample at 4°C.

HB9-WTI and HB9-WT1R primers were used for the first *HLXB9* PCR reaction and HB9-1994-F and HB9-WT-2R primers were used for the nested *HLXB9* PCR reaction *HLXB9* primer sequence: HB9-WT1 forward: 5'-CTTCCAGCTGGACCAGTGGCTG-3' and HB9-WT1 reverse: 5'-CGTCCTCGTCCTCGTCCTCC-3', HB9-1994 forward: 5'-TCCACCGCGGGCATGATCCTG-3' and HB9-WT2 reverse: 5'-GGCCCCAGCAGCTCCTCGGCTC-3'. β-actin gene expression was used as a loading control for *HLXB9* expression. β-actin primer sequences:β-actin forward: 5-'TGACGGGGTCACCCACACTGTGCCCATCTA-3'and β-actin reverse: 5'-CTAGAAGCATTGCGGTGGACGATGGAGGG-3'.

Primer	**Position**	**Exon**
HB9-WTI	808-830	Exon 1
HB9-WT1R	1278-1258	Exon 3
HB9-1994-F	836-857	Exon 1
HB9-WT-2R	1195-1173	Exon 3
β-actin forward	541-570	Exon 4
β-actin reverse	1201-1172	Exon 6

Table 2.2: Description of RT-PCR primers used in study.

PCR products were detected in 1.2% agarose gel containing 0.5µg/ml ethidium bromide. 5µl of PCR product was added to 1µl of loading dye then loaded in agarose gel. The gel was run in 1X TBE buffer (0.089 M Tris, 0.089 boric acid and 2Mm EDTA) at 80v for 45 minutes, then product size were determined by comparing to peqGOLD DNA-Sizer XI (Peqlab, Erlangen, Germany).

2.6. Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR)

Professor Saccone, Catania, Italy, provided results for the qRT-PCR experiments. The experiments were carried out using the method previously described in Leotta et al. 2014. Sybr green Real Time PCR experiments using primers for *HLXB9* and control gene *Tubulin* were carried out in non-cancer breast cells and breast cancer cell lines.

Primer	Sequence
HLXB9 forward	AACCTCCTGGGGAAGTGC
HLXB9 reverse	GGTGAGCATGAGCGAGG
Tubulin forward	GCCAAGAGTGAAGAACAG
Tubulin reverse	GAAGTCCAAGAACTTAGCTG

Table 2.3: Primer sequence used in qRT-PCR experiments

The C_t value of *HLXB9* was normalised using the C_t value of the endogenous control tubulin to obtain the ΔC_t value. The ΔC_t of the samples were then normalised to the calibrator to obtain the $\Delta\Delta\mathbf{C_t}$. Finally, the relative concentration of *HLXB9* in the samples of interest was obtained using the $2^{-\Delta\Delta Ct}$ formula.

2.7. Indirect immunofluorescence (IF)

Breast cancer cells at ≥70% confluency were split 1:4 and 250µl aliquot of cell suspension was added to each well of an 8-well Nunc™ Lab-Tek™ II Chamber™ System (Thermo-scientific, Loughborough, UK). Cells in chamber slides were grown at 37°C with 5% CO_2 in a humidified incubator a day before IF experiments. On the day of experiment, the medium was removed from each well and cells were washed twice in 500µl PBS. The cells were then fixed with 4% paraformaldehyde for 20 minutes at room temperature. After fixation, cells were washed in 500µl PBS then permeabilised by incubating in 500µl PBS containing 0.5% Triton X-100 for 15 minutes. Before antibody was added, aspecific binding sites in cells were blocked in 0.1% BSA in PBS for 1 hour to prevent non-specific binding/staining at 37°C in a humid chamber. Primary antibody: 500µl of anti HB9 antibody or anti tubulin antibody in blocking solution (Sigma-Aldrich, Dorset, UK; 1:100 dilution in blocking solution) was added to each chamber and the slides were incubated at 37°c for 1 hour. After incubation in primary antibody, the cells were washed three times in PBS for five minutes each. Secondary antibody: 500µl of antirabbit FITC for HLXB9, anti-mouse FITC for tubulin (Sigma-Aldrich, Dorset, UK; 1:100 dilution in blocking solution) and/or Alexa Fluor® 568 phalloidin in blocking solution (Invitrogen, Paisley, UK 1:200 dilution in blocking solution) for F-actin staining was added to the slide chamber and the slides were washed in PBS three times for five minutes each. The chamber box was removed; the slides were then mounted in Vectashield (Vector

Laboratories Ltd., Peterborough, UK) containing 4,6-Diamidine-2-phenylindole dihydrochloride (DAPI) and a coverslip was added to slide and sealed with bicycle glue.

Images for analysis were initially captured using *HF14 Leica DM4000 SOP v2* and *Leica AF6000 software* (data not show). Professor Saccone captured the images from the slide with a confocal microscope (Zeiss LSM 700 microscope) so that the cellular localisation of *HLXB9* could be determined.

2.8. Cytogenetic cell harvesting

Cell lines were harvested after reaching approximately 80% confluence by adding 0.01μg/ml Colcemid in Hank's Balanced Salt Solution (HBSS) (Gibco © KaryoMAX, Life Technologies, Paisley, UK) to tissue culture flask and incubating at 37ºC for 2.5 hours. The medium was then transferred into a sterile 15ml tube to prevent cell loss. The cells in the tissue culture flask were then washed in 2mls of PBS and the supernatant transferred into the same tube. 2 mls of TrypLE express a dissociation reagent (Life Technologies, Paisley, UK) was added to tissue flask to lift cells from flask then incubated at 37°C for 2 minutes, 1 ml of fresh medium was added to neutralise the TrypLE express action. The cell suspension was pipetted up and down to aid complete dissociation of cells from flask, the content of the flask was transferred into same 15ml tube as before. The cells were then centrifuged at 1500rpm for 5 minutes.

The supernatant was partially removed and the cells were resuspended in approximately 1ml of supernatant by vortexing. Pre-warmed hypotonic solution (0.075M KCl, Sigma-Aldrich, Dorset, UK) was added drop-wise whilst agitating the tube until 10mls of hypotonic solution was added; the cells were then incubated at 37ºC for 10 minutes. Ten drops of fixative solution (3:1 methanol: glacial acetic acid, Fisher Scientific, Paisley, UK) were added to the cells before the cells were spun at 1000rpm for 5 minutes. The supernatant was removed and pellet was resuspended in 1ml of fixative solution drop by drop. Whilst vortexing, the solution was topped up to 10mls by adding 9mls of fixative slowly, and then incubated for 15 minutes at room temperature before being spun at 1000rpm for 5 minutes. The supernatant was removed and 10mls of fixative was added in the manner described previously, then spun at 1000rpm for 5 minutes. This was repeated until the cell suspension was clear. The cell pellet was resuspended in about 1ml of fixative depending on size of pellet. The cell suspension was used immediately in slide preparation or was stored at -20°C until use.

2.9. Slide preparation

The fixed cell suspension was spun at 1000rpm for 5 minutes. The supernatant was removed then approximately 1ml of fresh 3:1 methanol: glacial acetic fixative was added depending on the size of the pellet. 8µl of cell suspension was added to SuperFrost glass slide (Thermo-scientific, Loughborough, UK). The slides were air-dried and the quality of metaphase and interphase cells on the slides was checked using a phase contrast microscope. The slides were stored in a box overnight at room temperature to age before being used for FISH or stored at -20°C.

2.10. G-banding of breast cancer cell lines

Fresh slides were prepared as previously described on the day there were to be banded. The slides were then aged for 1 hour 30 minutes in the oven at 70°C. The slides were allowed to cool, then placed in a coplin jar containing 2.5% trypsin (GIBCO, Life Technologies, Paisley, UK) in 0.9% NaCl (Sigma-Aldrich, Dorset, UK) mixture for 3-5 seconds then rinsed by twice transferring into a coplin jar containing 0.9% NaCl. The slide was then placed in the Gurrs Giemsa stain (GIBCO, Life Technologies, Paisley, UK) for 5 minutes for staining. After five minutes, the slide was rinsed twice in fresh coplin jars containing Gurrs buffer pH 6.8 (Gibco, Life Technologies, Paisley, UK). The slide was then allowed to air dry, and a coverslip containing cytoseal 60 mounting medium (Thermo scientific, Loughborough, UK) was added to the slide. The slide was allowed to air-dry then images were captured using Zeiss Axioskop 2 microscope and MetaSystems Isis software.

2.11. Preparation of PAC probe RP5-1121A15

P1 artificial chromosome (PAC) clone H_DJ1121A15 (RP5-1121A15) in *E.coli* was purchased from https://bacpac.chori.org. The pac clone DJ1121A15 contains the chromosome region 7q36 region encompassing the *HLXB9* gene (Tosi et al. 2003). The exact coordinates of the PAC Probe were unknown. However, the probe was selected as it has been used in other studies to detect the *HLXB9* gene (Simmons et al. 2002; Tosi et al. 2003; von Bergh et al. 2006; Baglobind et al. 2011; Leotta et al. 2014). Additionally, a map showing the approximate localisation of the probe on 7q is presented as figure A5 in the Appendix.

The PAC clone was grown in streaks on a Luria-Bertani (LB) agar plate with 12.5 µg/ml Kanamycin (Sigma-Aldrich, Dorset, UK) and incubated at 37°C for 24 hours. One colony was selected from the plate

and grown in 10mls of LB medium and 12.5µg/ml of kanamycin overnight. Spinning at 3500rpm for 15 minutes pelleted the bacterial cells. To extract the PAC DNA, the pellet was completely resuspended in 450µl of P1 solution (15mM Tris pH 8.0, 10mM EDTA, 100µg/ml Rnase A), then 450µl of P2 (0.2M NaOH, 1% SDS) was added and mixture was inverted. Finally 450µl of P3 (3M Potassium acetate) was added and inverting mixed the sample. The sample was placed on ice for 5 minutes, then spun at 10000rpm for 20 minutes at 4°C. 800µl of ice-cold isopropanol was added to a 2ml microcentrifuge tube and the supernatant from the sample was transferred into the ice-cold isopropanol. The sample was mixed by inversion and incubated at -20°C overnight. The next day, the probe was spun at 10000rpm for 30 minutes at 4°C to pellet out the DNA. The supernatant was completely removed and 500µl of ice cold 70% ethanol was added to the pellet, the tube was inverted to wash the pellet and tube was spun at 10000rpm for 5 minutes at 4°C. The supernatant was completely removed, and then the pellet was air dried before being resuspended in 40µl of TE buffer (10mM Tris and 1mM EDTA, pH 8.0). The quantity and size of the probe was checked using 1% agarose gel containing 0.5µg/ml ethidium bromide.

5 µl of probe was added to 1µl of loading dye then loaded in agarose gel. The gel was run in 1X TBE buffer (0.089 M Tris, 0.089 boric acid and 2Mm EDTA) at 80v for 1 hour, then product size and quantity were determined by comparing to Peqlab DNA Sizer III marker (Peqlab, Erlangen, Germany).

2.12. Nick Translation

1µg of DNA PAC probe DJ1121A15 was labelled using either 1µM digoxigenin (DIG) -11-dUTP (Roche, Hertfordshire, UK) or 20µM fluorescein (FITC) -12-dUTP (Roche, Hertfordshire, UK) by nick translation. Abbott Molecular nick translation kit (Abbott Molecular, Berkshire, UK) was used to label the probe. A 0.2ml microcentrifuge tube was placed on ice and allowed to cool, then the following components were added in the following order: 1µg of DNA PAC probe DJ1121A15, 2.5µl of either digoxigenin (DIG) -11-dUTP or 20µM fluorescein (FITC) -12-dUTP, 5µl of 0.1mM dTTP, 10µl of dNTP mix, 5µl of 10X nick translation buffer and 10µl of nick translation enzyme. Lastly nuclease free water was added to make a final reaction volume of 50µl. The mixture was quickly centrifuged, then vortexed.

The sample was incubated for 16 hours at 15°C and placed in ice to stop the reaction. The probe size was checked in 2% agarose gel containing 0.5µg/ml ethidium bromide. 5µl of nick translation sample was

added to 1µl of loading dye, and the mixture was then loaded into the agarose gel. The gel was run in 1X TBE buffer (0.089 M Tris, 0.089 boric acid and 2Mm EDTA) at 80v for 1 hour. The probe size and quantity were analysed by comparing to Peqlab DNA size XIII Marker (Peqlab, Erlangen, Germany).

2.13. Probe purification

The labelled probe from the nick translation reaction was added to Illustra microspin G-50 columns (GE Healthcare, Amersham, UK) to remove the unincorporated nucleotides. The spin column was then centrifuged at 6000rpm for 1 minute to collect the purified probe. 10µl of 1mg/ml of Human Cot-1 DNA, 10µl of 3M sodium acetate and 2.25V of 100% ice-cold ethanol was added to the purified probes. The DNA probe was allowed to precipitate overnight by incubating at -20°C. The sample was centrifuged at 13000rpm for 30 minutes at 4°c and the supernatant was discarded. 200µl of ice-cold ethanol was added to wash the DNA pellet; the sample was then centrifuged at 13000rpm for 15 minutes at 4°C. The supernatant was discarded and the DNA pellet was allowed to air dry. The purified probe DNA was resuspended in 5µl TE buffer (10 mM Tris-Cl, pH 7.5, 1 mM EDTA) and 45µl hybridisation buffer (50% (v/v) formamide, 10% (w/v) dextran sulphate, 1% (v/v) Tween 20 in 2 XSSC, pH7.0) and stored at -20°C until use.

2.14. Fluorescence in situ hybridisation

FISH was performed using commercially available whole chromosome 7 (XCP Orange XCyting Chromosome Paint, MetaSystems GmbH, Altlussheim, Germany) and PAC probe RP5-1121A15 labelled with either digoxigenin (DIG) or fluorescein (FITC).

Pre-treatment of Cells *(breast cancer cells only)*

In order to remove excess cytoplasm around metaphase and interphase cells on aged slides and to improve the efficiency of the FISH experiment, a pepsin pre-treatment of slide was carried out as described by Padilla-Nash et al, 2006. Cells on slides were equilibrated in 2X SSC without shaking for 5 minutes at room temperature, 120µl of 0.1mg/ml^{-1} of Rnase A in 2X SSC was added to the slides at room temperature, covered with parafilm and incubated at 37°C for 45minutes. The parafilm was removed and the slides were washed three times in 2X SSC for 5 minutes whilst shaking at room temperature. 10µl of

pepsin 100mg/ml^{-1} was added to an empty clean 100ml glass beaker followed by 100mls of pre-warmed 0.01M HCl and the solution was mixed thoroughly. The pH of the pepsin-HCl solution was adjusted to 2.0, and slides were incubated in the solution at 37°C for 10 seconds. The slides were transferred into petri dish containing 1X PBS and checked under the microscope to confirm there was no longer any cytoplasm surrounding the nucleus and metaphases and that the nuclei integrity was maintained. If there was still cytoplasm visible around the cells, incubation in Pepsin-HCl was repeated. The slides were washed twice for 5 minutes in fresh 1X PBS whilst shaking then transferred into a coupling jar containing 1X PBS/0.05M $MgCl_2$ for 5 minutes. The slides were then fixed in 1% (v/v) formaldehyde/1X PBS/$MgCl_2$ solution for 10 minutes at room temperature without shaking. After fixation was complete, the slides were washed for 5 minutes in 1X PBS at room temperature whilst shaking.

Probe Denaturation

Labelled 1121a15 probe was added to Hybridisation buffer in a 1:4 ratio. 10µl of either labelled 1121a15 (20ng/µl) or whole chromosome 7 paint was denatured 85°C for 8 minutes then at 37°C for 30 minutes to allow preannealing of repetitive sequences.

Sample Denaturation

The cells on slides were dehydrated through a series of 70%, 90% and 100% ethanol for 3 minutes each and allowed to air dry. The cells on slides were denatured in 70% formamide in 2X SSC, pH 7.0 at 72°C for 5 minutes. After denaturation, the slides were immediately transferred into ice-cold 2X SSC on a shaker for 5 minutes at room temperature before being dehydrated using a series of 70%, 90% and 100% ethanol then allowed to air dry.

Hybridisation

Denatured probe was added on slides, covered with a 22X22 glass coverslip and sealed with rubber cement. The rubber cement was allowed to air dry, and the slides were transferred into a humid chamber at 37°C for hybridisation overnight.

Post Hybridisation washes and detection of probe

Following overnight hybridisation, the rubber cement was carefully removed and slides were washed in 2X SSC for 5 minutes on the shaker in the dark to remove coverslip. The slides were transferred into 0.4X SSC at 72°Cf or 5 minutes then washed in 2X SSC for 5 minutes in the dark whilst shaking.

1121A15 probe labelled with FITC: these slides were washed three times in 2X SSC/0.05%Tween20 on the shaker for 5 minutes in the dark. A final wash was done in 1X PBS for 5 minutes on shaker in the dark; the slides were then mounted in Vectashield (Vector Laboratories Ltd., Peterborough, UK) containing DAPI.

1121A15 probe labelled with DIG: 80µl of blocking solution (0.45g BSA in 15ml of 4 X SSC/0.05%Tween20) was added to these slides, which were then covered in parafilm and incubated in a humid chamber at 37°Cfor at least 1 hour. Mouse anti-dig (Sigma-Aldrich, Dorset, UK) in blocking solution was added to slides (final dilution of 1:666). The slides were then covered with parafilm and incubated in humid chamber at 37°C for 30 minutes. After incubation, the slides were washed three times in 4X SSC/0.05%Tween20 for 5 minutes on the shaker. After the washes, rabbit anti-mouse-FITC (Sigma-Aldrich, Dorset, UK) in blocking solution (final dilution of 1:1000) was added to the slides, then the slides were covered in parafilm and incubated in a humid chamber at 37°C for 30 minutes. The slides were then washed three times in 4X SSC/0.05%Tween20 for 5 minutes on the shaker in the dark. Anti-rabbit-FITC (Sigma-Aldrich, Dorset, UK) in blocking solution was added to the slides (final dilution 1:100), and then the slides were covered in parafilm and incubated in a humid chamber at 37°C for 30 minutes. After incubation, the slides were washed three times in 4X SSC/0.05%Tween20 for 5 minutes on the shaker in the dark. A final wash using 1X PBS for 5 minutes on the shaker in the dark was carried out before the slides were mounted in Vectashield (Vector Laboratories Ltd., Peterborough, UK) containing DAPI, and covered with a coverslip and sealed with rubber cement.

Image capture

200-300 hybridised nuclei were captured using an Olympus BX41 fluorescence microscope, 100X oil immersion lens with Smart Capture Software v3 (Digital Scientific, Cambridge, UK).

2.15. Erosion analysis (Nuclear Localisation)

Images were processed using Adobe Photoshop to remove non-specific background signals. The processed images were run through an erosion script donated by Dr Boyle and Dr Bickmore; used with permission from Dr Bridger in IPLab spectrum software. The script outlines the nucleus by detecting the DAPI staining, and it divides the nucleus into five concentric shells of equal volume with the outermost (first) shell denoting the nuclear periphery and the innermost (fifth) shell denoting the nuclear interior.

The script measures the pixel intensity of the DAPI and FITC labelled probe in each shell. The probe signal is normalised by dividing the percentage probe intensity by the percentage DAPI intensity in each shell. In order to simplify the data for analysis, data from shell 1 and 2 have been pooled together to represent the nuclear periphery, data from shell 3 represents an intermediate position in the nucleus, and data from shell 4 and 5 have been pooled together to represent the nuclear interior.

Histograms were plotted from the average signal in each shell with error bars representing the standard error of the mean (average±sem). Statistical analyses were performed using either graphpad prism online tool (http://www.graphpad.com/quickcalcs/ttest1.cfm) or IBM SPSS statistics v18.

2.16. Determining the prognostic signifiance of *HLXB9* overexpression in Breast Cancer Patients.

The prognostic significance of aberrant *HLXB9* expression was evaluated in breast cancer patients using online databases: KM plotter and *BreastMark* (Gyorffy et al. 2013; Madden et al. 2013). In both databases, patient's cohorts were divided into two groups by median gene expression. Patients with lower gene expression than the median were classified as showing low gene expression whilst patients with greater gene expression than the median were classified as showing high gene expression. A survival curve showing the patients' outcome over time by level of gene expression was generated by the databases (as shown in figure 6.11). The difference in patients' survival between the two groups was assigned a prognostic value (hazard ratio with 95% confidence intervals). Hazard ratios of less than 1 indicated a positive impact on prognosis whilst hazard ratios greater than 1 indicated a negative impact on prognosis. The *P* value of the hazard ratio was also calculated. *P* values equal or less than 0.05 indicated the hazard ratio observed was statistically significant.

Patients' outcome were categorised into:

- overall survival (OS), which is the time from either diagnosis or treatment that patients are alive;
- relapse free survival (RFS) or disease free survival (DFS), which is the time after primary treatment that the patient survives without any sign and symptom of cancer returning in the original tumour site;

- disease free metastatic survival (DFMS) or distant disease free survival (DDFS), which is the time after treatment or diagnosis that patients are alive without metastasis to secondary tumour site(s) and
- post progression survival (PPS) which is the time after relapse or metastasis that patients are alive.

Two multi-gene expression tests are now used clinically to determine the possibility of disease reoccurrence in patients in low-grade tumours without lymph node involvement. (Madden et al. 2013; Marrone et al. 2015). These tests are Onco*type* DX, a 21-gene panel test (5 control genes and 16 oncogenes) and MammaPrint, a 70-gene panel test. The 16 oncogenes in the onco*type* DX panel is composed of five proliferation genes (*KI67, STK15, Survivin, CCNB1* and *MYBL2*), two invasion genes (*MMP11* and *CTSL2*), two HER2 associated genes (*GRB7* and *HER2)*, ER defining genes (*ER, PGR, BCL2* and *SCUBE2*) and three other genes (*GSTM1, CD68, BAG1*). The expression of these genes are normalised to the control genes and then scored with the proliferation genes having the highest impact on the overall score. In MammaPrint, 16 genes are associated with good prognosis whilst 45 genes are associated with poor prognosis (However, results could only be obtained on 43 out of 45 genes usng their Entrez ID in *BreastMark*).

The prognostic impact of gene expression was investigated by entering the Entrez ID in *BreastMark*. In KM plotter, Affymetrix ID's for the U133A and U133Plus2 microarray platform were generated from https://biodbnet-abcc.ncifcrf.gov/db/db2db.php. In cases were there were mutiple Affymetrix probes for a particular gene, an average of probe intensity was calculated. *HLXB9*'s Entrez ID is 3110 whilst its Affymetrix probe is 214614_at.

The Entrez ID and the Affymetrix ID/(s) for genes in the Onco*type* DX and MammaPrint tests are presented in tables below.

Entrez ID	**Gene Symbol**	**Affymetrix ID**
4288	*KI67*	212020_s_at; 212021_s_at; 212022_s_at; 212023_s_at
6790	*AURKA*	204092_s_at; 208080_at; 208079_s_at
332	*BICR5*	202095_s_at; 210334_x_at; 202094_at
891	*CCNB1*	214710_s_at; 228729_at
4605	*MYBL2*	201710_at
4320	*MMP11*	203878_s_at; 203876_s_at
1515	*CTSL2*	210074_at
2886	*GRB7*	210761_s_at
2064	*ERBB2*	216836_s_at; 210930_s_at

2099	*ESR1*	211233_x_at; 205225_at; 215552_s_at; 211234_x_at; 217163_at; 211235_s_at; 217190_x_at; 215551_at; 211627_x_at
5241	*PGR*	208305_at; 228554_at
596	*BCL2*	207004_at; 207005_s_at; 203684_s_at; 203685_at
57758	*SCUBE1*	219197_s_at
2944	*GSTM1*	204550_x_at; 215333_x_at
968	*CD68*	203507_at
573	*BAG1*	211475_s_at; 202387_at; 229720_at

Table 2.4: Entrez ID, Gene Symbol and Affymetrix ID of Onco*type* DX genes.

Entrez ID	Gene Symbol	Affy ID
Gene that indicate good prognosis		
8659	*ALDH4*	203722_at; 211552_s_at; 201612_at
8817	*FGF18*	206986_at; 214284_s_at; 206987_x_at; 231382_at; 211029_x_at; 211485_s_at
27113	*BBC3*	211692_s_at
57593	*EBF4*	233850_s_at
57758	*SCUBE2*	219197_s_at
146923	*RUNDC1*	235040_at; 226298_at
8840	*WISP1*	211312_s_at; 229802_at; 235821_at 206796_at
2947	*GSTM3*	202554_s_at; 235867_at
151126	*ZNF385B*	229019_at 1555801_s_at; 1555800_at
146760	*RTN4RL1*	230700_at
10455	*PECI*	218025_s_at; 221780_s_at; 215693_x_at; 219108_x_at
7043	*TGFB3*	209747_at
55351	*HSA250839*	219686_at
58475	*MS4A7*	224358_s_at; 223343_at; 223344_s_at
163	*AP2B1*	200615_s_at; 200612_s_at
79132	*LGP2*	219364_at
Genes that indicate good prognosis		
55321	*C20ORF46*	219958_at
11082	*ESM1*	208394_x_at
9134	*CCNE2*	211814_s_at; 205034_at
54583	*EGLN1*	224314_s_at; 221497_x_at; 223046_at; 223045_at
1058	*CENPA*	210821_x_at; 204962_s_at
9055	*PRC1*	218009_s_at
445815	*AKAP2*	202759_s_at; 202760_s_at
10874	*NMU*	206023_at
3488	*IGFBP5*	203424_s_at; 203425_s_at; 203426_s_at; 211958_at; 211959_at; 1555997_s_at
10531	*MP1*	205273_s_at; 229644_at; 204117_at; 37950_at; 217971_at
57110	*HRASLS*	219983_at; 219984_s_at
8577	*TMEFF1*	205122_at; 205123_s_at
4175	*MCM6*	201930_at; 238977_at
643008	*SMIM5*	229740_at
83879	*CDCA7*	224428_s_at
5984	*RFC4*	204023_at
23594	*ORC6L*	219105_x_at
6515	*SLC2A3*	202497_x_at; 202498_s_at; 216236_s_at; 202499_s_at; 222088_s_at
57211	*GPR126*	213094_at
79791	*FBXO31*	224162_s_at; 219784_at; 219785_s_at; 218875_s_at
1633	*DCK*	203302_at
51514	*L2DTL*	218585_s_at; 222680_s_at
1284	*COL4A2*	211964_at; 211966_at
9833	*KIAA0175*	204825_at
92140	*MTDH*	1559822_s_at; 212248_at; 212250_at; 212251_at; 227277_at
51377	*UCH37*	219960_s_at; 220083_x_at; 229248_at
51560	*RAB6B*	225259_at; 210127_at; 221792_at
160897	*GPR180*	231871_at; 1556019_at; 1556021_at; 1566257_at
79888	*FLJ12443*	201818_at
8293	*SERF1A*	219982_s_at
8476	*PK428*	213595_s_at; 214464_at; 215296_at; 203794_at

10403	*HEC*	204162_at
8833	*GMPS*	214431_at
1894	*ECT2*	234992_x_at; 219787_s_at
4318	*MMP9*	203936_s_at
5019	*OXCT*	202780_at
2781	*GNAZ*	204993_at
2321	*FLT1*	226497_s_at; 210287_s_at; 226498_at; 222033_s_at; 204406_at
2131	*EXT1*	201995_at; 214985_at; 220393_at; 218502_s_at
56942	*DC13*	218447_at
81624	*DIAPH3*	232596_at; 220997_s_at; 229097_at
169714	*QSOX2*	235239_at; 227146_at
286052	*LOC286052*	241370_at
51203	*LOC51203*	218039_at; 219978_s_at
85453	*TSPYL5*	213122_at

Table 2.5: Entrez ID, Gene Symbol and Affymetrix ID of MammaPrint genes.

Chapter 3: Detection of t(7;12)(q36;p13) in paediatric leukaemia using dual colour fluorescence in situ hybridisation (FISH)

This chapter was published as an article in Hematology and Leukemia ISSN 2052-434X.

Owoka T, Vetter M, Federico C, Saccone S and Tosi S. Detection of t(7;12)(q36;p13) in paediatric leukaemia using dual colour fluorescence in situ hybridisation. Hematol Leuk. 2015; 3:4. http://dx.doi.org/10.7243/2052-434X-3-4

3.1. Introduction

Chromosomal translocations are a recurrent finding in cancer. In leukaemia, the detection of particular translocations at diagnosis is used as a biomarker to inform treatment decisions and judge patients prognosis (Betz and Hess 2010; Braoudaki and Tzortzatou-Stathopoulou 2012). The t(7;12)(q36;p13) translocation is found in approximately 30% of paediatric acute myeloid leukaemia (AML) patients and its presence is associated with an unfavourable outcome (Tosi et al. 2003; von Bergh et al. 2006). The chromosomal breakpoint of this translocation on chromosome 7 is heterogeneous, while the breakpoint on chromosome 12 involves the *ETV6* gene at p13 (Tosi et al. 2000; Simmons et al. 2002; Tosi et al. 2003; Naiel et al. 2013).

The use of conventional methods for detecting chromosomal translocation such as G-banding or RT-PCR is not ideal for the detection of t(7;12) (q36;p13) because of the characteristics of the translocation. For instance, the translocation involves subtelomeric regions on 7q36 and 12p13. Subtelomeric regions are difficult to distinguish microscopically because of a lack of distinct characteristics. Therefore there is a possibility of misidentifying this translocation as deletions of 7q (Tosi et al. 2015). Another method commonly used to detect translocations is the amplification of the fusion transcript by RT-PCR. However, this method is not suitable for the detection of t(7;12) (q36;p13) as the fusion transcript is only present in 50% of cases (von Bergh et al. 2006).

Therefore, fluorescent in situ hybridisation (FISH) remains the most suitable method for detecting this abnormality (Tosi et al. 2015). Initially, our research group working in collaboration with MetaSystems GmbH, Germany developed a three-colour FISH probe to identify this translocation (Naiel et al. 2013). However this approach is not ideal as it requires the use of microscopes equipped with at least four fluorescence filter sets in order to distinguish the probe signals from the nuclei rather than the more commonly available three fluorescence filter set in laboratories.

3.2. Aim

The aim of this chapter is to validate the detection of t(7;12)(q36;p13) using a novel dual-colour FISH assay. This aim will be met by carrying out FISH experiments on archival patient samples with confirmed t(7;12)(q36;p13) translocations using a combination of two FISH probes directly labelled with different

fluorochromes. The FISH images will be analysed to determine the sensitivity and accuracy of this diagnostic assay.

3.3. Materials and Methods

Description of patient's samples

Four archival bone marrow cell suspensions in 3:1 methanol:glacial acetic acid from patients diagnosed with haematological disorders previously described as harbouring a t(7;12)(q36;p13) were used for the study (see table 3.1 for patients clinical and cytogenetic characteristics).

Description of dual colour Probe

The dual-colour FISH probe for the detection of t(7;12)(q36;p13) was produced by MetaSystems GmbH, Germany. This is a break-apart probe designed to flank *HLXB9* on 7q36 and *ETV6* on 12p13 (figure 3.1). Figure 3.2 shows typical hybridisation patterns expected in the interphase nucleus from using the probe in FISH experiments (Method described in Chapter 2).

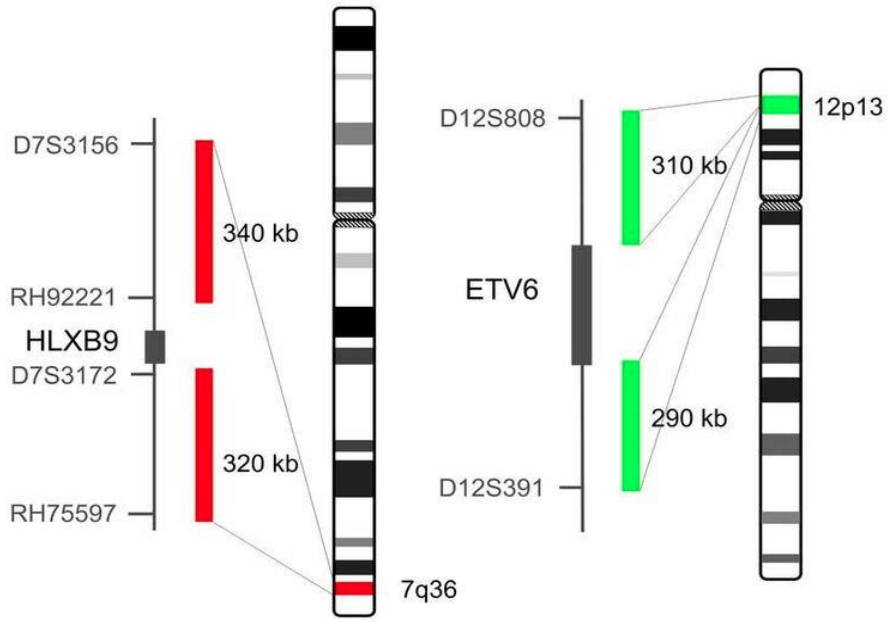

Figure 3.1: Ideograms of chromosome 7 and 12 showing the localisation of FISH probe. The probes were designed to flank the *HLXB9* gene (Red) and the *ETV6* gene (Green). The genomic positions of probe and probe size in kilobase (kb) are located to the left of the ideograms, whilst the localisation of probe at 7q36 and 12p13 are shown to the right of the ideograms.

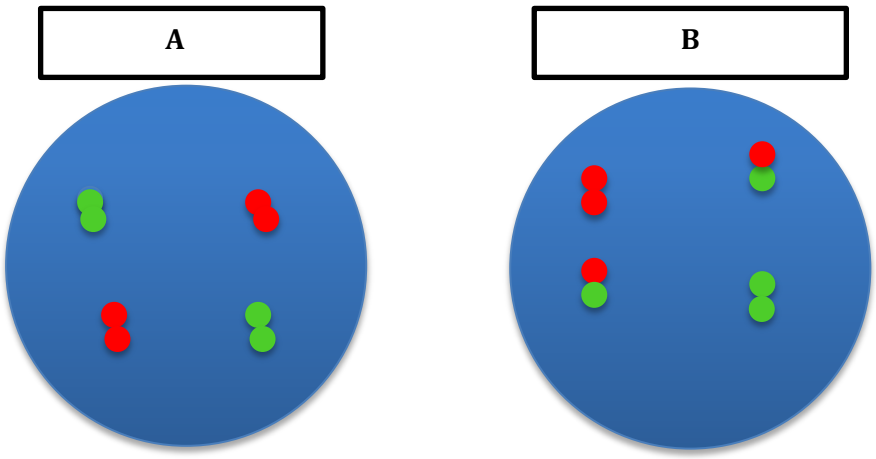

Figure 3.2: Schematic representation of FISH signals expected in Interphase Nuclei using dual colour t(7;12) (q36;p13) probe.
(A) Control nucleus without t(7;12): there are two different green gene signals (2G) representing two copies of *HLXB9* and two different red gene signals (2R) representing two copies of ETV6. As the probes are designed to break apart flanking the genes, these signals are observed as doublet signals.
(B) Typical nucleus harbouring t(7;12): there are two fused red and green signal (2RG) representing t(7;12), one green signal (1G) and one red signal (1R).

3.4. Results

All patients analysed were confirmed to have t(7;12) (q36;p13). Two different hybridisation patterns were observed in the nuclei and metaphases of patients' bone marrow cells. The expected pattern of hybridisation described in figure 3.2 was observed in three out of the four analysed patients (Pt. 1,2 and 3). In these patients, there were two fused red and green signals representing t(7;12), one red signal representing the non-translocated chromosome 7 and one green signal representing the non-translocated chromosome 12 (see figure 3.3).

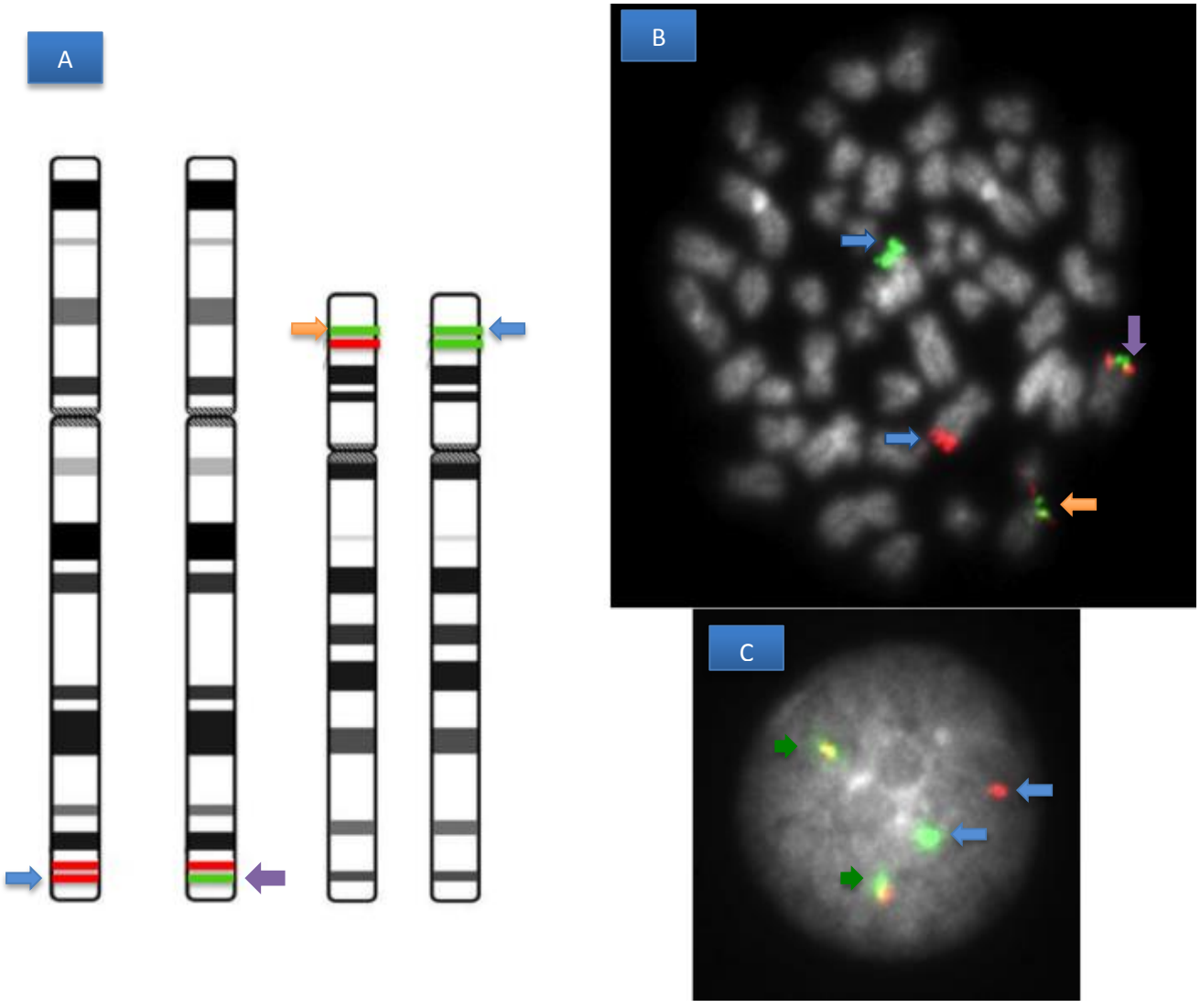

Figure 3.3: Ideograms and example of FISH hybridisation patterns observed in patients 1-3.
(A) Ideograms showing a schematic representation for t(7;12) (q36;p13) observed in patients 1,2 and 3, probe flanking *HLXB9* in red and probe flanking *ETV6* in green. **(B & C)** Dual colour FISH using a break-apart probe flanking *HLXB9* on chromosome 7 (spectrum orange: Red) and *ETV6* on chromosome 12 (FITC: green) on metaphase chromosome and interphase nucleus (counterstained in DAPI).
The non-translocated chromosome 7 and 12 are observed as doublet signals and are shown by blue arrows. The translocation was observed as two co-localised (fused) red and green signals. Purple arrows on ideogram and metaphase cell show the translocated derivative 7 but not in the interphase cell as the origin of the translocation could not be determined in the interphase cell. Orange arrows on ideogram and metaphase cells show the translocated derivative 12. Green arrows in the interphase cell show the translocated derivative 7 and 12.

However, a different pattern of hybridisation was observed in patient 4. This included: one fused red and green signal representing one copy of translocated t(7;12),one red signal representing the non-translocated chromosome 7 and two green signals representing one copy of the normal chromosome 12 and the second green signal representing the second copy of the translocation on chromosome 7. This was assumed as the translocated chromosome 7, although no red signal was observed indicating a deletion on chromosome 7. The deleted 7q could be identified in the interphase chromosome due to a reduction in size of signal and the loss of the doublet signal on chromosome 7 (see figure 3.4C).

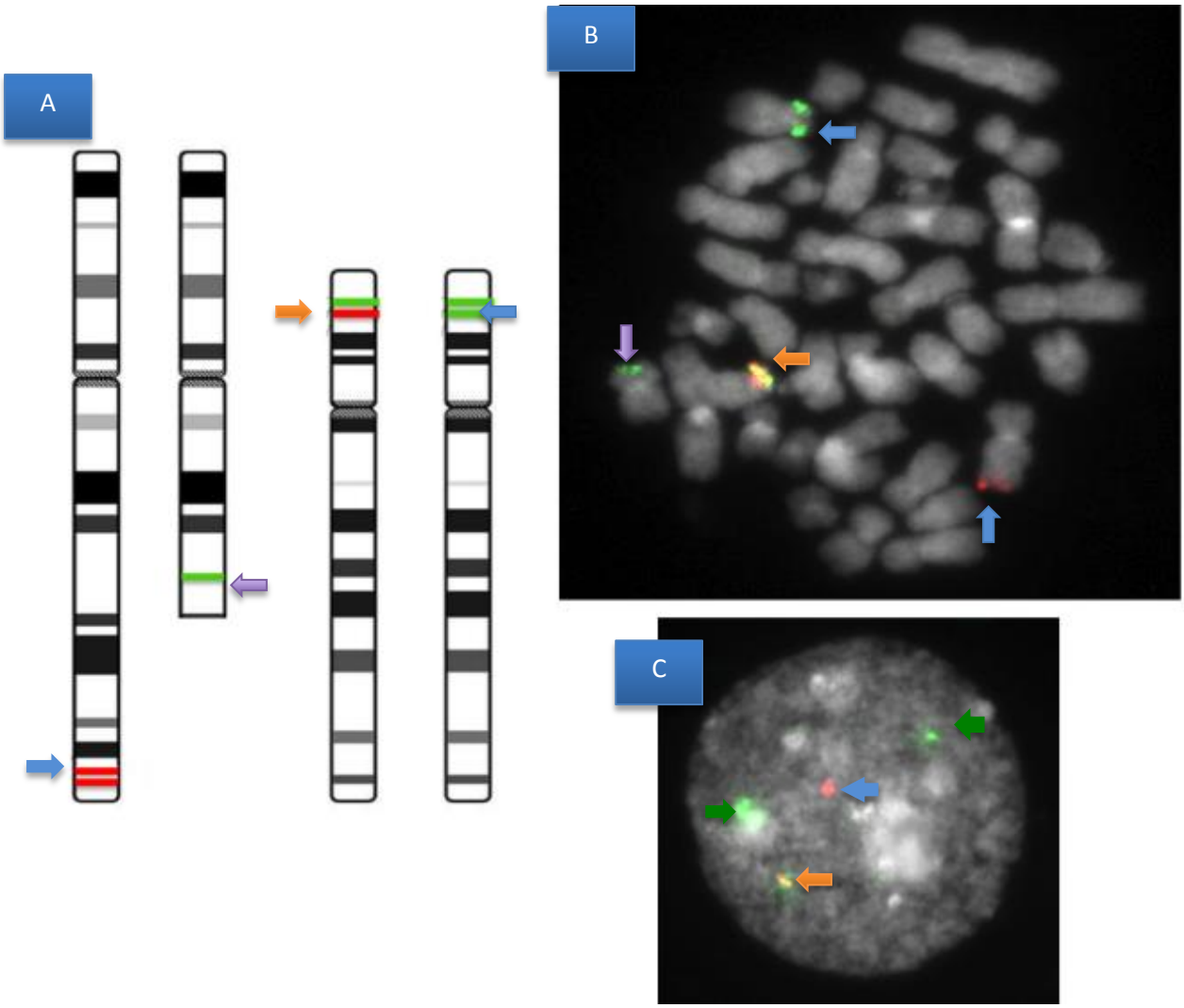

Figure 3.4: Ideograms and example of FISH hybridisation patterns observed in patient 4.
(A) ideograms showing a schematic representation for t(7;12) (q36;p13) observed in patient 4, probe flanking *HLXB9* in red and probe flanking *ETV6* in green. **(B &C)** Dual colour FISH using a break-apart probe flanking *HLXB9* on chromosome 7 (spectrum orange: Red) and *ETV6* (FITC: green) on chromosome 12 on metaphase chromosome and interphase nucleus (counterstained in DAPI). The non-translocated chromosome 7 and 12 were observed as doublet signals and are shown by blue arrows. There was one copy of the translocation observed as one co-localised (fused) red and green signals; this was confirmed as the derivative 12 by analysing the metaphase cell and is shown by an orange arrow. The other green signal was identified as a derivative 7 with del (7q) by analysing the metaphase cell and is shown on the ideogram and metaphase spread by a purple arrow. This distinction could not be made in the interphase cell therefore the copies of chromosome 7 are shown by green arrows in interphase cell.

Using the dual colour break-apart probe, identifying the chromosomal origin of the translocated chromosomes (derivatives) was possible in the metaphase spread but not in the interphase cells.

Patient	Age/Sex	Diagnosis	Karyotype	Total number of cells evaluated	Results
1	5 months/ female	T-ALL later revised as AML	49,XX,+6,t(7;12)(q36;p13),+8,+19[3]/ 50,idem,+18[7]	82	Confirmed, patient previously analysed by FISH as pt. 4 in Tosi et al., 2000
2	5 months/ male	AML-M4	48,XY,+8,+19	78	Confirmed, patient previously analysed by FISH as pt 6 in Tosi et al., 2000 and pt 4 in Tosi et al., 2003
3	3 months/ male	AML-M0	47,XY,t(7;12)(q36,p13),+der(19)	73	Confirmed, patient previously analysed by FISH as pt 5 in Tosi et al., 2000 and pt 3 in Tosi et al., 2003
4	7 months/ female	MDS	46,XX,der(7)t(7;12)(q36;p13)del(7)(q22q36)	155	Confirmed, patient previously analysed by FISH as pt 1 in Tosi et al., 2000; 2003 and Naiel et al., 2013.

Table 3.1. Clinical and cytogenetic data of patients analysed in this study. AML, acute myeloid leukaemia (AML);M4 and M0 are AML FAB classification subtypes, MDS, myelodysplastic syndrome; T-ALL, T-acute lymphoblastic, leukaemia; pt, patient

3.5. Discussion

The dual colour probe enabled the detection of t(7;12)(q36, p13) translocation in all four patients previously described as harbouring the t(7;12) translocation (Tosi et al. 2000; Tosi et al. 2003; Naiel et al. 2013). The probe also allowed the detection of the translocation in a case where a deletion of 7q had occurred; being able to detect this translocation in cases of t(7;12) with an accompanied deletion of 7q will help reduce the misidentification of this translocation as a del 7q. However, a limitation of this dual colour probe was the inability to identify the chromosomal origin of derivatives chromosomes in interphase cells.

In conclusion, this study proved that this dual colour FISH assay can be used to accurately identify the t(7;12)(q36;p13) using a microscope with three single band pass filters to distinguish the single probe signals and the dual colour signals in both metaphase chromosomes and interphase nuclei signal. This validation has proved that this probe set can be used in diagnostic laboratories that use fluorescence microscopes equipped with the conventional filter sets.

Chapter 4: Investigating the expression and the nuclear localisation of *HLXB9* in patients diagnosed with haematological malignancies

Some of the content of this chapter was submitted for publication as a research article to Chromosome Research:

Owoka T, Leotta C, Foster H, Federico C, Bridger J, Saccone S, Tosi S. Repositioning of the *HLXB9* gene in the nucleus of leukaemia cells depends on chromosome 7 rearrangements.

4.1. Introduction

Chromosomal abnormalities of chromosome 7, particularly deletions in the long arm of chromosome 7, (del 7q) are commonly observed in myeloid haematological malignancies (Tosi et al. 1996; van der Straaten et al. 2005; Hasle et al. 2007; Březinová et al. 2007; Döhner et al. 2010; Honda et al. 2015). Clinically, del 7q are associated with a poor prognosis in both MDS and AML with the only option for curative treatment being allogeneic haematological stem cell transplantation (alloSCT) (van der Straaten et al. 2005). However, a significant proportion of del 7q patients treated with alloSCT are prone to relapse, making further research into characterising the genes in this region crucial in the understanding of oncogenesis in these patients (van der Straaten et al. 2005; Březinová et al. 2007; Honda et al. 2015).

Studies carried out on AML infant patients with t(7;12) (q36;p13) have shown an overexpression of *HLXB9* in all cases (von Bergh et al. 2006; Ballabio et al. 2009; Park et al. 2009). The breakpoints in these patients are variable but do not disrupt the *HLXB9* gene, which might explain the observed overexpression of *HLXB9* (Simmons et al. 2002; Tosi et al. 2003; Naiel et al. 2013). In these patients, a fusion transcript was only observed in approximately 50% of cases (Almstrup et al. 2007; Ballabio et al. 2009). This suggests the formation of a fusion protein cannot be the mechanism for oncogenesis in these patients.

However, the fusion of *ETV6* with *HLXB9* has been shown to lead to an altered nuclear localisation of *HLXB9* from the nuclear periphery to the nuclear interior (Ballabio et al. 2009). This change in the nuclear localisation has been suggested to lead to the overexpression of *HLXB9* observed in t(7;12) AML patients (Ballabio et al. 2009). *HLXB9* overexpression has also been reported in the GDM-1 cell line, a cell line derived from an AML patient harbouring a translocation t(6;7) (q23;q36) (Nagel et al. 2005).

The observation that the disruption of 7q36 due to a translocation results in *HLXB9* overexpression poses the questions: whether an interstitial deletion in 7q might lead to the expression of *HLXB9* in a similar way and if an altered nuclear localisation of *HLXB9* accompanies its expression as seen in the translocations.

4.2. Aims

The aim of this chapter is to investigate the expression of *HLXB9* in patients diagnosed with various haematological malignancies. The nuclear localisation of *HLXB9* in patients positive for *HLXB9* expression or patients harbouring deletions in 7q will also be examined to determine if a change in the nuclear localisation of *HLXB9* is affected by its expression or an alteration in the structure of chromosome 7.

In order to meet these aims, the following objectives will be pursued:

- *HLXB9* expression will be investigated in patients diagnosed with various haematological malignancies with or without chromosome 7 abnormalities by RT-PCR,
- Nuclear localisation of *HLXB9* in patients diagnosed with haematological disorders either expressing *HLXB9* or harbouring interstitial deletions of 7q will be analysed by erosion analysis of FISH images,
- Finally, the results from the expression and nuclear localisation experiments will be analysed to determine if there is a link between *HLXB9* nuclear localisation, its expression or the presence of chromosome 7 abnormalities.

4.3. Methods and materials

The methods and materials used in this study are described in Chapter 2.

4.4. Results

4 4.1. Investigating *HLXB9* expression in haematological malignancies using RT-PCR

HLXB9 expression was examined in control (peripheral blood cells, bone marrow cells, lymphoblastoid cell lines) and cancer patients samples and cell lines by RT-PCR. RNA was extracted from samples fixed in methanol:acetic acid and converted into cDNA. RT-PCR experiments were carried out using the cDNA template and *β-actin* primers to check the quality and quanity of the cDNA template. When the cDNA template was determined to be of good quality, RT-PCR using primers for *HLXB9* were carried out. Amplified products from the RT-PCR experiments were electrophoresed in a 1.2% Agarose gel and analysed. Results from these experiments are presented and discussed below.

I. *HLXB9* expression in non-cancer blood cells

a) HLXB9 was expressed in a number of control peripheral blood cells

RT-PCR was carried out on peripheral blood cells from nine control (non-cancer) peripheral blood samples. Peripheral blood cells are mature cells therefore undergo very limited cell division. In order to culture them, phytohaemagglutinin (PHA), a plant mitogen that selectively stimulates lymphocyte proliferation is normally used (O'Donovan et al. 1995; Branco et al. 2008; Ioannou et al. 2015). However, stimulating cells with PHA alters the nuclear volume and the gene expression of lymphocytes (Branco et al. 2008; Ioannou et al. 2015). Therefore, the expression of *HLXB9* was analysed before and after PHA stimulation (when possible) in these experiments. Another reagent used in culturing peripheral blood cells is thymidine. Because obtaining metaphase cells is crucial for the evaluation of patients cytogenetic information; lymphocytes grown in culture are blocked in the DNA synthesis stage of the cell cycle (synchronisation) through the addition of thymidine which is then washed away to increase the proportion of cells harvested in mitosis (Bangs and Donlon 2005). The expression of *HLXB9* was also investigated in peripheral blood samples after thymidine addition to examine any alteration to *HLXB9* expression.

The peripheral blood samples used as controls were divided into two groups:

- **Non-stimulated and stimulated normal peripheral blood (unmatched samples)**: The expression of *HLXB9* was also analysed in six control peripheral blood (described below) samples with or without PHA and/or thymidine.
    - two of the samples were peripheral blood samples (figure 4.1A:lanes 1 and 2) and these samples showed no expression of *HLXB9*
    - two were peripheral blood cells stimulated with PHA (figure 4.1A: lanes 3 and 4) and these samples showed no expression of *HLXB9*
    - the last two samples were peripheral blood cells stimulated with PHA and synchronised with thymidine (figure 4.1A:lanes 5 and 6). These samples showed *HLXB9* expression as seen as faint bands on the gel.
- **Non-stimulated and stimulated normal peripheral blood (matched samples)**: The expression of *HLXB9* was analysed in three control peripheral blood samples (figure 4.1B). The samples were obtained and analysed by RT-PCR before and after stimulation with PHA (Sample ID: pbmc, con 41 and con 43). Two out of the three samples (pbmc and con 43) showed *HLXB9*

expression before stimulation with PHA. After PHA stimulation, all three patients showed *HLXB9* expression (figure 4.4B). The pbmc sample also showed an increase in the expression of *HLXB9* after stimulation with PHA, whilst con 43 showed the same level of expression before and after PHA stimulation.

In summary, 40% of peripheral blood cells (non-stimulated with PHA) showed *HLXB9* expression. In addition, 60% peripheral blood samples stimulated with PHA showed *HLXB9* expression. Both samples stimulated with PHA and synchronised with thymidine also showed *HLXB9* expression. Although the results obtained were mixed, they suggest *HLXB9* might be expressed in peripheral blood cells. There is also a possibility that *HLXB9* expression might increase after stimulation with PHA, although these results have to be confirmed by qRT-PCR. These results also show *HLXB9* is expressed upon thymidine synchronisation of lymphocytes in culture. However, due to the small number of samples (n=2), this cannot be stated conclusively.

b) HLXB9 was expressed in control bone marrow cells

Five non-cancer bone marrow samples were screened for *HLXB9* expression to ascertain if the expression of *HLXB9* in progenitor blood cells was different to mature blood cells circulating in the peripheral blood. All five samples studied show *HLXB9* expression (figure 4.1D). The reason for the variation in expression between bone marrow cells and peripheral blood cells is yet to be determined. However, the results obtained suggest that *HLXB9* may have a function associated with proliferation in blood cells. Therefore *HLXB9* was expressed uniformly in bone marrow cells, cells with an increased proliferative capacity than peripheral blood cells (Korbling and Anderlini 2001).

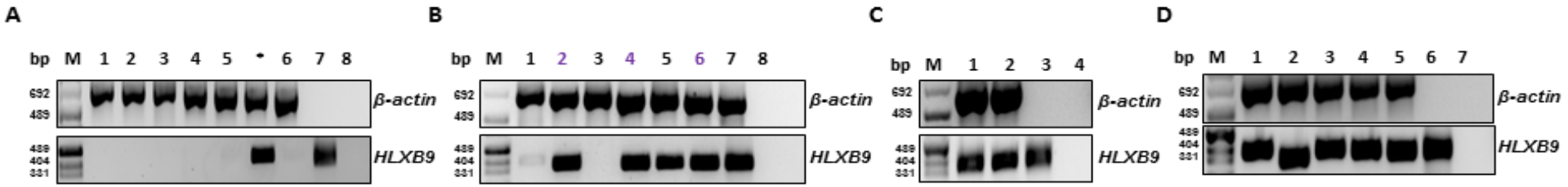

Figure 4.1: Agarose gel electrophoresis analysis of RT-PCR products in control samples.RT-PCR products were electrophoresed in 1.2% Agarose gel at 80 Volts. Amplified products at 660bp obtained using primers for *β-actin* was used as both quality and quantity control. The *β-actin* PCR products in each sample were analysed before the *HLXB9* RT-PCR was carried out to make sure cDNA template from samples did not show signs of degradation or contamination. The *β-actin* bands for each sample were also used to evaluate if the same starting amount of cDNA was used in the PCR reaction.

Figure 4.1A: Agarose gel electrophoresis analysis of RT-PCR products in control (non-cancer) peripheral blood samples. Whole blood sample (n=2); whole blood stimulated with phytohaemagglutinin (n=2); and whole blood samples stimulated with PHA and treated with thymidine (n=2).

LANES:
M= peqGOLD DNA SIZER XI, Lane 1= whole peripheral blood cells (WBC) sample 1, Lane 2= WBC sample 2, Lane 3= WBC sample 3 stimulated with PHA, Lane 4=WBC sample 4 stimulated with PHA, Lane 5= WBC sample 4 stimulated with PHA and treated with thymidine (thy), Lane 6= WBC sample 5 stimulated with PHA and treated with thymidine, Lane 7= (top) negative control- cDNA negative / (bottom) positive control K562 cell line, Lane 8 = negative control mastermix
Lane *= Sample not included as part of this study, patient with constitutional abnormality of 7q

(Bottom) Gel of RT-PCR reaction using *HLXB9* primers showing the amplified product at 359bp. Lanes 1-4 were negative for *HLXB9* expression (no bands in lane on gel), Lane 5 and 6 show faint bands. (Top) RT-PCR control gel showing the quality of cDNA samples using *β-actin* primers.

Figure 4.1B: Agarose gel electrophoresis analysis of RT-PCR products in matched control peripheral blood samples before and after stimulation with PHA.

LANES:
M= peqGOLD DNA SIZER XI, Lane 1= control peripheral blood mononuclear cells (pbmc), Lane 2= pbmc stimulated with PHA,
Lane 3= control peripheral blood (pb) sample-control 41, Lane 4= control 41 stimulated with PHA, Lane 5= control pb sample-control 43, Lane 6= control 43 (npb) stimulated with PHA
Lane 7= positive control-K562 cell line,
Lane 8= negative controls

(Bottom) Gel of RT-PCR reaction using *HLXB9* primers showing the amplified product at 359bp. Before PHA stimulation, two out of three pb control samples showed *HLXB9* expression (pbmc and 41). After PHA stimulation, all three peripheral blood samples showed *HLXB9* expression. (Top) RT-PCR control gel show the quality of cDNA samples using *β-actin* primers.

Figure 4.1C: Agarose gel electrophoresis analysis of RT-PCR products in lymphoblastoid cell lines GM00893 and GM17208B.

LANES:

M= peqGOLD DNA SIZER XI, Lane 1= GM00893, Lane 2= 17208B, Lane 3= (top) negative control-negative cDNA/ (bottom) positive control-K562 cell line, Lane 4= negative control- master mix

(Bottom) Gel of RT-PCR reaction using *HLXB9* primers showing the amplified product at 359bp. Both lymphoblastoid cell lines show (Top) RT-PCR control gel showing the quality of cDNA samples using *β-actin* primers.

Figure 4.1D: Agarose gel electrophoresis analysis of RT-PCR products in five bone marrow samples.

LANES:
M= peqGOLD DNA SIZER XI, Lane 1= control bone marrow cells (BM 1), Lane 2= BM2, Lane 3= BM3, Lane 4= BM4, Lane 5= BM5, Lane 6= (top) negative control-negative cDNA/(bottom) positive control-K562 cell line, Lane 7= negative control-master mix

(Bottom) Gel of RT-PCR reaction using *HLXB9* primers showing the amplified product at 359bp. All five bone marrow samples showed *HLXB9* expression. (Top) RT-PCR control gel showing the quality of cDNA samples using *β-actin* primers.

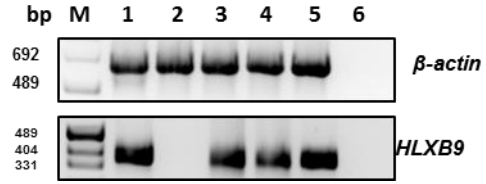

Figure 4.2: Agarose gel electrophoresis analysis of RT-PCR products in leukaemia (K562, GFD-8, GDM1) and lymphoma (CRL 2632 and 2630) derived cell lines. Lanes: M= peqGOLD DNA SIZER XI , Lane 1= K562, Lane 2 = GF-D8, Lane 3 = GDM-1, Lane 4 = CRL 2632, Lane 5 = CRL 2630, Lane 6 = cDNA negative control. *(Bottom) gel of RT-PCR reaction using HLXB9 primers showing the amplified product at 359bp. Four out of the five cell lines investigated showed HLXB9 expression.(Top) RT-PCR control gel showing the quality of cDNA samples using β-actin primers.*

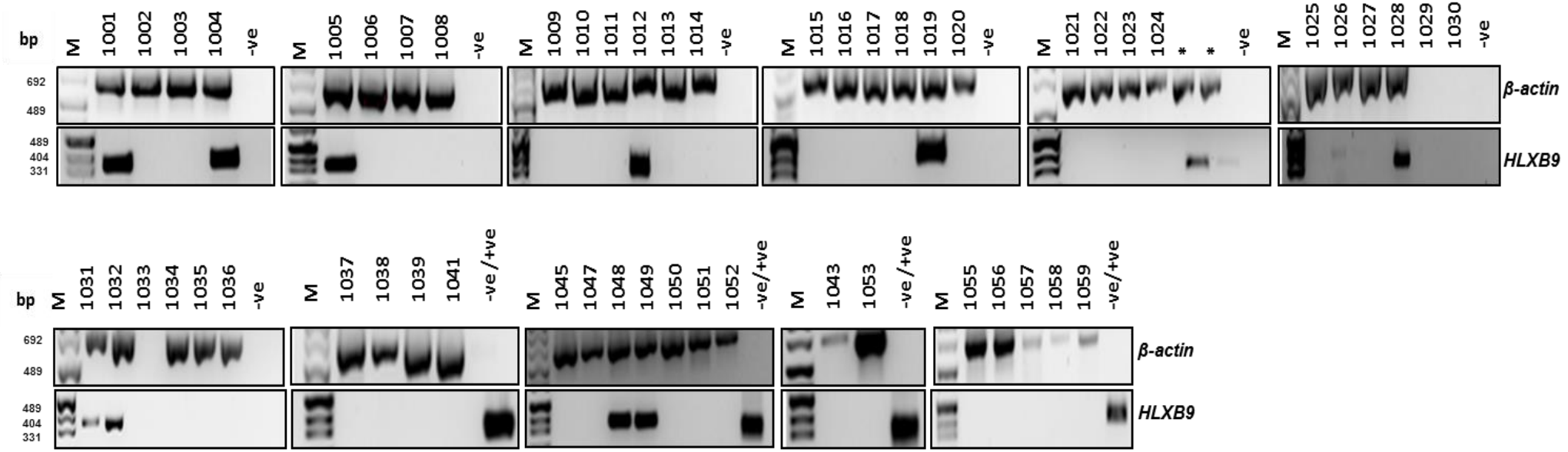

Figure 4.3: Agarose gel electrophoresis analysis of RT-PCR products in patients diagnosed with various haematological disorders. *12 out of the 48 patients diagnosed with various haematological malignancies showed HLXB9 expression.* Patients 1029, 1030, 1033, 1040, and 1042 were not included in the analysis due to low RNA yield and subsequent failure of cDNA synthesis and RT-PCR experiments. * Samples not included in this study as they are breast cancer cell lines. **Lanes:** M= peqGOLD DNA SIZER XI, negative controls (-ve) cDNA negative or master mix and K562 was used as positive controls (+ve)

c) HLXB9 was expressed in control lymphoblastoid cell lines: GM00893 and GM17208B

HLXB9 expression was also examined in two control lymphoblastoid cell lines, GM00893 and GM17208B. Lymphoblastoid cell lines are obtained from Epstein Barr Virus transformation of resting B cells from peripheral blood. They are commonly used as controls in place of lymphocytes in research (Hussain and Mulherkar 2012). Lymphoblastoid cell lines, unlike mature lymphocytes are able to undergo cell division and are grown in culture without become tumorigenic (Hussain and Mulherkar 2012). Both lymphoblastoid cell lines examined showed *HLXB9* expression (figure 4.1C). These results implies a function for *HLXB9* linked to proliferation in these cells

II. *HLXB9* expression in leukaemia and lymphoma derived cell lines

Four out of five cell lines investigated expressed HLXB9

HLXB9 expression was also examined in five leukaemia or lymphoma derived cell lines described in table 4.1 and shown in figure 4.2. The two lymphoma derived cell lines analysed had no chromosome 7 abnormality and both showed *HLXB9* expression. In contrast, the three leukaemia cell lines analysed had confirmed abnormality of chromosome 7. Two out of the three leukaemia derived cell lines showed *HLXB9* expression. These results indicate that *HLXB9* expression can occur independently of the presence of chromosome 7 abnormalities.

Cell line	Diagnosis	Description of Karyotype published in literature	Chr7 abnormality	*HLXB9* expression (Figure 4.2)
K562	Chronic Myeloid Leukaemia	Complex karyotype including 4 copies of chromosome 7 (Gribble et al. 2000)	Yes	Yes
GF-D8	Acute Myeloid Leukaemia (AML) -M1 FAB subtype.	Complex karyotype including 3 copies of chromosome 7 including)t(5;7)t(7;15),inv(7)x2, der(7)t(7;15)del(7) t(7;12) (Tosi et al. 1999)	Yes	No
GDM-1	AML	48,XX, der (2) t(2;12)(q36;q13);t(6;7)(q23;q36) (Nagel et al. 2005)	Yes	Yes
CRL 2632	Non-Hodgkin's B cell lymphoma	Complex karyotype including t(14:18) (Laulier et al. 2011)	No	Yes
CRL 2630	Non Hodgkin's B cell lymphoma	Trisomy of chromosome 11 as published in Gabay et al. 1999 however no trisomy of chromosome 11 was detected in cells in our lab (data not shown)	No	Yes

Table 4.1: Four out of the five cell lines investigated expressed *HLXB9*. *HLXB9* expression in cell lines derived from leukaemia/lymphoma patients showing karyotype published in literature or obtained experimentally.

III. *HLXB9* expression in patients with various haematological malignancies

25% of patients analysed showed HLXB9 expression

48 patients diagnosed with various haematological malignancies were screened for *HLXB9* expression (figure 4.3). 12 patients out of the 48 patients studied showed *HLXB9* expression (25%). Patient's diagnosis and karyotype were checked after obtaining the RT-PCR results to prevent bias in the results from the expression experiments. As the study was interested in expression of *HLXB9* in patients harbouring del 7q, patients were categorised into two groups by karyotype (figure 4.4):

- patients with chromosome 7 abnormalities (n=35)
- patients without chromosome 7 abnormalities (n=8)

The karyotype of five patients analysed were unknown. Seven out of 35 patients with confirmed chromosome 7 abnormality showed *HLXB9* expression (20%). Three out of eight patients without chromosome 7 abnormalities showed *HLXB9* expression (37.5%). Two out of five patients with an unknown karyotype showed *HLXB9* expression (40%). There was no significant difference in the proportion of patients showing *HLXB9* gene expression when patients with chromosome 7 abnormalities were compared to patients without chromosome 7 abnormalities (P=0.290, Pearson chi squared test). These findings suggest that *HLXB9* expression in haematological malignancies is not limited to patients harbouring chromosome 7 abnormalities.

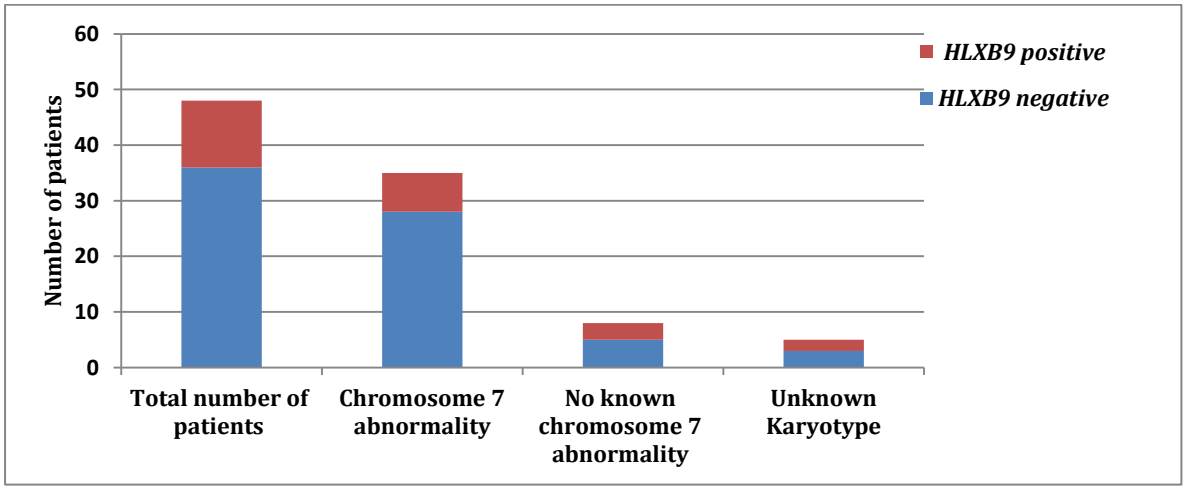

Figure 4.4: The proportion of patients with haematological disorders that showed *HLXB9* expression. (Total number of patients analysed n=48). Patients were categorised by karyotype into two groups; patients with chromosome 7 abnormalities (n=35) and patients without chromosome 7 abnormalities (n=8). The karyotypes of five patients were unknown. 25% of all patient sampled show *HLXB9* expression, 20% of patients with confirmed abnormality of chromosome 7 show *HLXB9* expression, 37.5% of patients without abnormality of chromosome 7 show *HLXB9* expression and 40% with an unknown karyotype showed *HLXB9* expression. There was no significant difference between the proportion of patients with chromosome 7 abnormality showing *HLXB9* expression and the patients without chromosome 7 abnormality showing expression (P=0.290, Pearson chi squared test). This implies that *HLXB9* expression in haematological malignancies can occur independently of the presence of chromosome 7 abnormalities.

During the study, a sample from two time points for one of the patients studied was screened. The two time points are represented by samples 1005 and 1007. Sample 1007 was taken at follow up a year after sample 1005 was taken. Sample 1005 showed *HLXB9* expression whilst sample 1007 was negative for *HLXB9* suggesting that *HLXB9* expression has been lost in this patient. Two different samples 1028 and 1032 were also taken from another patients obtained at the same time and both show *HLXB9* expression. The results from this patient validate the result obtained from the experimental design as the same results were obtained independently.

A greater proportion of patients diagnosed with MDS or MPD showed HLXB9 expression compared to patients diagnosed with other haematological malignancies.

The diagnosis was available for 39 of the patients analysed. Figure 4.5 below categorises the patients analysed by the type of haematological disorder diagnosed. Four out of the nine (44%) patients diagnosed with myelodysplastic/myeloproliferative neoplasms (MDS or MPD) showed *HLXB9* expression. Two out of six (33%) patients diagnosed with acute lymphoblastic leukaemia (ALL) show *HLXB9* expression. Two of the 23 (9%) patients with acute myeloid leukaemia (AML various FAB-subtypes) show *HLXB9* expression; and the only patient diagnosed with chronic myeloid leukaemia screened did not show *HLXB9* expression.

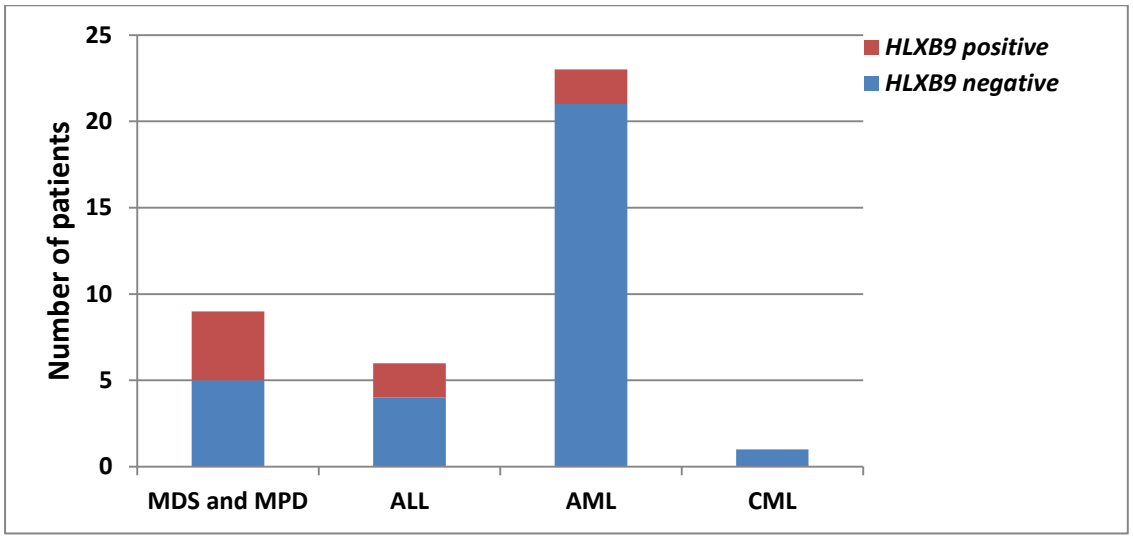

Figure 4.5: Proportion of patients with *HLXB9* expression categorised by the type of haematological disorder diagnosed. 44% of the patients diagnosed with MDS and MPD show *HLXB9* expression, 33% ALL patients show *HLXB9* expression, 6% AML patients studied show *HLXB9* expression and the one CML patient sampled did not show *HLXB9* expression.

The proportion of patients diagnosed with MDS/MPD expressing *HLXB9* was compared to the proportion of patients diagnosed with AML expressing *HLXB9* using a Pearson chi squared test. The difference in proportion was statistically significant (P=0.02), which suggests that *HLXB9* expression occurs more

frequently in MDS/MPD compared to AML. MDS or MPD can be described as premalignant conditions because a proportion of patients go on to develop acute myeloid leukaemia (Björkholm et al. 2014).

The MDS/MPD patients that showed *HLXB9* expression had different types of chromosome 7 abnormalities with heterogeneous breakpoints. Examination of the five patients diagnosed with MDS or MPD that did not show *HLXB9* expression revealed no overlap between the type of chromosome 7 abnormalities present or the breakpoints on chromosome 7. Interestingly, patient 1005 who was diagnosed with MDS had both an interstitial and terminal deletion of chromosome 7 and showed *HLXB9* expression. When the blood sample from this patient was collected a year later (1007), there was an evolution in the karyotype of this patient gaining additional abnormality of chromosome 3 and 12 and a loss of *HLXB9* expression.

The incidence in the proportion of *HLXB9* expression in MPD/MDS patients in addition to the evolution of patient 1005 karyotype suggests expression of *HLXB9* might occur early on in leukaemia development.

HLXB9 expression in patients diagnosed with haematological disorders was not limited to patients that harboured chromosome 7 abnormalities.

12 out of the 48 patients studied show *HLXB9* expression. The diagnosis and karyotype of patients that showed *HLXB9* expression are summarised in table 4.2.

Category	Patient ID	Karyotype	Type of 7q abnormality	Diagnosis
Patients with chromosome 7 abnormality (n=35) seven showed *HLXB9* expression	1001	47,XY, del(1)(p31.1) [12] 47,XY,idem,del(7)(q21q31) [13] 46,XY [5]	Interstitial deletion	MPD
	1005**	46,XY, add(7)(p10), add (7)(q?22)	Interstitial and terminal deletion	MDS
	1012	47,XY, r(7)(p22q22),del(20)(q11q13)	Ring Chromosome	MPD
	1019	46, XX, add (7), i(9) q10, t(10;11) (p12;q21) [20] 46, XX [10]	Add 7	ALL
	1031	46,XY,del(7)(q22q36)	Interstitial deletion	MDS-MPS
	1048	44-45,XX,ins(1;7)(q2?3;dup(q21q22)),-5, 7,+mar1+mar2,inc.ish,ins(1;7)(wcp7+;wcp7+)	Insertion	Unknown
	1049	Full karyotype unknown with t(3;7)(unknown breakpoint)	Translocation	Unknown
Patients with no known chromosome 7 abnormality (n=8): three showed *HLXB9* expression	1026	46, XY, t(5;11) (q31;q23)	None	ANLL-M2
	1027	46,XX, t(6;11) (q13;p13)	None	ALL- ref
	V1001	45-46,XY,-14,+mar[CP2]/46,XY (patient 10 in Ballabio et al.2009)	None	AML-M1
Patients with unknown Karyotype (n=5) two showed *HLXB9* expression	1004	Unknown	Unknown	Unknown
	1028*	Unknown	Unknown	Unknown
	1032*	Unknown	Unknown	Unknown

Table 4.2: Clinical and cytogenetic data of patients positive for *HLXB9* expression in our study. Results are categorised into three groups: patients with known chromosome 7 abnormalities (seven patients out of 35 showed *HLXB9* expression), patients without chromosome 7 abnormalities (three out of eight showed *HLXB9* expression) and patients with unknown karyotype and clinical details (two out of five showed *HLXB9* expression) * Patients 1028 and 1032 are two blood samples obtained from the same patient at the same time. ** Patient 1005 and patient 1007 are samples obtained from the same patient, however, sample 1007 was taken a year later; interestingly patient 1005 expresses *HLXB9* but a year later when sample 1007 was taken no longer show expression of *HLXB9*. Pt v1001 was published as pt 10 in Ballabio et al. 2009.

Three out of the eight patients without chromosome 7 abnormalities show *HLXB9* expression. The first patient, pt V1001 had been previously described as showing *HLXB9* expression by our group (Ballabio et al. 2009). The patient karyotype was: 45-46, XY,-14,+mar[CP2]/46,XY. In this study, two additional patients without chromosome 7 abnormality, pt 1026 and pt 1027 showed *HLXB9* expression. Both patients had a translocation-involving chromosome 11 with different breakpoints. Pt 1026 with t(5;11) (q31;q23) was diagnosed with ANLL-M2 (AML) whilst pt 1027 with t(6;11) (q13;p13) was diagnosed with ALL-ref. Two of five patients with an unknown karyotype showed *HLXB9* expression. In the group of patients with chromosome 7 abnormalities that showed *HLXB9* expression, the breakpoints on

chromosome 7 were heterogeneous and the type of chromosomal abnormality was varied. In addition, a correlation between patients that showed *HLXB9* expression and harboured deletions of 7q could not be identified

A hypothesis of this study was that patients with interstitial deletions of chromosome 7 might be more likely to express *HLXB9* compared to patients without interstitial deletions of chromosome 7. The proportion of patients with an interstitial deletion of 7q that expressed *HLXB9* (three out of seven) was compared to the proportion of patients without interstitial deletion of 7q that expressed *HLXB9* (six out of 36). There was no significant difference in the proportion of patients showing *HLXB9* expression indicating that patients with interstitial deletions of chromosome 7 are not more likely to express *HLXB9* (result of Pearson Squared chi test, P=0.119).

In summary, due to the variability in *HLXB9* expression shown in the controls and patient samples, the difference in the expression of *HLXB9* observed probably differs for reasons yet to be determined. Therefore, further research is required to understand why these patients showed *HLXB9* expression.

75% of patients diagnosed with haematological malignancies did not show HLXB9 expression

36 out of the 48 patients analysed were negative for *HLXB9* expression. The results, diagnosis and cytogenetic data from these patients are categorised in table 4.3 into three main groups: patients with unknown karyotype (n=5); patients without chromosome 7 abnormalities (n=8) and patients with abnormalities of chromosome 7 (n=35).

Description	Type of chromosomal abnormality	Patient ID	Karyotype	Diagnosis
3 out of 5 patients with unknown karyotype did not show *HLXB9* expression		1020	N/A	N/A
		1045	N/A	N/A
		1047	N/A	N/A
5 out of 8 patients without chromosome 7 abnormality did not show *HLXB9* expression		1021	46,XX, der(11)(q23)	ANLL-M1
		1022	6,XX, t(2;11) (q31;p15)	ANLL-M5
		1023	46,XY,t(4;11) (p12;p15)	AML-M1
		1024	46,XX,del (15q)	AML
		1025	45,XY,-Y, t(8;21)(q22q22)/45, idem, add(3)(q?26),del(5)(q?33)/46, XY	ANLL-M5
28 out of 35 patients harbouring abnormalities of 7q did not show *HLXB9* expression	9 had terminal deletions of 7 (q)	1010	47,XY,+Y,del(7)(q31),?dic(13;17)(q13;p11), +mar	AML
		1014	46,XY,del(7)(q22) [18] 46,XY,ins(11;X)(q23;q13q?23) [4] 46, XY,ins(11;X),t(8;10) (q21;q26) [2] 46, XY [16]	ANLL
		1017	46,XXadd (7)(q22)	AML
		1036	46,XY,del(7)(q22),t(8;21)	AML

		1038	46,XX,der(1)(p36),der(3)(q25?),der(7)(q22),der(13)(q13)	ANLL
		1041	46,XX,del(7)(q22?),del(9)(p22),del(12)(p12),-21+mar	ALL
		1043	Complex karyotype +del(7)(q13)	ALL
		1051	del(7q)(unknown breakpoint)	MDS to AML
		1052	Complex karyotype+del(7)(q11q32)	ALL
	4 had interstitial deletions of 7(q)	1003	47,XY,dic(5;17)(q11;p11),del (7)(q22q?),+11,+13,-15,+mar [1] /46,idem,-3 [5] /46,idem,-3,+8,-13 [22] /46,XY [2]	?
		1016	46, XX, del 7 (q22-? q34)	AML-M0
		1055	46,XY,?del (7)(q22q22) [16]/46,XY [4]	MPD
		K1001	46,XY,del(7) (q22q36)	MDS/AREB
	6 had interstitial and terminal deletion of 7(q)	1007**	46,XY,add (7) (p10), add (7) (q?22) [18] 46, idem,del (3) (q2?1), ins (12;?) (q13;?) [12]	MDS
		1008	46,XX, t(9;22) (q34;q11) [3]/ 46,XX, t(7;8) (q21;q11),t(9;22) (q34;q11) [27]	CML- M1 (interstitial deletion) CML-M2 (terminal deletion)
		1009	46?XX,dic (5;17)(q11; p11),-7,der(10)t(10;19)(p11;q11)-19,+2 +mar [7]/ 44,XX,-5,-16,-17, 18,add(21p),add(21p),+ring,+mar [21]/46,XX [2]	AML
		1011	46,XY,t(2;3;7;21)(p15;q21;q22;q22),del(5)(q15 q33)[28] /46, XY [2]	ANLL
		1013	46,XY,inv(11) (q13q22),inv (16)(p13q22) [2]/ 46, XY,idem,del (7) (q22) [20]/ 46,XY,idem, t(3;17) (q2?1;q21), del 7(q32) [8]	AML
		1051	Complex karyotype+del(7)(q ?)	AML
	1 had deletion in 7(q) with unknown breakpoint	1018	44,-Y,-3,del(5)(q13q33), del(7q)(unknown breakpoint),der (11)t(3;11)(q13;p13),del(12)(p11p12),add(13)(q22),-17,add (17)(p11),der(19)t(17;19)(q11;q13),-20,+mar 1,+mar 2	T-AML
	1 had an add7(q)	1034	46,XX,add(7) (q22)	AML
	-7 had other abnormalities of 7(q)	1002	46, XY, +mar.ish -7, del (7) (D7Z1+)	MDS
		1006	43,X,dic(?Y;16)(q11;q13),-5,der(7)inv (7)(p11q22) del (7) (q22q32),-12,der (15) t(15;?17)(p11;q21) add(19)(p13),-20,+2 mar [29]/46,XY?? [1]	MDS
		1015	46,XX,Inv (7) (q31q33-34)	AML
		1035	46, XX, der(7),+mar (ish)7WCP+ (unbalanced translocation of chromosome 7)	AML-M5
		1037	46,XY,+X,-7,add(9) (p22),der(12) t(7;12) (q11;p13),del(14)(q13)/47,XY,idem+mar/46,XY	ALL
		1039	Complex+del 7(q32)? +add 7q	N/A
		1053	46,XY,r(7)(p22q22)	ANLL-M1

Table 4.3: Table showing the clinical and cytogenetic data of patients negative for *HLXB9* expression. Results are categorised into three groups, patients with known chromosome 7 abnormalities (28 patients out of 35 in this group did not show *HLXB9* expression), patients without chromosome 7 abnormalities (5 out of 8 did not show *HLXB9* expression) and patients with unknown karyotype and clinical details (3 out of 5 did not show *HLXB9* expression). ** Patient 1005 and patient 1007 were samples obtained from the same patient however sample 1007 was taken a year later, interestingly the patient had *HLXB9* expression when sample 1005 was collected positive and a year later has no expression of *HLXB9*, this coincides with an evolution of the patient karyotype.

The group of patients with chromosome 7 abnormalities that did not show *HLXB9* expression included nine patients with terminal deletion of chromosome 7 (table 4.3). This result was expected as *HLXB9* maps to 7q36 and is likely to be lost during a deletion involving the terminal region of 7q. Others

abnormality of chromosome 7 observed included interstitial deletion of 7q and addition of 7q. The breakpoints of these abnormalities were heterogeneous.

Patients without chromosome 7 abnormalities that did not show *HLXB9* expression included: one patient with a structural abnormality of the *MLL* gene, two patients with translocations involving *NUP 98* and one patient with an *AML/ETO* translocation.

4 4.2. Nuclear Localisation of *HLXB9* in non-cancer samples and patients diagnosed with haematological malignancies

The nuclear localisation of *HLXB9* was analysed in patients diagnosed with various haematological malignancies identified as either expressing *HLXB9* or habouring interstitial deletion of 7q. The rationale for these experiments was to determine if the nuclear localisation of *HLXB9* was altered upon its expression or due to the deletion and rejoining of 7q as observed in interstitial deletions. 2D FISH experiments were carried out using a probe containing *HLXB9*, and erosion script analysis of the resulting FISH images was used to determine the nuclear localisation of the gene. The nuclear localisation of *HLXB9* was also determined in a number of controls commonly used in research.

I. Nuclear localisation of *HLXB9* in control (non-cancer) blood cells

The nuclear localisation of *HLXB9* was analysed in non-cancer peripheral blood cells, bone marrow cells and lymphoblastoid cell lines; these cells are commonly used as controls in leukaemia research. The nuclear localisation in these samples were analysed to determine the localisation of *HLXB9* in non-cancer cells.

a) HLXB9 had an intermediate nuclear positioning in non-cancer bone marrow cells

The nuclear localisation of *HLXB9* was examined in five mononuclear bone marrow cells, which are mostly comprised of progenitor blood cells. These cells were found to express *HLXB9* by RT-PCR. *HLXB9* was found to have an intermediate nuclear position in all five samples of control bone marrow mononuclear cells analysed (figure 4.6). There was no significant difference in the nuclear positioning of the *HLXB9* gene in the bone marrow samples studied using a one-way ANOVA and Bonferroni post hoc test.

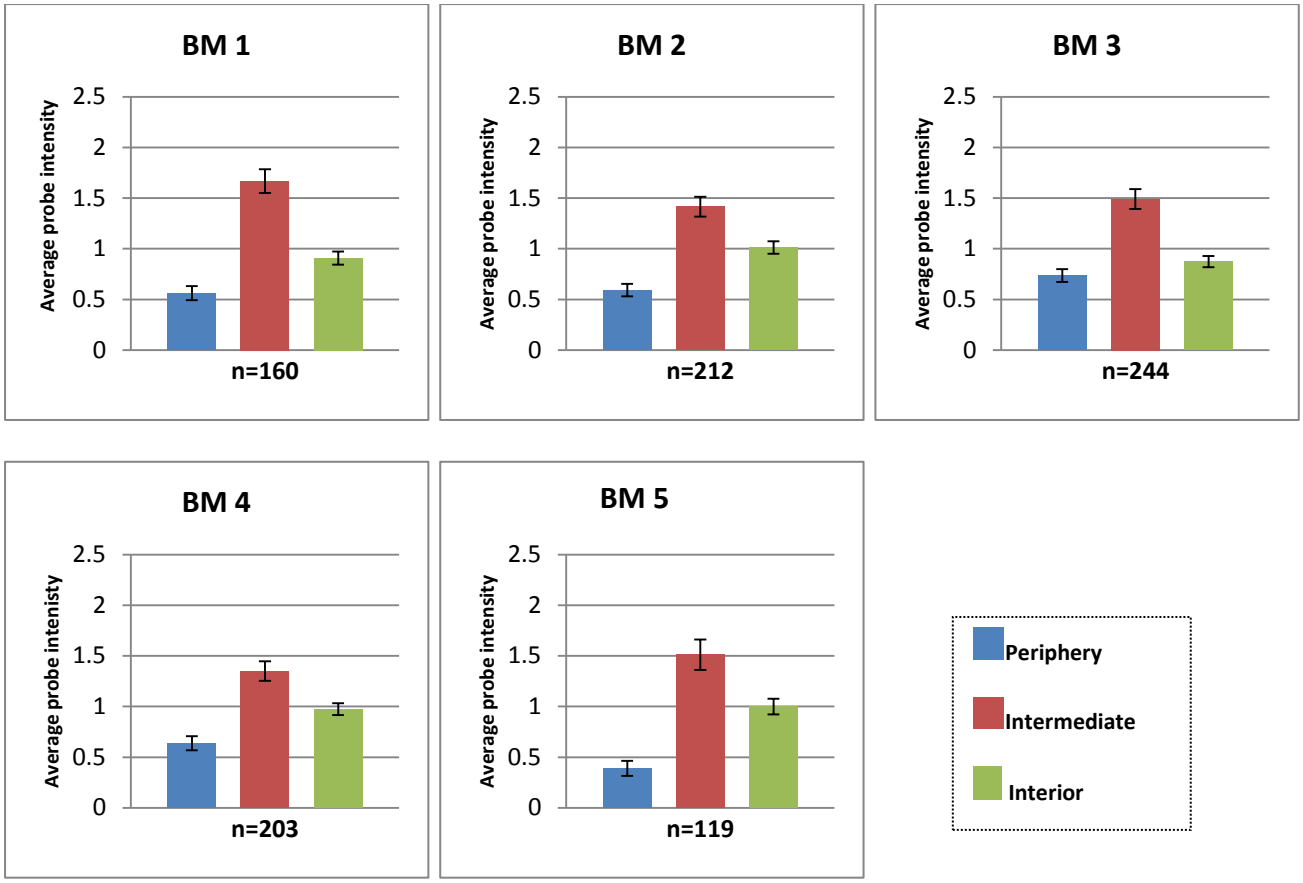

Figure 4.6: *HLXB9* had an intermediate nuclear localisation in bone marrow mononuclear cells. 2D FISH was performed on bone marrow samples fixed in methanol: acetic acid. FISH images were analysed by erosion script analysis. The histograms display the average probe intensity (normalised by DAPI staining) on the y-axis and the number of nuclei studied on the x-axis. The standard error bars represent the standard error of the mean for each nuclear position. There was no significant difference in the nuclear positioning of the *HLXB9* gene in the bone marrow samples studied using a one-way ANOVA and Bonferroni post hoc test.

b) HLXB9 had an intermediate nuclear positioning in control (non-cancer) peripheral blood cells

As previously discussed, due to their limited proliferative capacity, PHA (a plant mitogen) is regularly used to culture lymphocytes in peripheral blood cells in the laboratory. Upon stimulation with PHA, the nuclear volume of lymphocytes is known to increase as shown in figure 4.7 (Branco et al. 2008; Ioannou et al. 2015). Therefore, the nuclear localisation of *HLXB9* was analysed in peripheral blood samples before and after stimulation with PHA (figure 4.8).

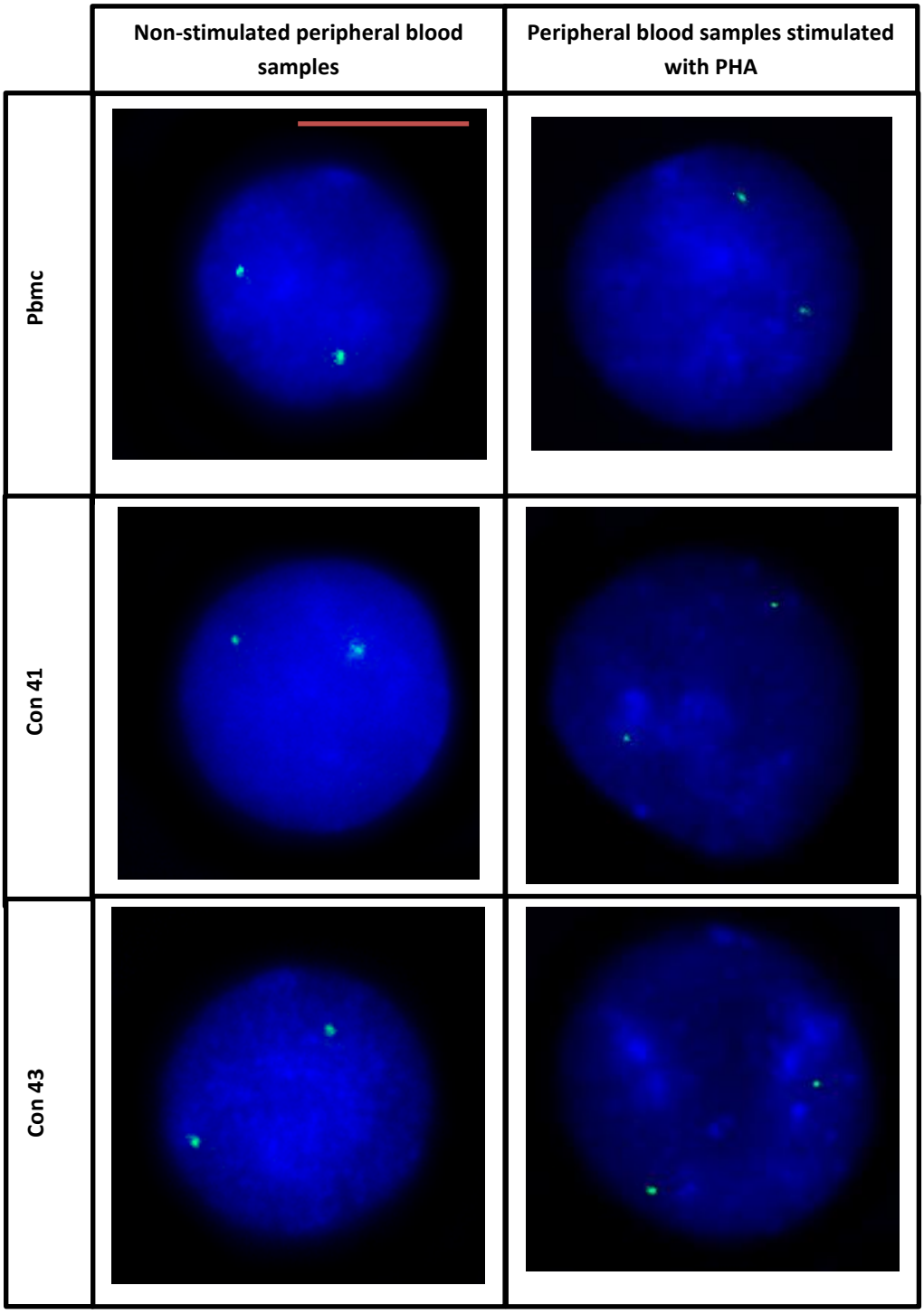

Figure 4.7: Nuclear localisation of *HLXB9* in control peripheral blood samples with or without stimulation with PHA. There was an increase in the nuclear volume of peripheral blood samples stimulated with PHA (Left): non-stimulated peripheral blood samples (right): stimulated peripheral blood samples. 2D FISH images of peripheral blood samples, nuclei counterstained in DAPI (blue), *HLXB9* was labelled with FITC (green). Scale bar= 5μM

I. <u>Control (non-cancer) peripheral blood cells (non-stimulated with PHA)</u>

HLXB9 had an intermediate nuclear positioning in the three control (non-cancer) peripheral blood samples analysed (pbmc, con 41, con 43) (figure 4.8). There was no significant difference

in the localisation of the *HLXB9* gene between the three control samples (results obtained from a one-way ANOVA and Bonferroni post hoc test).

In order to determine if there was a difference in the nuclear localisation of *HLXB9* between peripheral blood cells and bone marrow cells, their nuclear localisation of *HLXB9* were compared using a one-way ANOVA and Bonferroni post hoc test. The nuclear localisation of *HLXB9* in the non-stimulated peripheral blood cells was similar to the bone marrow cells analysed. This indicates the nuclear localisation of *HLXB9* was conserved in blood cell regardless of cell maturity and suggests that regulation of *HLXB9* expression in these cells is probably not a result of a position effect regulation of expression.

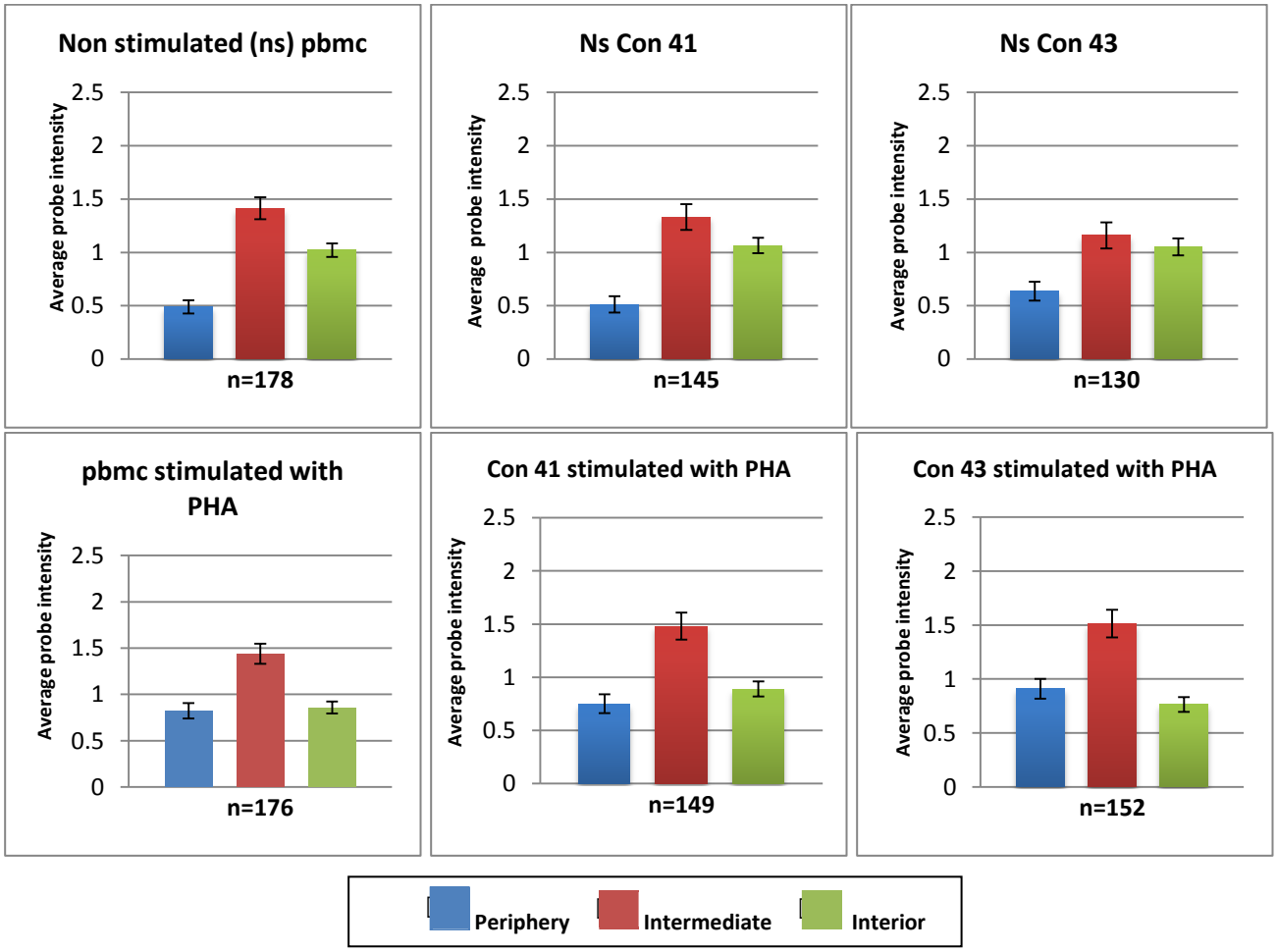

Figure 4.8: *HLXB9* had an intermediate nuclear localisation in peripheral blood mononuclear cells before and after PHA stimulation. After PHA stimulation, there was an increase in the proportion of gene signals in the nuclear periphery in the three samples. However, this increase was only statistically significant in sample pbmc and con 41 samples only. 2D FISH was performed on peripheral blood cells. FISH images were analysed by erosion script analysis. The histogram displays the average probe intensity (normalised by DAPI staining) on the y-axis and the number of nuclei studied on the x-axis the standard error bars represents the standard error of the mean for each nuclear position.

II. Stimulated control (non-cancer) peripheral blood cells

HLXB9 also had an intermediate nuclear positioning in the cells after stimulation with PHA (figure 4.8). There was no significant difference in the localisation of the *HLXB9* gene between the three stimulated peripheral blood samples (results obtained from a one-way ANOVA and Bonferroni post hoc test).

The nuclear localisation of *HLXB9* in peripheral blood samples before and after stimulation with PHA was compared by a paired student t-test. There was an increase in proportion of genes localised in the nuclear periphery in the pbmc and con 41 samples after stimulation with PHA (as shown in table 4.4). In these two samples, *HLXB9* expression increased post stimulation with PHA (figure 4.1B). There was no difference in the nuclear localisation in the con 43 before and after PHA stimulation. This may be because the level of *HLXB9* expression in the non-stimulated con43 sample and the stimulated sample con 43 was similar.

Non-stimulated controls vs.	PBMC	Con 41	Con 43
Stimulated with PHA (paired student t-test)	Increase in proportion of genes localised in peripheral shell upon stimulation with PHA ***.	Increase in proportion of genes localised in peripheral shell upon stimulation with PHA *.	No difference in gene localisation in nuclei.

Table 4.4: Comparison of the nuclear localisation of *HLXB9* in three peripheral blood samples before and after PHA stimulation using a paired student t-test. There was an increase in the proportion of *HLXB9* gene signal localised in the nuclear periphery in the PBMC and con 41 samples after stimulation with PHA. No difference was observed in the nuclear localisation of *HLXB9* in the con 43 before and after PHA stimulation. Results of unpaired student t-test, *$P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$

c) HLXB9 had a peripheral-Intermediate nuclear localisation in peripheral blood cells stimulated with PHA and synchronised with thymidine

The nuclear localisation of *HLXB9* in healthy peripheral blood cells stimulated with PHA and synchronised with thymidine were also analysed as *HLXB9* expression was observed in these samples. *HLXB9* had a peripheral-intermediate nuclear positioning in these cells (figure 4.9). There was no significant difference in the nuclear localisation of *HLXB9* between two samples analysed (results of an unpaired student t-test).

The nuclear localisation of *HLXB9* in these samples was then compared to the peripheral blood control samples: pbmc, con 41 and con 43 by a one-way ANOVA and Bonferroni post hoc test (table 4.5). Upon stimulation with PHA and addition of thymidine, there was an increase in the proportion of genes in the

nuclear periphery in both samples. This means there was a change in the nuclear localisation of *HLXB9* to the nuclear periphery after stimulation with PHA and treatment with thymidine.

The nuclear localisation of *HLXB9* in these controls: con 6 and con 7 were also compared to peripheral blood samples stimulated with only PHA. There was no significant difference in the nuclear localisation of *HLXB9* between the two conditions.

Samples stimulated with PHA and thymidine vs	PBMC	Con41	Con 43
Control 6	Increase in proportion of genes in the nuclear periphery ***	Increase in proportion of genes in the nuclear periphery ***	Increase in proportion of genes in the nuclear periphery. *
Control 7	Increase in proportion of genes in the nuclear periphery. ***	Increase in proportion of genes in the nuclear periphery. ***	Increase in proportion of genes in the nuclear periphery. **

Table 4.5: Comparison of the nuclear localisation of *HLXB9* in peripheral blood samples stimulated with PHA and thymidine (con 6 and con7) to three peripheral blood samples (pbmc, con 41 and con 43). The results indicate a thymidine specific change in gene positioning of *HLXB9*. There was an increase in the proportion of *HLXB9* gene signal localised in the nuclear periphery in pbmc, con 41 and con43. Results of a one-way ANOVA and Bonferroni post hoc test. *$P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$

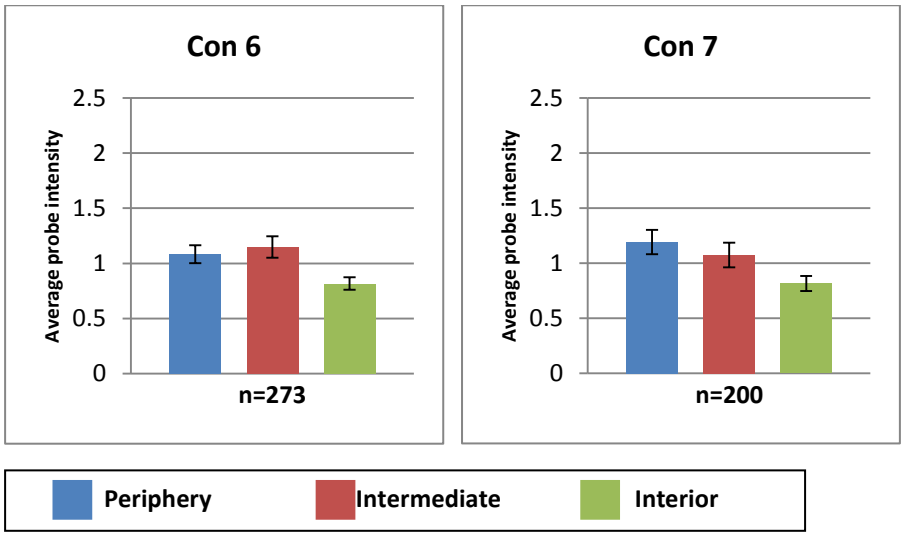

Figure 4.9: *HLXB9* had a peripheral-intermediate nuclear localisation in peripheral blood cells stimulated with PHA and synchronised with thymidine. *2D FISH was performed on peripheral blood cells in methanol: acetic acid fixative. FISH images were analysed by erosion script analysis.* The histogram displays the average probe intensity (normalised by DAPI staining) on the y-axis and the number of nuclei studied on the x-axis. The standard error bars represent the standard error of the mean for each nuclear position. There was no significant difference in the nuclear positioning of the *HLXB9* gene between the two samples stimulated with PHA and thymidine.

d) *HLXB9* had an intermediate nuclear positioning in the two lymphoblastoid cell lines analysed

HLXB9 had an intermediate nuclear positioning in the two lymphoblastoid cell lines analysed- GM17208B, GM00893 (figure 4.10). There was no significant difference in the nuclear positioning of *HLXB9* between the lymphoblastoid cell lines studied (results of an unpaired student t-test).

As lymphoblastoid cell lines are derived by EBV transformation of b-lymphocytes in the peripheral blood, the nuclear localisation of *HLXB9* in lymphoblastoid cell lines was compared to the peripheral blood cells analysed (pbmc, con 41 and con 43). There was no significant difference in the nuclear localisation of *HLXB9* between the two groups (results obtained from a one-way ANOVA and Bonferroni post hoc test).This implies the nuclear positioning of *HLXB9* in the lymphoblastoid cell lines was similar to the localisation of *HLXB9* in peripheral blood cells.

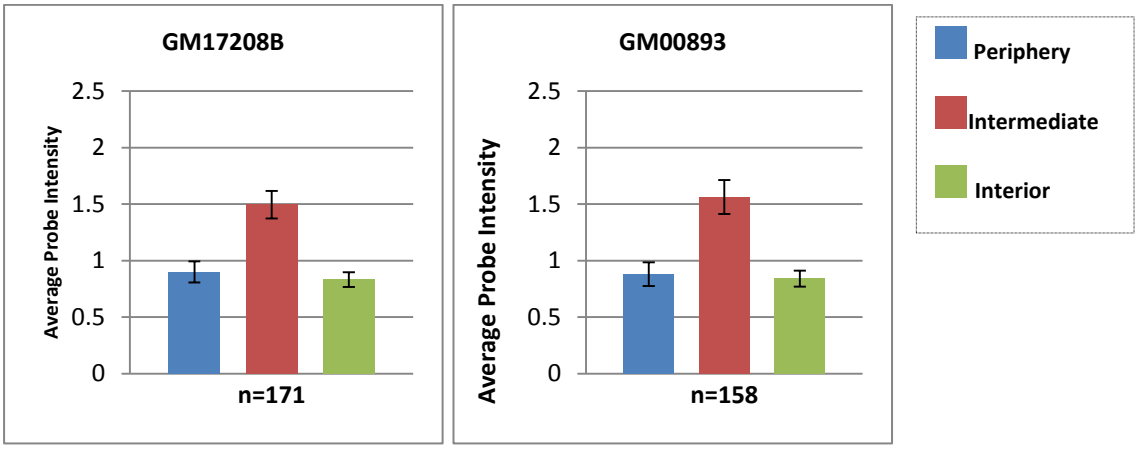

Figure 4.10: *HLXB9* had an intermediate nuclear localisation lymphoblastoid cell lines analysed. 2D FISH was performed on lymphoblastoid cells in methanol: acetic acid fixative. FISH images were analysed by erosion script analysis. The histogram displays average probe intensity (normalised by DAPI staining) on the y-axis and the number of nuclei studied on the x-axis. The standard error bars represent the standard error of the mean for each nuclear position. There was no significant difference in the nuclear positioning of the *HLXB9* gene between the two lymphoblastoid cell lines analysed (result of unpaired student t-test)

II. Nuclear localisation of *HLXB9* in patients with haematological disorders

The nuclear localisation of *HLXB9* was analysed in seven patients diagnosed with various haematological malignancies. The patients were selected based on either confirmed expression of *HLXB9* or the presence of interstitial deletion of 7q (refer to table 4.6). The selection of the patients was carried out in this manner in order to determine if a change in the nuclear localisation of *HLXB9* is dependent on its expression or the presence of chromosomal 7 abnormalities. Nuclear localisation of *HLXB9* was analysed in patients categorised into the following groups:

a. Patients found to express *HLXB9* without chromosome 7 abnormalities

b. Patients diagnosed with various haematological disorders with an interstitial deletion in 7q (del 7q).

The nuclear localisation of *HLXB9* in patients was only compared to control bone marrow samples as the patients' samples analysed were bone marrow samples.

Patient ID	Diagnosis	Karyotype	Type of chromosome 7 abn	*HLXB9* expression	Nuclear Localisation of *HLXB9*
1026	ANLL-M2	46, XY, t(5;11) (q31;q23)	None	+ve	Intermediate
V1001	AML	45-46, XY,-14+mar	None	+ve *(Pt 10 in Ballabio et al. 2009)*	Intermediate
1027	ALL	46,XX, t(6;11) (q13;p13)	None	+ve	Peripheral
C1001	AML-M5	45,XX,del 7(q22q?)/ 46,xx, del 7(q22q?)	Interstitial deletion of 7q	N/A	Interior
K1001	MDS/ AREB	46, XY, del (7) (q22q36)	Interstitial deletion of 7q	-ve	Intermediate to interior
S1001	MDS	45,XX,del(5) (q14),del (7) (q22),del (9)(q22),-17	Interstitial deletion of 7q	N/A	Intermediate to interior
9001	CMML	46, del 7q [25/30]/ 45,-7 [5/30]	Interstitial deletion of 7q and monosomy 7	N/A	Interior

Table 4.6: Clinical and cytogenetic characteristics of patients analysed for nuclear localisation of *HLXB9*.

a. Patients without chromosome 7 abnormality with confirmed expression of *HLXB9*

The nuclear localisation of *HLXB9* was analysed in three patients without chromosome 7 abnormalities that showed *HLXB9* expression by RT-PCR. Two of these patients were diagnosed with AML whilst the third patient was diagnosed with ALL. The nuclear localisation of *HLXB9* in these patients was investigated to determine if a change in the nuclear localisation of *HLXB9* occurred due to expression of *HLXB9*.

- ***HLXB9 had an intermediate nuclear localisation in AML patients without chromosome 7 abnormalities***

  HLXB9 had an intermediate nuclear position in the bone marrow cells of the two AML patients (pt 1026 and ptV1001) without chromosome 7 abnormalities analysed. There was no significant difference in the nuclear localisations between the two patients (result of an unpaired student's t-test).

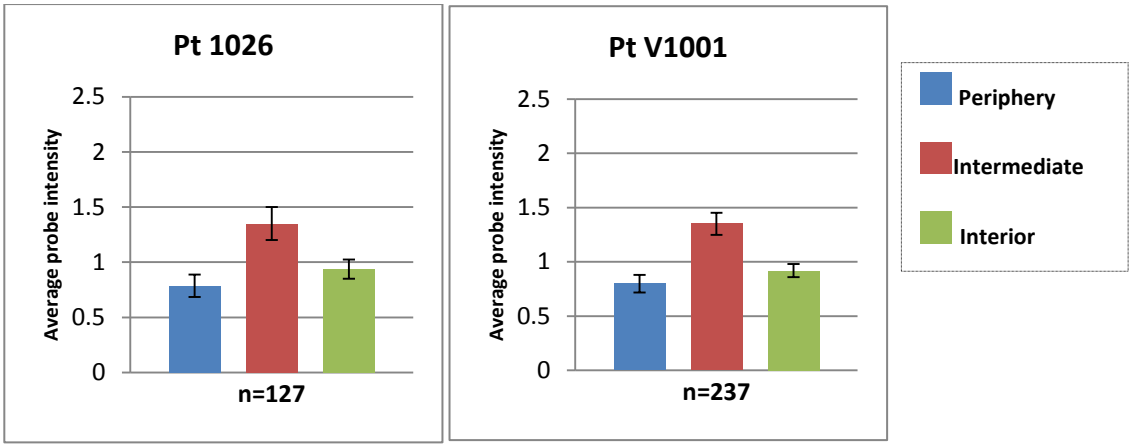

Figure 4.11: *HLXB9* had an intermediate nuclear localisation in the two AML patients without chromosome 7 abnormalities with confirmed expression of *HLXB9*. 2D FISH images were analysed by erosion script analysis. The histogram displays average probe intensity (normalised by DAPI staining) on the y-axis and the number of nuclei studied on the x-axis. The standard error bars represent the standard error of the mean for each nuclear position. There was no significant difference in the nuclear positioning of the *HLXB9* gene between the two AML patients without chromosome 7 abnormalities analysed (result of student t-test)

Additionally, *HLXB9* had a similar nuclear localisation in these patients (pt 1026 and pt V1001) to the control bone marrow cells analysed (result obtained from a one-way ANOVA and Bonferroni post hoc test).

- ***HLXB9* had a peripheral nuclear localisation in an ALL patient without chromosome 7 abnormalities**

An ALL patient without chromosome 7 abnormalities was found to express *HLXB9* by RT-PCR. Although *HLXB9* was expressed in this patient, *HLXB9* was localised in the nuclear periphery.

The nuclear localisation of *HLXB9* in this ALL patient was statistically different to the control bone marrow cells (result obtained from a one-way ANOVA and Bonferroni post hoc test). There was a statistical significant increase in the proportion of *HLXB9* gene signals in the nuclear periphery in this patient compared to the control bone marrow samples. This implies the *HLXB9* nuclear localisation in this patient was different from the nuclear localisation of *HLXB9* observed in control bone marrow cells This suggests an ALL specific nuclear localisation of *HLXB9*. However as only one sample was analysed, further studies with a larger number of samples will be required to confirm this finding.

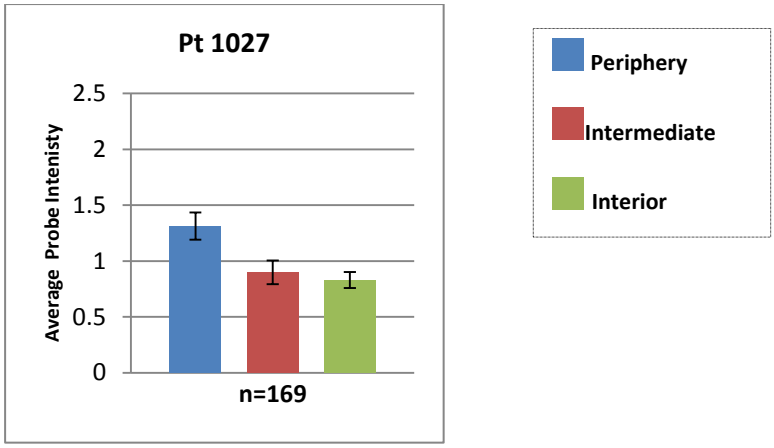

Figure 4.12: *HLXB9* had a peripheral nuclear localisation in patient 1027, an ALL patient without chromosome 7 abnormalities. This patient had a different nuclear localisation of *HLXB9* compared to control bone marrow and peripheral blood samples analysed. 2D FISH images were analysed by erosion script analysis. The histogram displays average probe intensity (normalised by DAPI staining) on the y-axis and the number of nuclei studied on the x-axis. The standard error bars represent the standard error of the mean for each nuclear position.

b. Nuclear localisation of *HLXB9* in patients diagnosed with haematological disorders harbouring an interstitial deletion of chromosome 7

In a previous study, the nuclear localisation of *HLXB9* was analysed in paediatric AML patients with a translocation t(7;12) (q36; p13). In the study, the translocated *HLXB9* was localised in the nuclear interior compared to the non-translocated *HLXB9* (Ballabio et al. 2009). Due to that finding, the nuclear localisation of *HLXB9* was analysed in AML patients with interstitial deletions of chromosome 7. This is because, similar to a translocation event, an interstitial deletion also results in the re-joining of chromosome, which might alter the nuclear localisation of *HLXB9*.

In this study, the nuclear localisation of *HLXB9* was analysed in bone marrow cells of four AML patients with interstitial deletions of chromosome 7. The deletions had various breakpoints (as shown in table 4.8).

All four patients (K1001, C1001, 9001 and S1001) analysed had a similar distribution of *HLXB9* gene signals in the nuclei (result obtained from a one-way ANOVA and Bonferroni post hoc test).

Only one patient, C1001 showed a different nuclear localisation of *HLXB9* compared to the bone marrow cells investigated (results in table 4.9). There was a significant reduction in the proportion of gene signals in the nuclear periphery of pt C1001 compared to three out of the five bone marrow samples analysed (BM2, BM3 and BM4). There was also a reduction in the proportion of the genes signals localised in the nuclear intermediate in pt C1001 compared to all five bone marrow samples analysed. This patient had

an interstitial deletion with an unknown breakpoint therefore the reason for the change in nuclear localisation could not be determined.

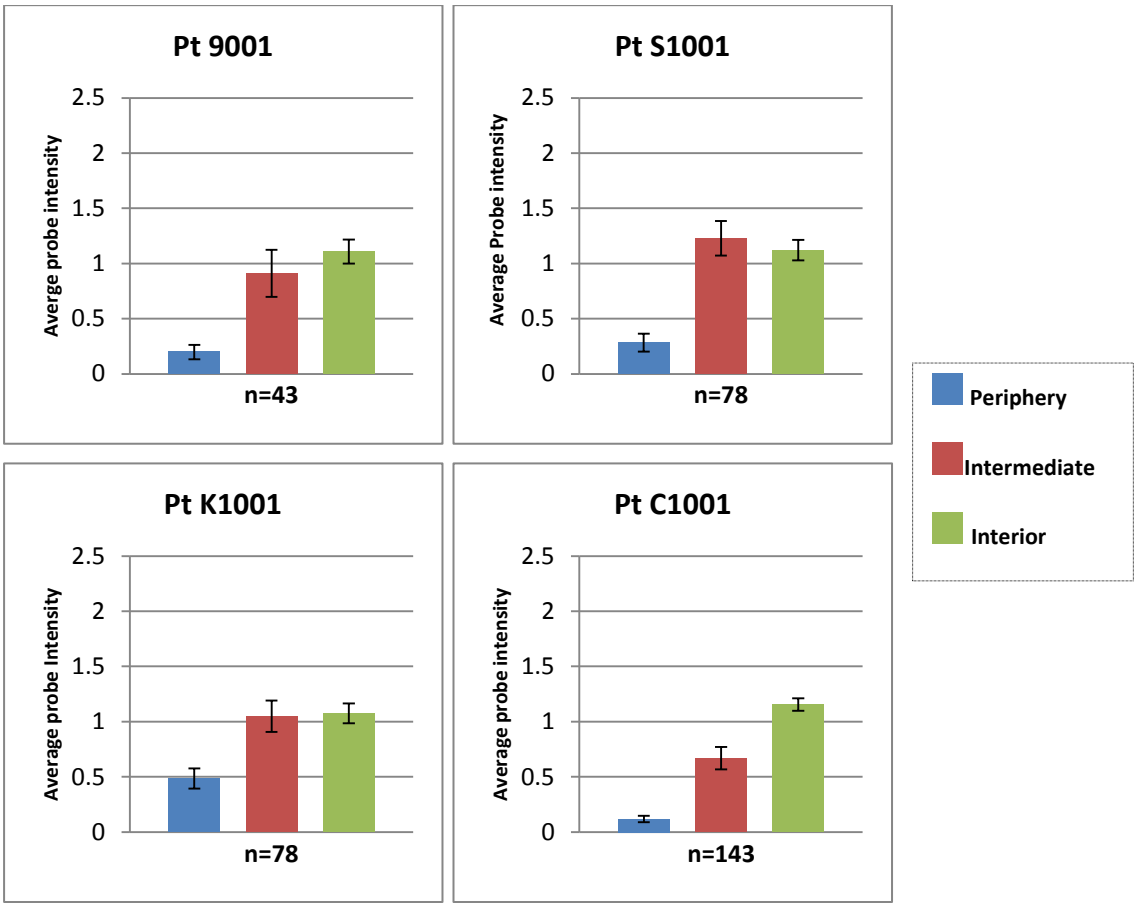

Figure 4.13: Nuclear localisation of HLXB9 in AML patients with interstitial deletions of chromosome 7 (del7q). Four AML patients with interstitial deletions of *HLXB9* (various breakpoint see table 8) were analysed by erosion script analysis of 2D FISH. The histogram displays average probe intensity (normalised by DAPI staining) on the y-axis and the number of nuclei studied on the x-axis. The standard error bars represent the standard error of the mean for each nuclear position. Three of the four patients (C1001, pt 9001 and S1001) analysed had the same distribution of *HLXB9* gene signals in the nuclei (results of an unpaired student t-test). Patient K1001 had an increased proportion of *HLXB9* gene signal in the nuclear periphery compared to the other three patients analysed.

C1001 vs	Periphery	Intermediate	Interior
BM1	No significant difference	Reduction in intermediate***	No significant difference
BM2	Reduction in nuclear periphery*	Reduction in intermediate***	No significant difference
BM3	Reduction in nuclear periphery***	Reduction in intermediate***	No significant difference
BM4	Reduction in nuclear periphery**	Reduction in intermediate**	No significant difference
BM5	No significant difference	Reduction in intermediate**	No significant difference

Table 4.7: Comparison of the nuclear localisation of *HLXB9* in Pt C1001 to control bone marrow samples. There was a significant reduction in the proportion of *HLXB9* gene signals in the nuclear periphery and intermediate in pt C1001 compared to control bone marrow samples; results of a one-way ANOVA and Bonferroni post hoc test. *$P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$

4.5. Discussion

One of the main hypothesis for this study was the deletion in 7q might lead to expression of *HLXB9* as seen in translocations involving 7q such as t(7;12) (q36; p13) and t(6;7) (q23, q36) (Nagel et al. 2005; von Bergh et al. 2006; Ballabio et al. 2009; Wildenhain et al. 2010; Wildenhain et al. 2012). To investigate this hypothesis, *HLXB9* expression was investigated in leukaemia and lymphoma derived cell lines and patients diagnosed with various haematological malignancies. However, prior to the investigation of *HLXB9* expression in cancer samples, *HLXB9* expression was examined in control (non-cancer) bone marrow and peripheral blood cells.

HLXB9 was expressed in all control bone marrow samples and a proportion of control peripheral blood samples

Although the *HLXB9* gene was first discovered in B-lymphocytes, there remains a lack of consensus about whether *HLXB9* is expressed by bone marrow cells or peripheral blood cells (Harrison et al. 1994). Deguchi and colleagues previously reported *HLXB9* expression in CD34 positive enriched bone marrow cells and in unfractionated bone marrow cells (Deguchi and Kehrl 1991). However since their study, other groups have been unable to detect *HLXB9* expression in healthy $CD34^+$ bone marrow cells or $CD34^+$peripheral blood cells by RT-PCR (Nagel et al. 2005; von Bergh et al. 2006; Wildenhain et al. 2012). In this study, *HLXB9* was expressed by all five control unfractionated bone marrow samples analysed. This finding was in line with results obtained by Deguchi and Kehrl (Deguchi and Kehrl 1991). The function of *HLXB9* in blood cell development in the bone marrow is unknown, however, it has been suggested to be associated with maintaining stem cell niche by regulating cell adhesion or cell-to-cell interaction genes (Wildenhain et al. 2012)

HLXB9 was also expressed in two out of five of the peripheral blood cells analysed. All samples analysed were unfractionated white blood cells isolated from peripheral blood. Therefore, the variability in expression of *HLXB9* observed might indicate either variability in the expression of *HLXB9* between healthy individuals or differences in the cellular population of the samples.

Reagents routinely used in culturing lymphocytes from peripheral blood might have an effect on HLXB9 expression

PHA and thymidine are two reagents routinely used in the laboratory to prepare lymphocytes from peripheral blood for cytogenetic analysis. As PHA stimulation has been observed to alter the transcriptional profiles of lymphocytes, the expression of *HLXB9* was investigated in control peripheral blood samples after stimulation with PHA (Branco et al. 2008; Ioannou et al. 2015). *HLXB9* expression was found to increase in two of the three patients analysed. However quantitative PCR experiments are needed to confirm or accurately quantify the change in expression observed.

HLXB9 expression was also examined after peripheral blood cells were treated with thymidine. The two samples treated with thymidine showed *HLXB9* expression. However, this observation was made analysing only two samples, therefore further studies using more samples will be required to confirm this finding.

These results suggests that *HLXB9* expression is induced by reagents used in culturing lymphocytes from blood and implies that *HLXB9* may have a function associated with proliferation in blood cells.

HLXB9 was expressed by four out of five leukaemia/lymphoma cell lines analysed

HLXB9 was found expressed in four of the five cell lines investigated. Three of the five cell lines investigated harboured chromosome 7 abnormalities (GDM-1, K562 and GF-D8). Out of these three cell lines, two showed *HLXB9* expression (GDM-1 and K562). *HLXB9* was expressed in the GDM-1 cell line, a cell line derived from an AML-M4 patient with a complex karyotype that includes the translocation t(6;7) (q23;q36). This cell line had been previously shown to overexpress *HLXB9* (Nagel et al. 2005). *HLXB9* was also found expressed in the K562 cell line; this cell line was developed from a CML patient and has a complex karyotype with ~ 66-72 chromosomes with four copies of chromosome 7. Chromosome 7 specific chromosomal abnormalities in this cell line include i(7)(q10) (del(7) (q31.2;q36);del (7) (q21;q36) (Gribble et al. 2000). This cell line was chosen for investigation because the deletion of chromosome 7 in this cell line is proximal to *HLXB9* but does not disrupt the gene. Another leukaemia cell line chosen for this study was the GF-D8 cell line. This cell line was derived from an AML-M4 patient with a complex karyotype. This cell lines has three abnormal copies of chromosome 7; two copies have an inversion of 7 (breakpoint distal to 7q36) and the third copy has an unbalanced translocation between

chromosome 7 and 15. As the *HLXB9* gene is not deleted, this cell line was expected to express *HXLB9*; however, this was not the case. This might be because the deletion or rearrangement of chromosome 7 in GF-D8 does not result in the juxtaposition of *HLXB9* with an enhancer or a regulatory element.

Two non-Hodgkin lymphoma derived cell lines were also analysed as part of this study (CRL 2630 and CRL 2632). CRL 2630 is a cell line derived from non-Hodgkin B cell lymphoma patient with a diploid karyotype plus trisomy 11, whilst CRL 2632 is a cell line from a non-Hodgkin B cell lymphoma patient with t(14;18)(q32;q21). Both cell lines showed *HLXB9* expression, although there were no known chromosome 7 abnormalities in the cell lines. These results suggest that *HLXB9* expression can occur independently of the presence of chromosomes 7 abnormality.

However, as these results were obtained from cell lines grown in culture, the possibility of *HLXB9* expression being a result of an evolution in the cell lines or by introduction of artefacts during culturing could not be ruled out. Therefore, *HLXB9* expression in patient's samples was analysed.

25% of patients diagnosed with haematological malignancies showed HLXB9 expression.

Bone marrow cells of patients diagnosed with various haematological disorders were investigated for *HLXB9* expression. 12 of the 48 patients investigated showed *HLXB9* expression. This result indicates that *HLXB9* was not expressed by a majority of patients diagnosed with haematological malignancies. 35 of the 48 patients analysed harboured chromosome 7 abnormalities, and 20% of these patients showed *HLXB9* expression. However, a larger proportion of patients without chromosome 7 abnormalities showed *HLXB9* compared to these patients (37.5%). This finding confirms that *HLXB9* expression is not limited to patients harbouring chromosome 7 abnormalities.

44% of patients diagnosed with MPS/MPD (pre-malignant haematological disorders) showed *HLXB9* expression compared to 14% of patients diagnosed with leukaemia suggesting that *HLXB9* expression might occur early on in leukaemogenesis. Only 6% of AML patients showed *HLXB9* expression compared to 33% of ALL patients. This result was contrary to expectation as *HLXB9* expression in leukaemia had be mainly associated with AML patients. Moreover, in previous research, *HLXB9* expression had only been specifically associated with a subset of infant AML patients harbouring the translocation t(7;12) (q36;

p13) (von Bergh et al. 2006; Ballabio et al. 2009; Park et al. 2009; Wildenhain et al. 2012). These results thereby indicate that *HLXB9* expression in AML is not limited to t(7;12) patients.

As translocation involving 7q had previously been associated with expression of *HLXB9*, patients with interstitial deletion of 7q were investigated for *HLXB9* expression. The rationale for selecting these patients was because, similar to a translocation event, the deletion of 7q results in the re-joining of chromosome 7 that might lead to expression of *HLXB9* as observed in the translocation previously discussed. It was expected that a significant proportion of patients harbouring interstitial deletions of 7q (del 7q) would show *HLXB9* expression compared to patients without del7q. However, this was not the case. In fact a larger proportion (40%) of patients with other chromosome 7 abnormalities showed *HLXB9* expression compared to del 7q patients. A unifying characteristic between the three patients with interstitial deletions of 7q that showed *HLXB9* expression could not be identified. Additionally, a common characteristic could also not be identified in patients with interstitial deletion of 7q that did not show *HLXB9* expression. These results suggest the interstitial deletions of 7q do not always give rise to expression of *HLXB9.* This may be because the re-joining event on chromosome 7 does not result in a juxtaposition of *HLXB9* with an enhancer or regulatory element.

Finally, an unsuccessful attempt was made to identify a common characteristic between patients that showed *HLXB9* expression or between patients that did not show *HLXB9* expression. These patients analysed showed different abnormalities and various breakpoints. The lack of consensus also suggests a chromosome 7 independent mechanism of gene expression in patients that showed *HLXB9* expression.

HLXB9 has an intermediate nuclear positioning in control (non-cancer cells)

The second hypothesis for this study was that the nuclear localisation of *HLXB9* might become altered due to a chromosome 7 abnormality as seen in t(7;12) AML patients. Therefore, the nuclear localisation of *HLXB9* was examined and compared in control samples, patients expressing *HLXB9* without chromosome 7 abnormality and patients harbouring interstitial deletion of 7q.

HLXB9 was found to have an intermediate nuclear positioning in the control peripheral blood cells ,bone marrow cells and lymphoblastoid cells analysed. The nuclear localisation of *HLXB9* in bone marrow cells

was similar to the peripheral blood cells analysed. This suggests a conserved nuclear localisation of *HLXB9* in blood cells independent of cell maturity. The results also indicates a conserved nuclear localisation of *HLXB9* between healthy indiduals. This result was in line in findings from two studies that showed no variation in the nuclear localisation of genes in healthy individuals (Wiech et al. 2005; Meaburn et al. 2009).

Reagents routinely used in culturing lymphocytes from peripheral blood affected HLXB9 nuclear localisation

The nuclear localisation of *HLXB9* was also examined in peripheral blood cell after treatment with lymphocyte culturing reagents PHA and thymidine. This was to investigate if *HLXB9* nuclear positioning changes upon lymphocyte stimulation. *HLXB9* was found to have an intermediate nuclear positioing in cells stimulated with PHA. In addition, there was also a significant increase in the proportion of gene signals in the nuclear periphery in two out of three samples indicating a shift in *HLXB9* localisation after stimulation with PHA. This finding was not suprising as previous research had identified an increase in the nuclear volume of lymphocytes upon stimulation with PHA (Branco et al. 2008; Ioannou et al. 2015). Ioannou and colleagues also found a PHA-stimulated reorganisation of the nucleus (Ioannou et al. 2015).

The addition of thymidine resulted in a significant shift in the nuclear localisation of *HLXB9* to the nuclear periphery. *HLXB9* in peripheral blood cells stimulated with PHA and treated with thymidine showed a peripheral- intermediate nuclear positioning of *HLXB9*.

HLXB9 had an intermediate nuclear localisation in AML patients without chromosome 7 abnormalities that expressed HLXB9

The nuclear localisation of *HLXB9* in bone marrow cells of two AML patients that showed *HLXB9* expression without chromosome 7 abnormalities was investigated. *HLXB9* had an intermediate nuclear positioning in these patients. These patients were also shown to express *HLXB9* by RT-PCR. When compared to the control bone marrow cells, *HLXB9* was found to have a similar nuclear localisation in these patients. This indicates there was no change in the nuclear localisation of *HLXB9* in these patients.

HLXB9 had a peripheral nuclear localisation in an ALL patient without chromosome 7 abnormalities that expressed HLXB9

HLXB9 had a peripheral nuclear positioning in bone marrow cells of the only ALL patient investigated. This ALL patient showed *HLXB9* expression by RT-PCR. This finding was interesting as *HLXB9* was previously suggested to be silenced in the nuclear periphery in patients bone marrow cells (Ballabio et al. 2009). The nuclear localisation of *HLXB9* in this patient was different when compared to the control bone marrow and peripheral blood cells analysed. This result indicates an ALL specific nuclear localisation of *HLXB9*. However, due to the small sample size, further studies with a larger number of samples will be needed to confirm this finding.

Only one patient (C1001) with an interstitial deletion of 7q showed a nuclear relocalisation of HLXB9

Only one of the four patients harbouring interstitial deletion of 7q showed a significant change in the nuclear localisation of *HLXB9* compared to the bone marrow cells. This interstitial deletion in this patient had an unknown breakpoint so the reason for the alteration in nuclear localisation of *HLXB9* could not determined.

HLXB9 nuclear positioning and expression

Aberrant expression of *HLXB9* has been found to be a unifying characteristic of AML patients harbouring t(7;12), who showed dismal clinical outcomes. These patients have also been observed to have an altered nuclear localisation of *HLXB9*, therefore it has been hypothesised that a change in the nuclear localisation of this gene is responsible for the altered expression in *HLXB9* observed (Ballabio et al. 2009). Additionally, alteration in *HLXB9* nuclear localisation linked to increased expression has also been observed in neuronal cells where increased expression of *HLXB9* is associated with cellular differentiation (Leotta et al. 2014). Therefore, the main hypothesis for this study was that a change in the nuclear localisation of *HLXB9* could result in the ectopic expression of this gene.

Although there was variability in *HLXB9* expression in the control peripheral blood and bone marrow samples analysed, *HLXB9* had a similar nuclear localisation. Therefore, no link could be found between expression of *HLXB9* and the nuclear localisation in the control samples.

In peripheral blood cells, stimulation with PHA was observed to result in an increased expression of *HLXB9* with an associated increase in the proportion of *HLXB9* gene signal in the nuclear periphery. The change in nuclear localisation of *HLXB9* might be indicative of an increased expression of *HLXB9*. However, there remains a possibility that a change in the nuclear localisation of *HLXB9* could be a result of PHA specific genome reorganisation. A previously study has showed that stimulation with PHA induced a random PHA specific nuclear reorganisation of the genome of lymphocytes (Ioannou et al. 2015).

Only one patient diagnosed with a myeloid lineage haematological disorder showed a difference in the nuclear localisation of *HLXB9*, unfortunately the expression status of *HLXB9* in this patient was unknown. In other patients, there were differences in the expression of *HLXB9,* but no change in nuclear localisation of *HLXB9*. This suggests that *HLXB9* expression is not dependent on its nuclear localisation and there are other mechanism/(s) for the regulation of *HLXB9* gene expression. This finding, although contrary to expectation is not completely surprising as previous studies have shown epigenetic regulation of *HLXB9* in other cancers (Tommasi et al. 2009; Lian et al. 2012).

4.6. Conclusion

In conclusion, a majority of patients diagnosed with haematological disorders do not express *HLXB9*. The expression of *HLXB9* was found to occur independently of the presence of chromosome 7 abnormalities in cell lines and patients diagnosed with various haematological disorders. *HLXB9* expression occurred more frequently in MDS/MPD patients than AML patients, which indicates the possibility of aberrant *HLXB9* expression being an early event in leukaemogenesis.

The nuclear localisation of *HLXB9* was conserved in healthy individuals even though variability in gene expression was observed. The nuclear localisation of *HLXB9* in control samples investigated was also similar to most of the patients analysed although differences in expression was also observed. Therefore, a clear indication from this study is that *HLXB9* expression was not solely regulated by a position effect mechanism. Therefore, further investigation is required to fully understand the regulation of *HLXB9* in leukaemia patients that show aberrant expression of this gene. The role of *HLXB9* in blood cell development or leukaemogenesis remains unknown. Although this study attempted to understand the

mechanism by which *HXLB9* becomes deregulated in leukaemia, the impact of its deregulation is unknown. Therefore, further studies to understand the biological significance of *HLXB9* deregulation particularly in MDS/MPD patients is recommended. Investigating the expression and function of *HLXB9* in these patients might help determine if *HLXB9* could be used as a biomarker for leukaemia prognosis or progression.

Chapter 5: Investigating the differential expression of *HLXB9* as a potential cancer biomarker by datamining online transcriptomics databases

5.1. Introduction

The number of homeobox genes discovered to be aberrantly expressed in cancer continues to increase and is particularly significant in cancer research as the deregulation of these genes can initiate a cascade of events that supports cancer development and progression (Miller et al. 2016; Abate-Shen 2002; Olsen et al. 2016). *HLXB9*, a homeobox gene, has been found to be aberrantly expressed in various cancers (Almstrup et al. 2005; Ballabio et al. 2009; Wilkens et al. 2011; Hu et al. 2012; Shi et al. 2013; Bailey et al. 2016).

The development of methods that can be used to measure whole genome gene expression such as microarray, RNA Sequencing (RNA-Seq) and next generation sequencing (NGS) has led to the deposition of data from these experiments into data banks. These data banks such as gene expression omnibus (GEO), the cancer genome atlas (TCGA) and the international cancer genome consortium (ICGC) can be accessed by the wider scientific community for data analysis. Databases have now been developed, and these can be used to interrogate datasets from various sources to aid the study of differential gene expression in disease states. Whole genome gene expression has been particularly useful in cancer research leading to the identification of potential targets for therapy, genomic subtypes of diseases and genes whose expression correlate with cancer development and progression (Rhodes et al. 2009; Bailey et al. 2016).

5.2. Aims

The main aim of this study is to determine whether *HLXB9* could be used as a biomarker in cancer. In order to achieve the aim, the following objectives will be pursued by:

I. Identifying cancers with differential expression of *HLXB9* using a number of freely available transcriptomics databases;
II. Analysing the identified differential expression of *HLXB9* to determine if it is associated with particular cancer types and stages;
III. Evaluating the results obtained from objectives I and II to determine if aberrant *HLXB9* expression is an indicator of prognosis or disease progression.

5.3. Methods

In order to investigate *HLXB9* expression in various cancers, the following gene expression databases were analysed and compared: Oncomine, Gene expression across normal and tumour tissue (GENT), FIREBROWSE and BioXpress.

Oncomine (Rhodes et al. 2004; Rhodes et al. 2007)

Oncomine (Research Edition) is a web-based cancer microarray database. It contains datasets on approximately 18,000 microarray studies for more than 20 different cancer types. Oncomine has three freely available analytical tools: the normal to cancer analysis tool can be used to compared gene expression in non-cancer and cancer samples; the cancer histological tool can be used to compare differential gene expression between histological subtypes of cancer; and the gene outlier tool which can be used to identify subpopulations of low and high expression in a cancer type. Used together, these tools can provide insights into the differential expression of genes in cancer.

In this study, to reduce the possibility of including false positives and artefacts, the parameters used in the normal to cancer analysis, cancer histology analysis and outlier analysis was set stringently at *P* value of ≤ 0.001, gene expression fold change of ≥2 and gene rank of top 10 per cent over/under expressed genes. In the outlier analysis, an outlier of 90-percentile expression value was categorised as overexpression and an outlier of 5-percentile expression value was categorised as underexpression.

GENT (Shin et al. 2011)

GENT is a web based cancer microarray database that can be used to identify outliers of gene expression, especially overexpression in a subset of the cancer populations. The database has microarray data across two different Affymetrix platforms i.e. the U133A microarray platform, which has a smaller number of gene probes, and the U133 Plus 2 platform, which has a more comprehensive number of gene probes. The gene expression values for *HLXB9* could be found on both platforms. In GENT, raw expression and transformed values are presented as boxplots. The transformed $\log_2$ value is more useful compared to the raw values as the data has been normalised without removing the relationship between the datasets. This enables a direct comparison of the minimum, lower quartile, median, upper quartile maximum expression and outlier expression values can be made between normal and cancer samples. By comparing these

expression values between the normal and cancer samples, cancers with outliers of expression of *HLXB9/MNX1* can be identified. The GENT database holds over 34,000 tissue samples with 9,000 cancer samples and 4,000 normal samples on the U133A platform and over 18,000 cancer samples and 3400 normal samples on the U133 Plus 2 platform. Datasets on *HLXB9* gene expression from both microarray platforms were downloaded and analysed using IBM SPSS Statistics v20 (see table 5.3 and table 5.4). Fold change was calculated as average expression levels of cancer samples divided by average probe expression level of control samples. The log2 values of the obtained fold change was also calculated and presented in the aforementioned tables.

FIREBROWSE http://firebrowse.org/ (no associated publication)

FIREBROWSE is also a web based gene expression viewer database developed by the Broad Institute and is based on the cancer genome atlas (TCGA) datasets. This database has experimental data from RNA-Seq experiments on samples from over 11,000 cancer patients. Utilising a 'view expression profile' tool, gene expression can be compared between normal and cancer samples in the database. The results from the analysis are presented as box plots with values for the lower quartile, median, upper quartile and number of samples in analysis. In this database, cancers with a fold change of ≥2 compared to normal samples were considered to overexpress *HLXB9* whilst cancers with a fold change of ≤ 0.5 were considered to underexpress *HLXB9*.

BioXpress (Wan et al. 2015)

BioXpress v 1.0 is a gene expression database with RNA-Seq data on 64 cancer types and over 6000 patient samples from different resources such as the TGCA, ICGC, the expression atlas and research publications. This database can be used to compare differential gene expression between normal and cancer samples and the frequency of the observed change by cancer types. The results are presented in a graph showing the percentage of overexpression and underexpression observed. *HLXB9* expression in a cancer sample was compared to its matched normal sample pair and classified into overexpression or under expression using the following method: overexpression was determined to be a fold change of greater than zero compared to its matched normal sample pair and underexpression was determined to a fold change of less than zero compared to its matched normal sample pair. The change in expression

observed was then analysed to determine its statistical significance. The frequency of statistically significant change in expression was presented in a graph as shown in figure 5.7.

5.4. Results

HLXB9 expression in various cancers was analysed using the previously mentioned online databases in order to identify cancers with differential expression of *HLXB9*. Results from these analyses are divided into two sections. Section I, provides an overview on *HLXB9* expression in various cancers according to database whilst Section II discusses the results obtained according to cancer type.

I. Overview on *HLXB9* expression in cancer according to database

Results on differential expression of *HLXB9* expression in various cancers are discussed by the database they were obtained from in this section.

a. Oncomine

Using the normal to cancer tool in Oncomine, eleven cancer types with altered expression of *HLXB9* were identified. Out of the eleven cancer types, eight (brain and central nervous system cancer, breast cancer, cervical cancer, oesophageal cancer, kidney cancer, lung cancer, lymphoma and melanoma) showed upregulation of *HLXB9* indicating a pro-oncogenic role for *HLXB9* in most cancers identified. Downregulation of *HLXB9* was found in kidney cancer, leukaemia and pancreatic cancer suggesting a loss of tumour suppressor function of *HLXB9* in these cancers. These results are summarised in Table 5.1.

Interestingly, both upregulation and downregulation of *HLXB9* was found in kidney cancer implying that accurate regulation of *HLXB9* expression in the kidney is important to prevent cancer development and progression. Previous studies on *HLXB9* expression by cancer type have also identified both overexpression and underexpression of *HLXB9* in patients (Tosi et al. 2003, von Bergh et al. 2006; Ballabio et al. 2009; Park et al. 2009, Ferguson et al. 2011). In leukaemia, overexpression of *HLXB9* was found in a subset of AML patients whilst underexpression was identified in ALL. Additionally, in pancreatic cancer, overexpression of *HLXB9* has been reported in insulinomas, whilst underexpression has been found in pancreatic adenocarcinomas (Shi et al. 2013; Desai et al. 2014, Bailey et al. 2016). These findings suggest *HLXB9* might function differently in different cell types originating from a common multipotent progenitor. This difference in expression levels in subtypes within cancer types might mask

some of the results obtained from this study. For instance, from Oncomine, it could be incorrectly deduced that *HLXB9* was underexpressed in leukaemia, which is most likely an artefact of the small population of AML patients that overexpress *HLXB9*. Therefore an attempt was made to analyse cancers by subtypes where possible.

Other genes that have been observed to show overexpression in some cancers and underexpression in other cancers include *HOXA9*, which has been found to be upregulated in colorectal carcinomas but downregulated in breast cancer and *HOXD8* which is downregulated in colorectal cancer whilst being overexpressed in lung cancer (Bhatlekar et al. 2014).

Cancer Type	Number of studies where *HLXB9* expression was altered by Cancer type ($P \leq 0.001$, fold change ≥ 2 and gene rank of top 10%)
Brain and Central Nervous System Cancer	1 out of 13 datasets showed *HLXB9* overexpression
Breast Cancer	5 out of 11 datasets showed *HLXB9* overexpression
Cervical Cancer	1 out of 6 datasets showed *HLXB9* overexpression
Oesophageal Cancer	3 out of 7 datasets showed *HLXB9* overexpression
Kidney Cancer	1 out of 7 datasets showed *HLXB9* overexpression 2 out of 7 datasets showed downregulation of *HLXB9*
Leukaemia	2 out of 11 datasets showed downregulation of *HLXB9*
Lung Cancer	3 out of 12 datasets showed *HLXB9* overexpression
Lymphoma	1 out of 8 datasets showed *HLXB9* overexpression
Melanoma	1 out of 4 datasets showed *HLXB9* overexpression
Ovarian Cancer	1 out of 8 datasets showed *HLXB9* overexpression
Pancreatic Cancer	1 out of 9 datasets showed downregulation of *HLXB9*

Table 5.1: Eleven cancer types where analysis identified populations of patients with aberrant expression of HLXB9. (Oncomine) . Datasets in Oncomine are obtained from published or unpublished microarray experiments deposited into the database. All datasets available per cancer type were analysed. Normal to cancer analyses were conducted on datasets available using the following parameters gene expression fold change of ≥2, $P \leq 0.001$ and gene rank of the top 10% of gene altered. Studies with increased fold change of ≥2 are defined as overexpression whilst studies with decreased fold change of ≥2 are defined as showing underexpression/downregulation.

The Oncomine database also has a cancer histology tool, which compares gene expression changes in different histological subtypes within a cancer type; eight cancers with altered *HLXB9* expression between cancer subtypes were identified using this tool. They were sarcoma, leukaemia, melanoma, breast cancer, kidney cancer, leukaemia, melanoma, ovarian cancer and sarcoma. Some of these cancer types: breast cancer, kidney cancer, leukaemia, and ovarian cancer were previously identified using the normal to cancer tool. This suggests in these particular cancer types, not only does differential expression of *HLXB9* occur in cancer states from normal; there are also subtypes that show significant differential expression. The results from these analyses are examined in more detail in table 5.7.

This database also allows users to carry out another analysis known as 'outlier analysis'. This analysis can be used to identify genes with very low or very high expression in a subset of samples (sub population) within the total population for a cancer type. This analysis is crucial as differential expression of genes might occur in a small subset of a cancer type, which could go undetected using a standard normal to cancer analysis. The cancer types with outlier expression of *HLXB9* were brain and central nervous system cancer, adrenal carcinoma, kidney cancer, leukaemia, lymphoma, lung cancer, sarcoma, liver cancer, breast cancer, ovarian cancer, prostate cancer. The results from some of these analyses are categorised by cancer type and examined in detail in table 5.7.

In summary, *HLXB9* expression was altered in a minority of cancers types in the Oncomine database, with most of the cancer types showing upregulation of *HLXB9*. Combining the analyses from the three freely available tools in Oncomine, differential expression of *HLXB9* might be particularly significant in breast cancer, kidney cancer, leukaemia and ovarian cancer.

b. Gene Expression across Normal and Tumour Tissue- GENT

The GENT database has samples from two different microarray platforms so the results were split into two to represent each platform; please see figure 5.1 and 5.3 for samples on the U133A platform and figure 5.2 and 5.4 for samples on the U133Plus 2 platform. In this database, comparing boxplots of normal and cancer samples can be used to identify outliers of gene expression in various cancers. Raw expression values are plotted as boxplots for normal and cancer samples in figure 5.1 and 5.2.

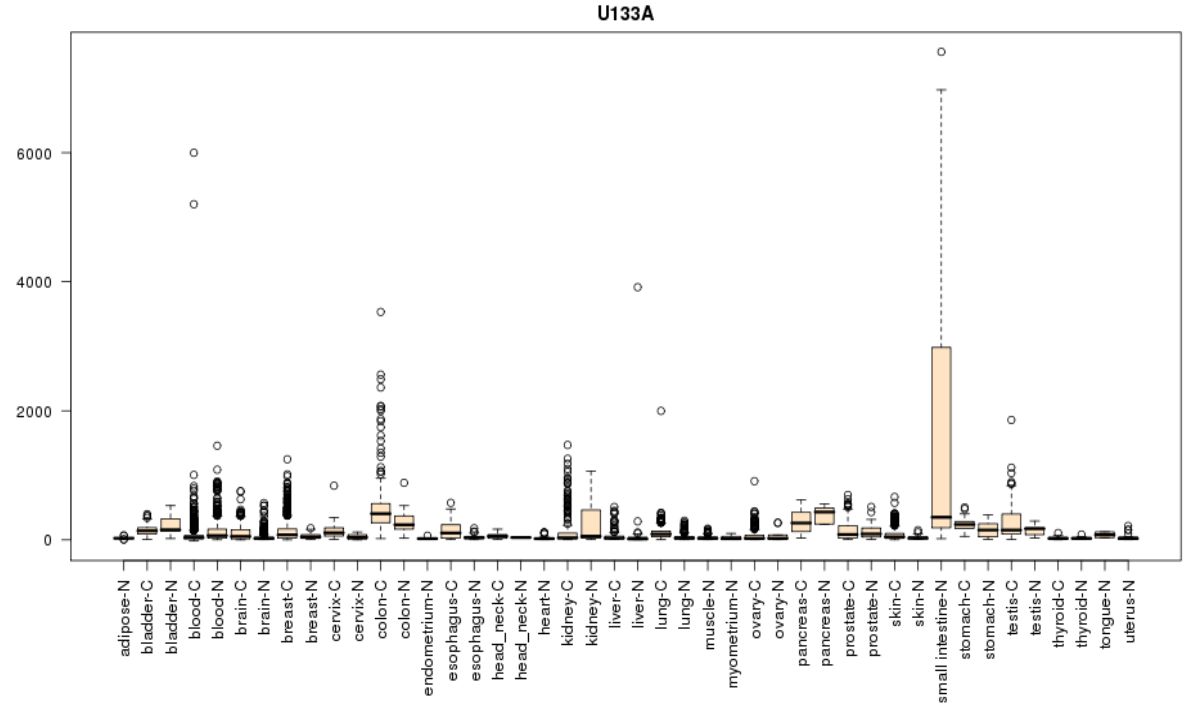

Figure 5.1: Boxplots showing differential expression of *HLXB9* in cancer originating from various tissues. These boxplots are raw values of *HLXB9* using the datasets under the U133A platform from the GENT database. The plot shows nineteen different normal and cancer matched samples.

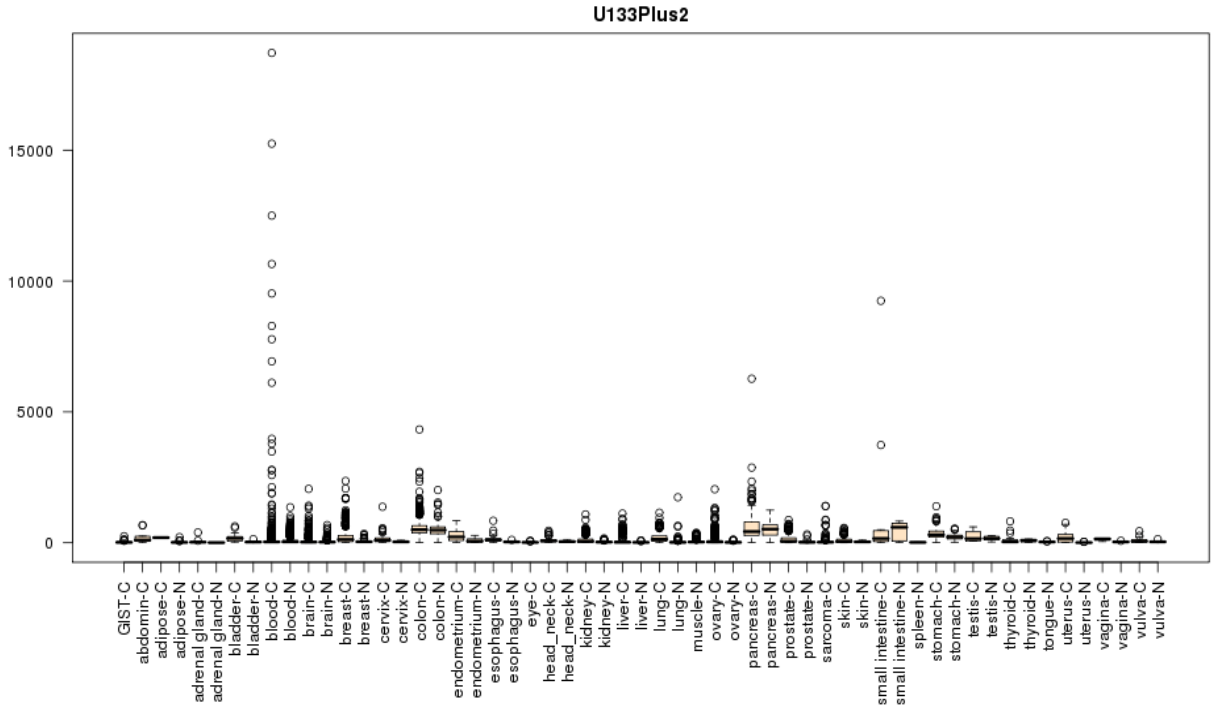

Figure 5.2: Boxplots showing differential expression of *HLXB9* in cancer originating from various tissues. These boxplots are raw values of *HLXB9* using the datasets under the U133 Plus2 platform from the GENT database. The plot shows twenty-five different normal and cancer matched samples.

Although comparing box plots of the raw values provides a suggestion of cancers with differential expression of *HLXB9*, comparing boxplots of the $\log_2$ transformed value between normal and cancer

samples is more meaningful as transformation of the raw values normalises the variation between datasets enabling an accurate judgment of the differential expression between normal and cancer samples (figures 5.3 and 5.4; tables 5.3 and 5.4).

It should be noted that some of the cancers originating from tissues such as adrenal gland and vaginal in this database had small sample size; therefore results obtained on this cancer type on *HLXB9* expression may not be truly representative of the cancer type. For details on sample size of cancers in this database please see table 5.2.

Origin	**U133Plus2**		**U133A**	
	Cancer	**Normal**	**Cancer**	**Normal**
Adipose	1	59	0	12
Adrenal gland	14	5	0	0
Bladder	39	14	87	15
Blood	7786	847	3130	1099
Brain	838	667	592	1627
Breast	2662	267	2635	91
Cervix	113	12	64	34
Colon	1994	287	256	27
Endometrium	72	75	0	9
Oesophagus	48	9	24	28
Eye	63	0	0	0
Gastrointestinal stromal tumours	66	0	0	0
Head and Neck	202	14	21	2
Kidney	573	130	366	66
Liver	194	50	156	52
Lung	547	336	582	364
Muscle	0	217	0	331
Ovary	902	51	341	9
Pancreas	174	62	13	8
Prostate	314	51	244	83
Sarcoma	520	0	0	0
Skin	302	141	499	59
Small intestine	13	6	0	22
Spleen	0	6	0	0
Stomach	311	57	46	18
Testis	4	9	184	13
Thyroid	62	25	44	25
Tongue	0	11	0	4
Uterus	155	12	0	24
Vagina	3	5	0	0
Vulva	21	14	0	0
Heart	0	0	0	41
Myometrium	0	0	0	24
Total	**18,004**	**3,439**	**9,284**	**4,087**

Table 5.2: Number of samples in the GENT database across the U133A and U133 Plus 2 platforms

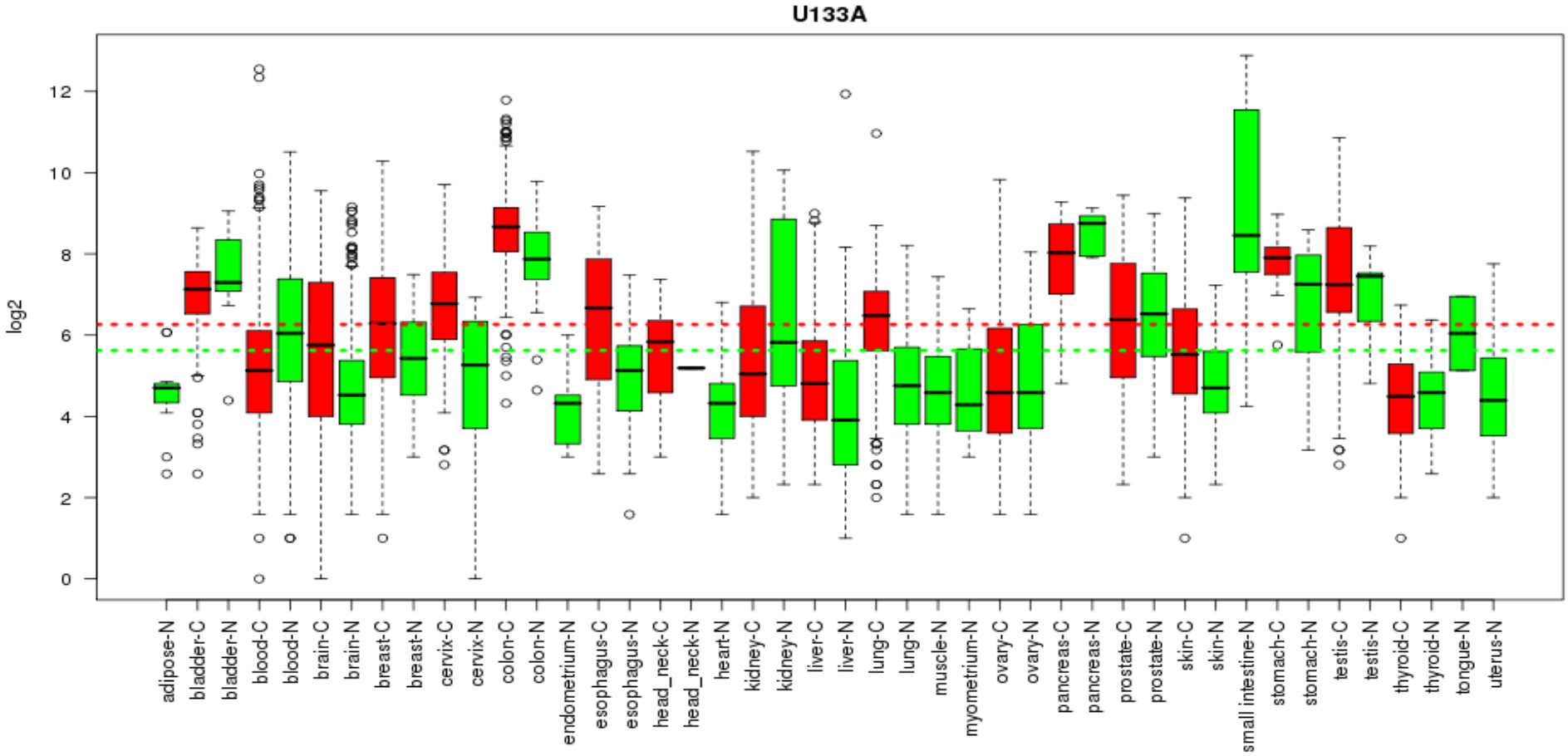

Figure 5.3: Boxplots showing differential expression of *HLXB9* in cancer originating from various tissues (U133A microarray platform). Box plot are $\log_2$ transformed raw values for the same nineteen normal cancer samples in figure 5.1. Please see tables 5.3a and b for detailed analysis of box plots.

Origin	Controls				Cancer samples			
	n	Lower Quartile	Median	Upper Quartile	n	Lower Quartile	Median	Upper Quartile
		log2 values				log2 values		
Breast	271	4.09	5.21	6.13	2658	5.32	6.95	8.11
Bladder	15	7.01	7.29	8.45	87	6.51	7.13	7.56
Brain	1627	3.81	4.52	5.39	592	4	5.75	7.29
Blood	1099	4.81	5.93	7.38	3165	4.09	5.13	6.13
Cervix	34	3.7	5.27	6.37	64	5.88	6.78	7.57
Colon	27	7.29	7.87	8.62	256	8.06	8.67	9.14
Oesophagus	28	3.97	5.13	5.75	24	4.88	6.67	7.9
Head and Neck	2	5.17	5.19	NA	21	4.45	5.83	6.45
Kidney	66	4.75	5.82	8.86	366	4	5.04	6.72
Liver	52	2.81	3.91	5.42	165	3.86	4.81	5.87
Lung	364	3.81	4.75	5.7	582	5.61	6.49	7.08
Ovary	9	2.85	4.58	7.16	341	3.58	4.58	6.18
Pancreas	8	7.94	8.75	9.01	13	6.71	8.03	8.96
Prostate	83	5.46	6.52	7.56	244	4.95	6.39	7.77
Skin	59	4.09	4.7	5.64	499	4.52	5.52	6.64
Stomach	18	5.49	7.25	7.99	46	7.47	7.9	8.17
Testis	13	6.29	7.45	7.64	184	6.56	7.25	8.66
Thyroid	25	3.64	4.58	5.11	44	3.52	4.49	5.3

Table 5.3a: Box plots of HLXB9 expression in control and cancer samples from U133A microarray platform in the GENT database.

Origin	Control samples				Cancer samples				Fold change	log 2 fold change	*P* value
	n	Mean	SD	SEM	n	Mean	SD	SEM			
		log2 values				log2 values					
Breast	271	5.08	1.37	0.08	2658	6.62	1.9	0.04	2.19	1.13	<0.0001
Bladder	15	7.48	1.12	0.29	87	6.87	1.21	0.13	0.67	-0.58	0.07
Brain	1627	4.62	1.23	0.03	592	5.68	1.87	0.08	2.79	1.48	<0.0001
Blood	1099	5.98	1.69	0.05	3165	5.13	1.41	0.03	0.51	-0.98	<0.0001
Cervix	34	4.98	1.55	0.27	64	6.57	1.39	0.17	2.91	1.54	<0.0001
Colon	27	7.83	1.08	0.21	256	8.59	1.06	0.07	1.78	0.83	0.0005
Oesophagus	28	4.99	1.37	0.26	24	6.36	1.94	0.4	3.46	1.79	0.0045
Head and Neck	2	5.19	0.03	0.02	21	5.58	1.19	0.26	1.72	0.78	0.65
Kidney	66	6.48	2.12	0.26	366	5.43	1.95	0.1	0.54	-0.90	<0.0001
Liver	52	4.16	1.95	0.27	165	4.97	1.49	0.12	0.55	-0.86	0.002
Lung	364	4.8	1.35	0.07	582	6.31	1.19	0.05	2.49	1.32	<0.0001
Ovary	9	4.92	2.39	0.8	341	4.88	1.74	0.09	0.76	-0.39	0.95
Pancreas	8	8.54	0.52	0.19	13	7.74	1.37	0.38	0.75	-0.41	0.13
Prostate	83	6.34	1.42	0.16	244	6.25	1.75	0.11	1.13	0.18	0.67
Skin	59	4.74	1.09	0.14	499	5.65	1.45	0.06	2.37	1.24	<0.0001
Stomach	18	6.7	1.68	0.39	46	7.85	0.57	0.08	1.53	0.62	0.0001
Testis	13	7.07	0.96	0.27	184	7.36	1.58	0.12	1.70	0.77	0.51
Thyroid	25	4.48	0.97	0.19	44	4.35	1.24	0.19	1.01	0.02	0.65

Table 5.3b: Comparison of HLXB9 expression in control samples to cancer samples using the U133A microarray platform in the GENT database. *P* ≤0.05 was considered significant, results of independent student test. Fold change calculated as average expression levels of cancer samples divided by average probe expression level of control samples. Fold change of ≥2 in *HLXB9* expression between control and cancer samples are defined as overexpression. Fold change of ≤0.5 in *HLXB9* expression between control and cancer samples are defined as underexpression. Rows highlighted in red represent cancer types with statistically significant overexpression of *HLXB9* whilst rows highlighted in blue represent statistically significant underexpression of *HLXB9.*

From the fold change and t-test analysis, cancer originating from eight organs showed differential expression of *HLXB9*. Six of the eight showed statistically significant overexpression of *HLXB9*: these were breast, brain, cervix, oesophagus, lung and skin (fold change ≥2, *P*≤0.05). Two showed underexpression of *HLXB9*: blood, kidney (fold change of ≤.0.5, *P*≤0.05).

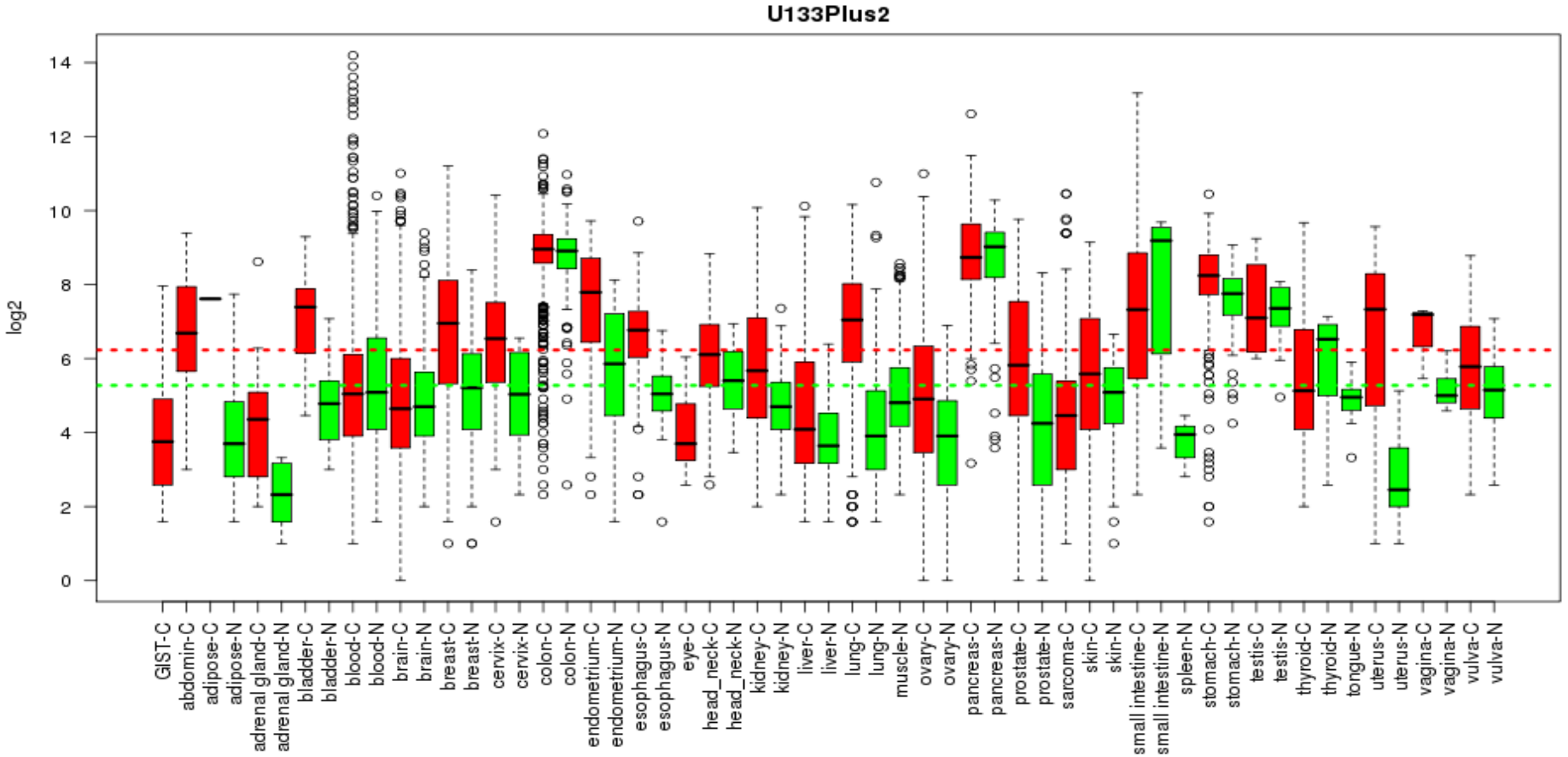

Figure 5.4: Boxplots showing differential expression of *HLXB9* in cancer originating from various tissues (U133 Plus 2 microarray platform). Box plot are $\log_2$ transformed expression raw values for the same twenty-five normal cancer samples as in figure 5.2. See tables 5.4a and b for more detailed analysis of box plots.

Origin	Controls				Cancer samples			
	n	Lower quartile	Median	Upper quartile	n	Lower quartile	Median	Upper quartile
		log2 values				log2 values		
Adrenal gland	5	1.29	2.32	3.25	14	2.81	4.36	5.25
Bladder	14	3.75	4.78	5.4	39	6.13	7.39	7.91
Blood	855	4.09	5.09	6.55	7836	3.91	5.04	6.11
Brain	668	3.91	4.7	5.64	874	3.58	4.64	6.00
Breast	91	4.52	5.43	6.32	2635	4.95	6.29	7.41
Cervix	12	3.82	5.04	6.23	113	5.36	6.54	7.55
Colon	293	8.43	8.91	9.24	1839	8.58	8.96	9.35
Endometrium	75	4.39	5.86	7.22	72	6.43	7.79	8.72
Oesophagus	9	4.2	5.04	5.72	48	6	6.77	7.29
Head and Neck	14	4.48	5.4	6.2	202	5.24	6.11	6.92
Kidney	130	4.09	4.7	5.36	573	4.39	5.67	7.1
Liver	50	3.17	3.64	4.54	194	3.17	4.09	5.91
Lung	381	3	3.91	5.13	547	5.91	7.04	8.04
Ovary	51	2.59	3.91	4.86	902	3.46	4.91	6.34
Pancreas	62	8.19	9.02	9.42	174	8.13	8.73	9.64
Prostate	51	2.59	4.25	5.61	314	4.44	5.82	7.54
Skin	141	4.17	5.09	5.75	302	4.09	5.59	7.08
Small Intestine	6	5.49	9.19	9.58	13	5.16	7.32	8.9
Stomach	57	7.17	7.75	8.24	311	7.73	8.24	8.81
Testis	9	6.41	7.36	7.92	4	6.09	7.1	8.89
Thyroid	25	4.76	6.52	6.93	62	4.07	5.13	6.78
Uterus	12	2	2.45	3.88	155	4.64	7.33	8.3
Vagina	5	4.7	5	5.83	3	5.46	7.19	NA
Vulva	14	4.19	5.15	5.78	21	4.45	5.78	7.13

Table 5.4a: Box plots of HLXB9 expression in normal samples to cancer samples from the U133Plus 2 microarray platform in the GENT database.

From the U133Plus 2 microarray platform, overexpression of *HLXB9* was found in cancers originating from the adrenal gland, bladder, breast, cervix, endometrium, oesophagus, kidney, liver, lung, ovary, prostate, skin, uterus and vagina (fold change ≥2, $P \leq 0.05$, as shown in table 5.4b).

The results obtained from both microarray platforms on GENT were compared. Five cancer types from both platforms showed statistically significant overexpression of *HLXB9*; these were breast, cervix, oesophagus, lung and skin. Additionally underexpression of *HLXB9* was found in leukaemia in both platforms (U133A, fold change 0.51. U133Plus2, fold change 0.58), although only the result obtained from the U133A platform met the parameters of this study (underexpression defined as fold change ≤ 0.5). Conflicting results were obtained on expression of *HLXB9* in two cancer types, kidney cancer and prostate cancer. Underexpression of *HLXB9* was obtained in the U133A platform whilst overexpression was observed in the U133Plus2 platform. This result was similar to the results obtained from the Oncomine database and will be explained in the discussion section. For prostate cancer, the result from the U133A platform the fold change was insignificant (n=244), whilst the result from the U133Plus2 platform was significant (n=314) which may be due to an increase in sample size.

A number of the cancer types had samples in only the U133Plus2 platform such as adrenal gland, endometrium, uterus and vagina, therefore could not be compared across both platform.

Tissue/Organ	Control samples				Cancer samples				Fold change	log2 fold change	*P* value
	n	Mean	SD	SEM	n	Mean	SD	SEM			
		log2 values				log2 values					
Adrenal gland	5	2.28	0.45	1.00	14	4.42	0.47	1.77	8.82	3.14	0.022
Bladder	14	4.7	1.07	0.29	39	7.03	1.23	0.2	5.20	2.38	<0.0001
Blood	855	5.43	1.91	0.07	7836	5.06	1.53	0.02	0.67	-0.58	<0.0001
Brain	668	4.82	1.29	0.05	874	4.87	1.86	0.06	1.72	0.78	0.55
Breast	91	5.42	1.07	0.11	2635	6.13	1.64	0.03	3.81	1.93	<0.0001
Cervix	12	4.91	1.42	0.41	113	6.34	1.57	0.15	3.14	1.65	0.003
Colon	293	8.77	0.83	0.05	1839	8.88	0.87	0.02	1.09	0.12	0.043
Endometrium	75	5.65	1.6	0.18	72	7.31	1.77	0.21	3.17	1.67	<0.0001
Oesophagus	9	4.81	1.47	0.49	48	6.46	1.52	0.22	3.39	1.76	0.004
Head and Neck	14	5.25	1.09	0.29	202	5.97	1.24	0.09	1.79	0.84	0.035
Kidney	130	4.73	0.98	0.09	573	5.74	1.7	0.07	3.04	1.60	<0.0001
Liver	50	3.8	1.04	0.15	194	4.76	2.21	0.16	5.71	2.51	0.003
Lung	381	4.1	1.46	0.07	547	6.79	1.65	0.07	5.29	2.40	<0.0001
Ovary	51	3.7	1.68	0.24	902	5.07	2.02	0.07	4.00	2.00	<0.0001
Pancreas	62	8.52	1.52	0.19	174	8.8	1.25	0.1	1.26	0.33	0.15
Prostate	51	3.88	2.12	0.3	314	5.82	2.1	0.12	3.54	1.83	<0.0001
Skin	141	4.86	1.18	0.1	302	5.44	1.88	0.11	2.33	1.22	0.0008
Small Intestine	6	7.89	2.49	1.02	13	7.2	3.2	0.89	2.43	1.28	0.65
Stomach	57	7.55	1.07	0.14	311	8.07	1.27	0.07	1.47	0.56	0.004
Testis	9	7.13	1.06	0.35	4	7.36	1.49	0.74	1.45	0.54	0.75
Thyroid	25	5.87	1.36	0.27	62	5.29	1.65	0.21	0.98	-0.04	0.12
Uterus	12	2.8	1.32	0.38	155	6.46	2.25	0.18	18.13	4.18	<0.0001
Vagina	5	5.21	0.64	0.29	3	6.65	1.03	0.6	2.86	1.52	0.047
Vulva	14	4.95	1.2	0.32	21	5.79	1.67	0.37	2.35	1.23	0.11

Table 5.4b: Comparison of HLXB9 expression box plots in normal samples to Cancer samples using the U133Plus 2 microarray platform in the GENT database. *P* ≤0.05 was considered significant, results of independent student test. Fold change calculated as average expression levels of cancer samples divided by average probe expression level of control samples. Fold change of ≥2 in *HLXB9* expression between control and cancer samples are defined as overexpression. Fold change of ≤0.5 in *HLXB9* expression between control and cancer samples are defined as underexpression. Rows highlighted in pink represent cancer types with statistically significant overexpression of *HLXB9* whilst rows highlighted in blue represent statistically significant underexpression of *HLXB9*.

c. FIREBROWSE

Cancer types identified as having overexpression of *HLXB9* (fold change ≥ 2) from RNA-Seq experiments in this database were rectum, stomach, kidney, head and neck, lung, adrenal, breast, prostate, bladder, endometrial. Underexpression of *HLXB9* (fold change of ≤ 0.5) was found in pancreatic and kidney cancer (as shown in figure 5.5 and table 5.5).

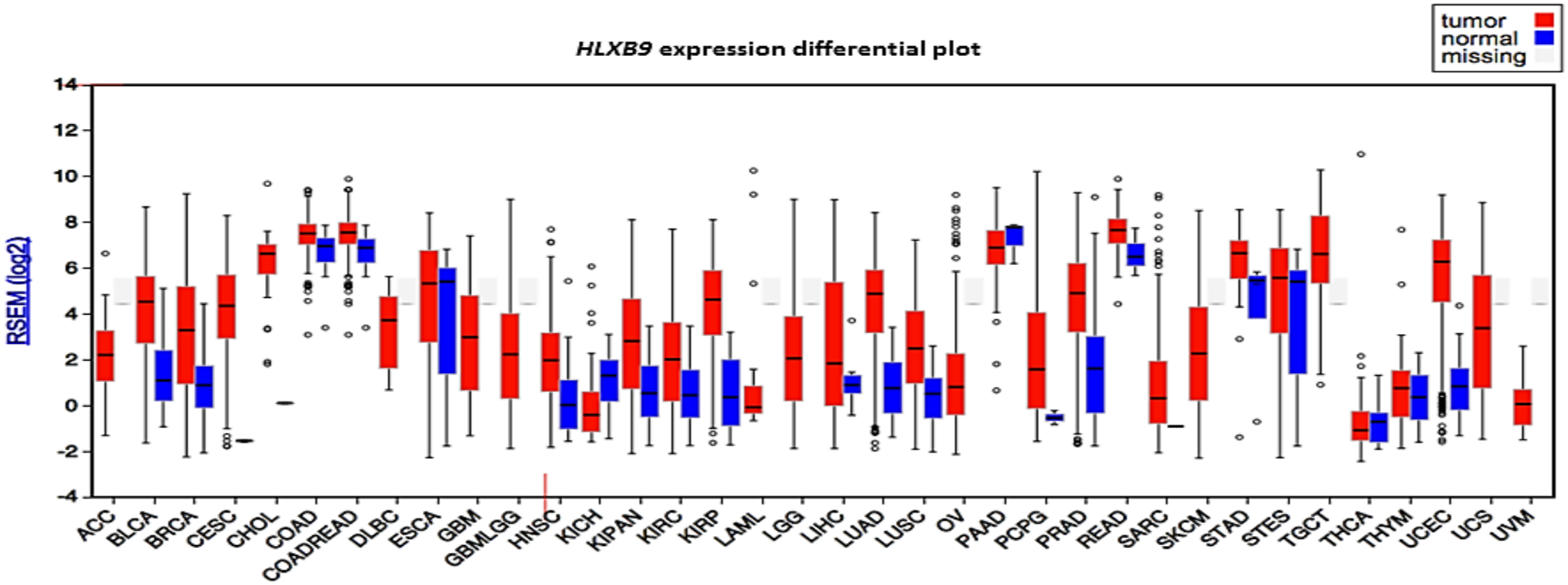

Figure 5.5: Boxplots showing differential expression of *HLXB9* (MNX1) in normal and cancer samples (FIREBROWSE). Overexpression of *HLXB9* was found in stomach adenocarcinoma (STAD), rectum adenocarcinoma (READ), kidney cancer cohort (KIPAN), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), head and neck squamous cell carcinoma (HNSC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), pheochromocytoma and paraganglioma (PCPG), breast invasive carcinoma (BRCA), prostate adenocarcinoma (PRAD), uterine corpus endometrial carcinoma (UCEC) and bladder urothelial carcinoma (BLCA). Both pancreatic adenocarcinoma (PAAD) and kidney chromophobe (KICH), a kidney cancer subtype showed underexpression of *HLXB9*. Other cancers analysed included colon adenocarcinoma (COAD), colorectal adenocarcinoma (COADREAD), oesophageal carcinoma (ESCA), liver hepatocellular carcinoma (LIHC), stomach and esophageal carcinoma (STES), thyroid carcinoma (THCA) and thymoma (THYM), see table 5.5 for fold change for these cancer types.

Control Samples	(n)	Lower Quartile	Median	Upper Quartile	Cancer Type	n	Lower Quartile	Median	Upper Quartile	Fold change
		RNA-Seq by Expectation Maximization RSEM (log_2)					RNA-Seq by Expectation Maximization RSEM (log_2)			
Normal sample (BLCA)	12	0.21	1.1	2.42	Bladder urothelial carcinoma (BLCA)	384	2.72	4.53	5.64	10.8
Normal sample (BRCA)	94	-0.1	0.88	1.74	Breast invasive carcinoma (BRCA)	992	0.94	3.29	5.2	5.3
Normal sample (COAD)	41	6.25	6.97	7.32	Colon adenocarcinoma (COAD)	459	7.05	7.52	7.93	1.5
Normal sample (COADREAD)	51	6.24	6.89	7.27	Colorectal adenocarcinoma (COADREAD)	626	7.06	7.56	7.99	1.6
Normal sample (ESCA)	11	1.38	5.42	6.02	Oesophageal carcinoma (ESCA)	180	2.75	5.33	6.78	0.9
Normal sample (HNSC)	28	-1.02	0.03	1.12	Head and neck squamous cell carcinoma (HNSC)	452	0.61	1.98	3.17	3.9
Normal sample (KICH)	23	0.17	1.31	1.99	Kidney chromophobe (KICH)	45	-1.15	-0.41	0.6	0.3
Normal sample (KIPAN)	98	-0.5	0.54	1.74	All kidney cohort (KIPAN)	775	0.73	2.81	4.65	4.8
Normal sample (KIRC)	52	-0.54	0.45	1.55	Kidney renal clear cell carcinoma (KIRC)	442	0.18	2.01	3.62	3
Normal sample	23	-0.89	0.36	2.01	Kidney renal papillary cell carcinoma (KIRP)	288	3.08	4.63	5.9	19.3
Normal sample (LIHC)	11	0.53	0.9	1.31	Liver hepatocellular carcinoma (LIHC)	170	-0.02	1.84	5.39	1.9
Normal sample (LUAD)	49	-0.34	0.75	1.88	Lung adenocarcinoma (LUAD)	504	3.17	4.88	5.92	17.4
Normal sample (LUSC)	42	-0.56	0.51	1.2	Lung squamous cell carcinoma (LUSC)	452	0.97	2.5	4.13	4
Normal sample (PAAD)	3	6.99	7.79	7.84	Pancreatic adenocarcinoma (PAAD)	179	6.16	6.9	7.65	0.5
Normal sample (PCPG)	2	-0.69	-0.54	-0.38	Pheochromocytoma and paraganglioma (PCPG)	149	-0.14	1.58	4.07	4.3
Normal sample (PRAD)	41	-0.34	1.61	3.02	Prostate adenocarcinoma (PRAD)	492	3.21	4.91	6.22	9.9
Normal sample (READ)	10	6.1	6.5	7.08	Rectum adenocarcinoma (READ)	167	7.07	7.66	8.15	2.2
Normal sample (STAD)	4	3.8	5.46	5.67	Stomach adenocarcinoma (STAD)	37	5.54	6.66	7.21	2.3
Normal sample (STES)	15	1.38	5.42	5.92	Stomach and esophageal carcinoma (STES)	217	3.15	5.58	6.87	1.1
Normal sample (THCA)	23	-1.61	-0.71	-0.31	Thyroid carcinoma (THCA)	189	-1.53	-1.08	-0.25	0.8
Normal sample (THYM)	2	-0.62	0.36	1.33	Thymoma (THYM)	107	-0.5	0.76	1.54	1.3
Normal sample (UCEC)	21	-0.19	0.84	1.62	Uterine corpus endometrial carcinoma (UCEC)	508	4.53	6.29	7.24	43.9

Table 5.5: Comparison of *HLXB9* expression box plots in normal samples to cancer samples (FIREBROWSE). This table shows cancer type, sample size (n), lower quartile, median, upper quartile expression values and fold change in expression between normal and cancer values. Row highlighted in pink represent cancer types with differential overexpression of *HLXB9* whilst rows highlighted in blue represent cancer types with differential underexpression of *HLXB9*. Overexpression of *HLXB9* was found in stomach adenocarcinoma (STAD), rectum adenocarcinoma (READ), all kidney cancer cohort (KIPAN), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), head and neck squamous cell carcinoma (HNSC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), pheochromocytoma and paraganglioma (PCPG), breast invasive carcinoma (BRCA), prostate adenocarcinoma (PRAD), uterine corpus endometrial carcinoma (UCEC), bladder urothelial carcinoma (BLCA). Pancreatic adenocarcinoma (PAAD) and kidney chromophobe (KICH) a type of kidney cancer showed underexpression of *HLXB9*.

As seen in the previous databases, most of the cancers identified as having differential expression of *HLXB9* show overexpression of the gene suggesting a potential pro-oncogenic role for *HLXB9* in most cancers identified. Additionally, some of the cancers showed very substantial increase in the expression of *HLXB9* in the cancerous state such as prostate adenocarcinoma (PRAD, fold change=9.86), bladder urothelial carcinoma (BLCA, fold change=10.8), lung adenocarcinoma (LUAD, fold change=17.4), kidney renal papillary cell carcinoma (KIRP, fold change=19.3) and uterine corpus endometrial carcinoma (UCEC, fold change=43.9).

One of the interesting findings from this database was the identification of different patterns of expression of *HLXB9* within a cancer type such as in kidney cancer (also seen in Oncomine and GENT). In kidney cancer, two out of three subtypes; kidney renal clear cell carcinoma (KIRC) and kidney renal papillary cell carcinoma (KIRP) showed *HLXB9* overexpression suggesting a possible pro-oncogenic function of *HLXB9* for these kidney cancer subtypes. However, the third subtype kidney chromophobe (KICH) showed underexpression of *HLXB9* suggesting a tumour suppressive function of *HLXB9* in this kidney cancer subtype. This finding illustrates the importance of studying cancers by subtype when differential expressions of genes are being considered between normal and cancerous states.

d. BioXpress v 1.0.

In this database**,** *HLXB9* expression in a cancer sample is compared to its non-cancer, matched pair sample. The change in expression between the two states was classified into overexpression or under expression. Overexpression was determined to be a fold change of greater than zero compared to its normal pair expression level and underexpression is a fold change less than zero compared to its normal pair; the results of this calculation are represented as a graph with a percentage frequency of over or under expression in observed in patients samples.

The database then compares the expression of *HLXB9* in the cancer sample to the non-cancer sample and calculates if the change observed between the two states was statistically significant. The frequency of statistically significant expression by cancer type was then represented graphically (figure 5.6). Six out of the eighteen cancer types analysed showed a statistical significant difference in the expression of

HLXB9. *HLXB9* was found to be significantly upregulated in 10% of cases of prostate cancer (PRAD), 6.67% of kidney renal papillary cell carcinoma (KIRP), 2% of cases of liver cancer (LIHC), 1.75% of cases of lung cancer (LUAD) and 0.88% of cases of breast cancer (BRCA). *HLXB9* was also found to be significantly downregulated in 33% of pancreatic patient samples (PAAD).

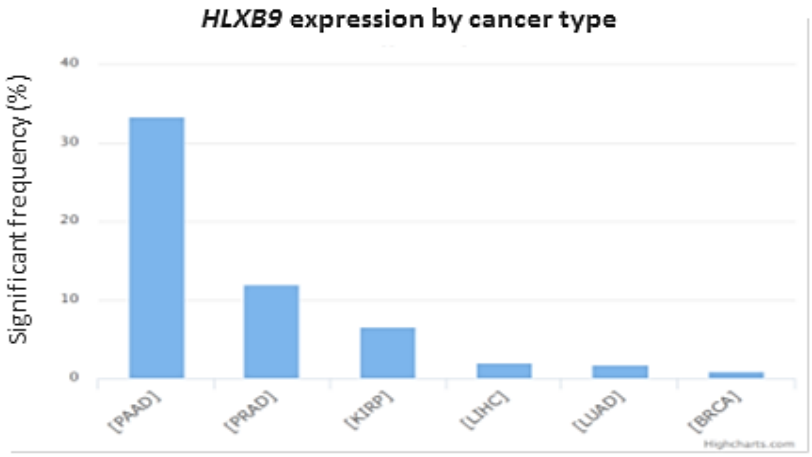

Figure 5.6: The frequency of significant differential expression of *HLXB9* in cancer types (adjusted p value≤ 0.05). *HLXB9* was found to be significantly upregulated in 10% of cases of prostate cancer (PRAD), 6.67% of kidney renal papillary cell carcinoma cases (KIRP), 2% of cases of liver cancer (LIHC), 1.75% of cases of lung cancer (LUAD), and 0.88% of cases of breast cancer (BRCA). *HLXB9* was found to be significantly downregulated in 33% of pancreatic patient samples (PAAD).

Although only six out of eighteen cancer types analysed showed a statistical significant difference in the expression of *HLXB9*, it is worth noting that some of the other cancer types analysed may not have reached statistical significance due to the small number of samples analysed, as shown in table 5.6. Examples include cervical cancer and rectum cancer where the sample size was less than ten. However, there was a significant difference observed in pancreatic cancer although only four samples were analysed. This indicates that downregulation of *HLXB9* might occur more frequently in pancreatic cancer which is not surprising as the pancreas is an organ with endogenous *HLXB9* expression.

Cancer Type	No of patient (normal and cancer pairs)	Overexpression Frequency (%)	Underexpression Frequency (%)	Frequency (significance %)
DOID:5041_Esophageal cancer [ESCA]	10	77.78	22.22	0
DOID:4362_Cervical Cancer [CESC]	4	100	0	0
DOID:363_Uterine cancer [UCEC]	18	82.35	17.65	0
DOID:3571_Liver Cancer [LIHC]	51	34	66	2
DOID:263_Kidney cancer [KIRP]	128	96.67	3.33	6.67
DOID:263_Kidney cancer [KIRC]	128	72.22	27.78	0
DOID:263_Kidney cancer [KICH]	128	16	84	0
DOID:219_Colon Cancer [COAD]	27	88.46	11.54	0
DOID:1993_Rectum cancer [READ]	6	83.33	16.67	0
DOID:1793_Pancreatic Cancer [PAAD]	4	33.33	66.67	33.33
DOID:1781_Thyroid cancer [THCA]	60	32.2	67.8	0
DOID:1612_Breast Cancer [BRCA]	114	69.03	30.97	0.88
DOID:1324_Lung Cancer [LUSC]	108	68	32	0
DOID:1324_Lung Cancer [LUAD]	108	94.74	5.26	1.75
DOID:11934_Head and neck cancer [HNSC]	41	78.05	21.95	0
DOID:11054_Urinary bladder cancer [BLCA]	20	89.47	10.53	0
DOID:10534_Stomach cancer [STAD]	31	76.67	23.33	0
DOID:10283_Prostate Cancer [PRAD]	51	78	22	10

Table 5.6: Cancer types, frequency of overexpression or underexpression and the frequency of statistical significance differential expression of *HLXB9* in cancer samples.

II. *HLXB9* expression by cancer types

Six cancer types were identified as having differential expression of *HLXB9* from all four databases analysed. These were breast cancer, kidney cancer, liver cancer, lung cancer, prostate cancer and endometrial cancer. More details on the pattern of expression of *HLXB9* in these cancer types are presented in table 5.7 and figure 5.7.

HLXB9 expression: Overexpression (Up) or Underexpression (Down)							
Cancer Type	Oncomine (% of analysis that met parameters)			GENT (Fold change)		FIREBROWSE (Fold change)	BioXpress Frequency of expression (%)
	Normal to Cancer Analysis	Cancer Histology Analysis	Outlier Analysis	U133A	U133 Plus2		
Brain and central nervous cancer	Up (3%)		Up (8%) Down (11%)				
Head and neck cancer				Up (1.7)	Up (1.7)	Up (3.9)	Up
Adrenal gland/cortex carcinoma			Up (3%) Down (3%)		Up (8.8)		
Pheochromocytoma and paraganglioma						Up (4.3)	
Thymoma						Up (1.3)	
Kidney cancer (all types)	Up (5%) Down (10%)	Up (19%) Down (11%)	Up (20%)	Down (0.5)	Up (3.0)	Up (4.8)	
Kidney chromophobe (KICH)		Up (3%) Down (3%)				Down (0.3)	Down
Kidney renal clear cell carcinoma (KIRC)	Down (5%)	Down (3%)	Up (7%)			Up (3.0)	Up
Kidney renal papillary cell carcinoma (KIRP)	Up (5%) Down (5%)	Up (5%)				Up (19.3)	Up (6.7%)
Leukaemia	Down (7%)	Up (1%) Down (1%)	Up (2%) Down (2%)	Down (0.5)	Down (0.6)		
Lymphoma	Up (3%)		Up (4%)				
Lung Cancer	Up (11%)		Up (4%)	Up (2.5)	Up (5.3)		
Lung adenocarcinoma (LUAD)	Up (11%)					Up (17.4)	Up (1.8%)
Lung squamous cell carcinoma (LUSC)						Up (4.0)	Up
Melanoma	Up (20%)	Up (33%)		Up (2.4)	Up (2.3)		
Sarcoma			Up (4%) Down (4%)				
Bladder Cancer				Down (0.7)	Down (5.2)	Up (10.8)	Up
Pancreatic cancer	Down (8%)			Down (0.7)	Up (1.3)	(0.54)	Down (33%)
Liver cancer			Up (7%)	Down (0.6)	Up (5.7)	Up (1.9)	Down (2%)
Oesophageal cancer	Up (27%)			Up (3.5)	Up (3.4)	(0.9)	Up
Stomach cancer				Up (1.5)	Up (1.5)	Up (2.3)	Up
Stomach and oesophageal carcinoma (STES)						Up (1.1)	
Small intestine					Up (2.4)		
Colon cancer				Up (1.8)	Up (1.1)	Up (1.5)	Up
Colorectal adenocarcinoma						Up (1.6)	
Rectum adenocarcinoma						Up (2.2)	Up
Breast cancer	Up (11%)	Up (4%) Down (4%)	Up (14%)	Up (2.2)	Up (3.8)	Up (5.3)	Up (0.9%)
Cervical cancer	Up (10%)	Up (8%)		Up (2.9)	Up (3.1)		Up
Endometrium cancer/ uterine corpus endometrial carcinoma		Up (3%) Down (3%)			Up (3.2)	Up (43.9)	Up
Ovarian cancer	Up (7%)	Up (9%) Down (2%)	Up (11%) Down (11%)	Up (0.8)	Up (4.0)		
Prostate cancer			Up (12%)	Up (1.1)	Up (3.5)	Up (9.7)	Up (11%)
Testicular cancer				Up (1.7)	Up (1.5)		
Vagina cancer					Up (2.9)		
Vulva cancer					Up (2.4)		

Table 5.7: Overview on HLXB9 expression in various cancers by databases. Up represents overexpression, down represents underexpression. Blank cells indicates lack of data for cancer type in database.

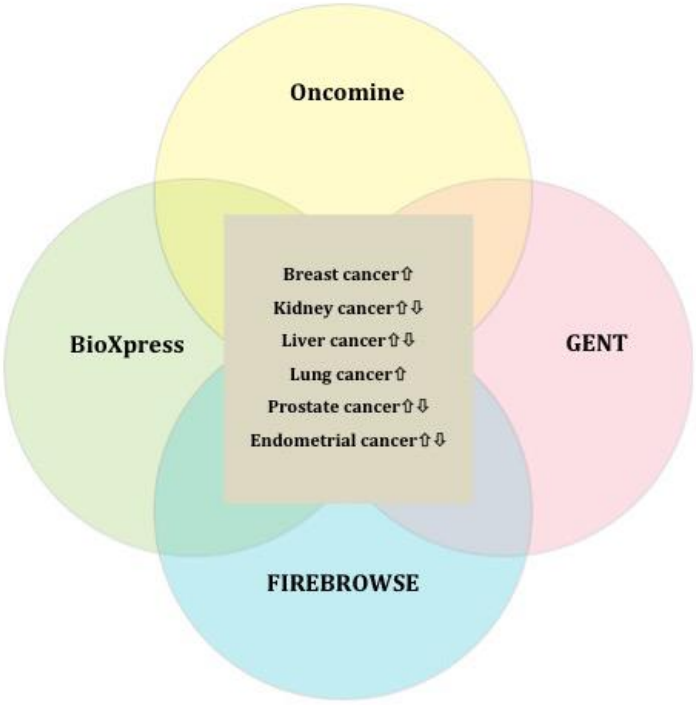

Figure 5.7: *HLXB9* expression in various cancer types. *HLXB9* was aberrantly expressed in six cancers types from all four databases; these were breast, kidney, liver, lung, endometrial and prostate cancer. Up arrows indicate up regulation of *HLXB9* whilst down arrows indicate downregulation of *HLXB9.*

Upregulation of *HLXB9* was identified in a proportion of breast cancer patients. This indicates the possibility of *HLXB9* overexpression being pro-oncogenic in some breast cancer patients. In two of the databases (Oncomine and FIREBROWSE), *HLXB9* overexpression was associated with invasive breast carcinoma suggesting a potential role for *HLXB9* in breast cancer metastasis. In the Oncomine database only, *HLXB9* overexpression was associated with invasive ductal breast carcinoma (IDC), when compared to other breast cancer subtypes in a study by Bitter 2005. This study also showed underexpression of *HLXB9* in mucinous breast carcinoma compared to the other breast cancer subtypes studied. *HLXB9* expression in breast cancer will be further investigated in Chapter 6.

Upregulation of *HLXB9* was also identified in a proportion of lung cancer patients. *HLXB9* overexpression was particularly significant in lung adenocarcinoma, a subtype of lung cancer (as shown in table 5.7). Although *HLXB9* overexpression was also found in lung squamous cell carcinoma patients, the change in expression was not as significant when compared to the change observed in lung

adenocarcinoma patients. This implies a pro-oncogenic role for *HLXB9* overexpression the lung adenocarcinoma subtype.

Upregulation of *HLXB9* in liver cancer was found in three out of the four databases investigated (Oncomine, GENT- U133Plus2 platform and FIREBROWSE) suggesting a pro-oncogenic role for *HLXB9*. Upregulation of *HLXB9* has previously been reported in advanced stage (III, IV) heptocellular carcinoma, the most common type of liver cancer (Wilkens et al. 2011). However, downregulation of *HLXB9* was found in BioXpress and GENT- U133A platform. This finding was contradictory to the aforementioned results and most likely indicates the heterogeneity of gene expression within subpopulations of a cancer type such as difference in tumour grade. Alternatively, the conflicting results obtained might indicate that *HLXB9* expression was not altered in liver cancer.

The expression of *HLXB9* in kidney cancer was interesting. Three of the databases (Oncomine, Gent and FIREBROWSE) analysed all types of kidney cancers together. From analysing all kidney cancer patients together, *HLXB9* was found to be mostly upregulated in kidney cancer. Three of the databases also analysed kidney cancer by subtypes (Oncomine, FIREBROWSE, and BioXpress). However, when the subtypes were investigated separately, a different pattern of expression emerged. Upregulation of *HLXB9* was mostly observed in KIRP and KIRC, and downregulation of *HLXB9* was mostly found in KICH (as shown in table 5.7).

Finally *HLXB9* was found to be upregulated in prostate and endometrial cancer, suggesting that *HLXB9* overexpression might be pro-oncogenic in these cancers.

5.5. Discussion

Experimental data from scientific literature suggested that differential expression of *HLXB9* would be identified in a large proportion of cancer types and patients. However, the results obtained using a number of transcriptomic databases were contrary to this expectation and imply that *HLXB9* is not differentially expressed in a majority of cancer types and patients.

It was also expected that by using this approach, patients previously characterised as overexpressing *HLXB9* such as infant AML patients harbouring translocation t(7;12)would be identified in this study. However, this was not the case and may be explained by the low incidence of these patients. Infant AML

patients harbouring t(7;12) represent approximately 4.8 cancer cases per million diagnosed (Ballabio et al. 2009; Tosi et al. 2015). Another tumour type previously described as showing *HLXB9* overexpression are insulinomas (Shi et al. 2013). Insulinoma is a rare type of pancreatic cancer, with an incidence of 1-4 cases per million cancer cases diagnosed (Okabayashi et al. 2013). Patients diagnosed with poorly differentiated grade III and IV hepatocellular carcinoma (HCC) patients have also been previously reported as showing *HLXB9* overexpression (Wilkens et al. 2011). However, apart from a review by Paradis who reported that most diagnosed HCC cases are grade II and III tumours, the incidence of grade III or IV tumours in HCC could not be determined (Paradis 2013).

There were limitations encountered in some of the database used. In Oncomine and for some cancer types in BioXpress, there were a small number of non-cancer and cancer samples, which means results obtained from this analysis may not be truly representative of the whole population. There was also an overlap in datasets used. For example, the TGCA datasets is freely available and is used in all databases analysed. This might introduce a bias in some of the results obtained, as there was a possibility of the same samples being interrogated repeatedly. Another limitation of this study was the lack of datasets on a number of cancer types in the databases analysed (as shown in table 5.7). Therefore, it is possible that differential expression of *HLXB9* important in the development and progression of cancers have not been identified in this study. In fact, aberrant expression of *HLXB9* previously reported in leukaemia (Ballabio et al. 2009; Wildenhain et al. 2010; Ferguson et al. 2011) was only found in two of the four databases analysed.

Despite these limitations, aberrant expression of *HLXB9* was identified in a proportion of patients diagnosed with six different cancer types. This suggests *HLXB9* may contribute to cancer development and progression from two different mechanisms. These are underexpression indicating a loss of tumour suppressor activity or overexpression indicating a pro-oncogenic role. This finding demonstrates that accurate regulation of this gene is crucial. In this study, differential *HLXB9* expression was found in cancer types previously reported in literature as showing aberrant expression of *HLXB9* these were: breast cancer and hepatocellular carcinoma.

HLXB9 was found upregulated in breast cancer suggesting a pro-oncogenic role for *HLXB9* in breast cancer. However, in contrast to this finding, earlier studies analysing *HLXB9* expression in breast cancer

had reported *HLXB9* as being downregulated suggesting loss of *HLXB9* reduces its tumour suppressor function in breast cancer (Tommasi et al. 2009; Lian et al. 2012; Siletz et al. 2013b; Siletz et al. 2013a). Our finding although contradictory to previous research may be attributed to a difference in the subtypes or subpopulations of breast tumours investigated. This is because results from both Oncomine and GENT suggest a subpopulation or subtypes of breast tumours might underexpress *HLXB9*.

HLXB9 was found upregulated in liver cancer (hepatocellular carcinoma) in three of the four database analysed. This result suggests a pro-oncogenic function of *HLXB9* in a subpopulation of hepatocellular carcinoma (HCC) patients. This finding was supported by previously published research by Wilkens and colleagues: in their study, *HLXB9* was found to be upregulated in a subset of HCC patients (grade III, IV) (Wilkens et al. 2011). In these patients, upregulation of *HLXB9* was associated with dedifferentiated HCC subtypes and poor patient prognosis (Wilkens et al. 2011).

This study also uncovered differential expression of *HLXB9* in cancer types not previously associated with aberrant expression of *HLXB9* such as lung cancer, kidney cancer and endometrial cancer. These findings highlight a new avenue of research in understanding these cancer types. The result obtained from kidney cancer analyses was particularly interesting.

There are three main subtypes of kidney cancer: kidney clear cell carcinoma (KIRC) representing 65% of diagnosed cases, kidney papillary cell carcinomas (KIRP) representing 20% of diagnosed cases and kidney chromophobe (KICH) representing 5% of diagnosed cases (Cairns 2011; Chen et al. 2016). Out of these subtypes, kidney chromophobe is a low-grade tumour subtype and has a favourable prognosis while KIRC and KIRP are mostly high-grade tumours, which have very poor prognosis. Additionally, most diagnosed metastatic kidney cancers are KIRC (Cairns et al 2011). *HLXB9* was upregulated in KIRP, a subtype of kidney cancer associated with poor prognosis. *HLXB9* expression can also be suggested to be upregulated in KICH; a subtype of kidney cancer with poor prognosis and associated with metastasis, although a subpopulation of patients diagnosed with this kidney cancer subtype also showed a significant downregulation of *HLXB9* (Oncomine only). This discrepancy in expression of *HLXB9* observed in KICH might be indicative of difference in tumour grade therefore more research is needed to fully understand *HLXB9* expression in this kidney subtype.

HLXB9 expression was found to be downregulated in KICH, a kidney cancer subtype with favourable outcome. Therefore, these results imply *HLXB9* overexpression is pro-oncogenic in KIRC and KIRP, two kidney cancer subtypes that are associated with poor prognosis, whilst underexpression in KICH is associated with a favourable outcome. This finding implies *HLXB9* expression in kidney cancer should be further explored to elucidate its function in the development and/or progression in this tumour type.

This study also highlighted the heterogeneity of cancers and the importance of studying tumours by subtypes during gene expression profiling experiments. For instance, *HLXB9* has been previously reported to be overexpressed in insulinomas, a subtype of pancreatic cancer, however in this study *HLXB9* was found underexpressed in pancreatic adenocarcinoma, the most common subtype of pancreatic cancer (Shi et al. 2013; Desai et al. 2014). Insulinomas, which are endocrine tumours, originate from b-cells from the Islets of Langerhans and are usually benign tumours. Whereas pancreatic adenocarcinomas are mostly exocrine tumours mostly originating from cells lining the pancreatic ducts are generally invasive and associated with a poor prognosis.

Additionally, the lack of categorisation of cancers types into subtypes as seen in the GENT database and for some cancer types in BioXpress and FIREBROWSE database was found to obscure the differential expression of *HLXB9* in some cancer types. For example, in kidney cancer by investigating only GENT or the all kidney cancer cohort in FIREBROWSE, an inaccurate conclusion could be reached that *HLXB9* was overexpressed in kidney cancer. However, when kidney cancer was sub-divided into subtypes like in BioXpress and FIREBROWSE, *HLXB9* was shown to be overexpressed in two subtypes (KIRP and KIRC) whilst underexpressed in one subtype (KICH).

Another finding from this study was that differential expression of *HLXB9* observed in cancers of organs/tissue solely or partially derived from the endodermal germ layer such as pancreas, rectum, lung, liver, bladder, stomach, and oesophagus. This indicates that *HLXB9* expression has to be correctly regulated in endodermal tissue derivatives after foetal development to maintain health (Langham 2010).

5.6. Conclusion

In conclusion, this study highlighted that a small proportion of cancer types and patients showed differential expression of *HLXB9*. However, there is a possibility that *HLXB9* expression may be implicated in cancer progression in some cancer types e.g. breast cancer, kidney cancer. Thereby, further research into characterising *HLXB9* expression in these specific cancer types and identifying *HLXB9* target genes and pathways maybe prove beneficial for improving patients' outcome.

Another avenue for further research would be to interrogate patient outcome in cases where *HLXB9* differential expression occurs, as this would provide insight into whether *HLXB9* expression could have prognostic significance in these patients. Although differential expression of *HLXB9* was identified in a number of cancer types, there remains a gap in knowledge of the function of *HLXB9* in these tissues and organs. Therefore it would be useful to understand the function of this gene in relation to the identified tissues and organs and how dysregulation may contribute to cancer development and progression.

6. Investigating *HLXB9* expression as a biomarker in breast cancer

6.1. Introduction

In the UK, breast cancer is the most commonly diagnosed cancer and the second leading cause of cancer deaths in women (Cancer Research UK). Genetically, breast cancer is a heterogeneous disease, with aberrations in expression of many genes. The identification of the increased expression of oestrogen (ER), progesterone (PR) and human epidermal growth factor 2 (HER2) receptors, proteins that mediate the increase in proliferation of breast cells observed in breast cancer has led to the categorisation of breast tumours into four main molecular subtypes: luminal A, luminal B, basal and HER2 amplified. These subtypes are now used routinely in making treatment decisions in clinic (Perou et al. 2000; Sorlie et al. 2003; Parker et al. 2009).

Apart from these molecular subtypes, other factors that have prognostic significance in breast cancer treatment include tumour grade, lymph node involvement and tumour response to treatment. However, there remains a need to identify new biomarkers to improve patients' outcome. For instance, biomarkers that can be used for early diagnosis, evaluate prognosis, monitor disease progression and provide targets for new therapy will be beneficial to patients.

6.2. Aim

The aim of this chapter is to investigate *HLXB9* as a possible biomarker in breast cancer by:

- Using quantitative RT-PCR and indirect immunofluorescence to investigate the expression of *HLXB9* in non-tumorigenic breast cells and breast cancer cell lines
- Using FISH and erosion analysis of FISH images to compare the nuclear localisation of *HLXB9* in non-tumorigenic and breast cancer cell lines to determine if there is a difference in the nuclear localisation of *HLXB9* in breast cancer
- Investigating the effect of the nuclear localisation of *HLXB9* in non-cancer and breast cancer cell lines on expression of *HLXB9*
- Examining *HLXB9* expression in terms of breast cancer molecular subtypes to determine if there is a relationship between these subtypes and *HLXB9* expression

- Evaluating the prognostic significance of *HLXB9* expression in breast cancer patients' outcome using widely available survival databases.

6.3. Materials and Methods

The materials and methods used in this chapter are described in Chapter 2.

6.4. Results

I. *HLXB9* expression in non-cancer mammary cells and breast cancer cell lines by RT-PCR

Several breast cancer cell lines were screened to identify cell lines that could be studied to investigate the function of *HLXB9* overexpression in breast cancer. RT-PCR using primers for *HLXB9* and *β-actin* (loading control) were carried out on the following cells/cell lines:

- Non-cancer cells/cell lines: two primary human mammary epithelial cells (HMEC) 184D and 240L, and a non-tumorigenic epithelial cell line MCF 10A were used as non-cancer controls in this study.

- Breast cancer cell lines: breast cancer cell lines screened in this study were GI-101, MDA-MB-361, MDA-MB-157, T47D, SKBR3, MDA-MB-231, and BT474.

- Transformed derived cell lines from the non-tumorigenic MCF 10A cell line; MCF 10A-T1K and MCF 10A-CA1h were also analysed.

After the RT-PCR experiments, agarose gel electrophoresis analysis of amplified PCR products was carried out on a 1.2% agarose gel run at 80 volts.

HLXB9 was expressed by non-cancer mammary cells

Samples from six different passages of HMEC 184D cells (primary mammary epithelial cells) were provided by Dr Yasaei and investigated for *HLXB9* expression. A sample of HMEC 184D cells was taken at early passage (passage 7); at midpoint passage (passage 10) and at late passage (passage 14) when these cells start to senescence due to p16 accumulation (stress induced senescence) (Garbe et al. 2014;

Garbe et al. 2009). All three samples showed *HLXB9* expression and there was no observable difference in the level of expression of *HLXB9* (figure 6.1a, lane 1, 2 and 3).

The other three HMEC 184D samples were obtained after the cells were treated with Nickel (II) Chloride-$NiCl_2$, a carcinogen that rescues cell growth by helping them bypass senescence at passage 14. The samples at passage 14, passage 19 and passage 25 were also examined to investigate how *HLXB9* expression changes as cells bypass their normal cell division control mechanism. These samples of HMEC 184D also show *HLXB9* expression and there was no observable difference in the level of expression of *HLXB9* (figure 6.1a, lane 4, 5 and 6).

HMEC 240L, the other human mammary epithelial cell analysed also expressed *HLXB9* (figure 6.1b).

Finally, *HLXB9* expression was investigated in MCF 10A, an immortalised non-tumorigenic epithelial cell line widely used as a control (non-cancer) sample in breast cancer studies. MCF 10A cells also express *HLXB9* (figure 6.1c).

All breast cancer cell lines screened expressed HLXB9

Seven breast cancer cell lines were screened for *HLXB9* expression (as shown in figure 6.2a and b). The molecular characteristics of the cell lines investigated are described in table 6.1. The cell lines investigated were GI-101, MDA-MB-361, MDA-MB-157, T47D, SKBR3, MDA-MB-231, and BT474. All cell lines examined expressed *HLXB9* (figure 6.3a and b).

Cell Line	Molecular subtypes	*HER2*	*PR*	*ER*	*HLXB9* expression confirmed by RT-PCR experiments
GI-01	Basal	+	-	+/-*	Yes
MCF 10A	Basal	-	-	-	Yes
MDA-MB-361	Luminal A	+	+	+	Yes
MDA-MB-157	Basal	-	-	-	Yes
MDA-MB-157+BRK	Basal	-	-	-	Yes
T47D	Luminal A	-	+	+	Yes
SKBR-3	HER2 amplified	++	-	-	Yes
MDA-MB-231	Basal	-	-	-	Yes
BT474	Luminal B	++	+	++	Yes

Table 6.1: Table showing the molecular subtypes, expression profile of the biomarkers currently used in categorising breast tumours on cell lines studied and results of HLXB9 expression experiments. Cell line phenotype determined by the genes expressed by cell line. Human epidermal growth factor receptor 2 (HER2), Progesterone receptor (PR), Estrogen receptor (ER), (-) no expression, (+) expression, (++) overexpression, (+/-*) differing reports in research articles with expression observed in some studies (Neve et al. 2006,Mackay et al. 2009, Subik et al. 2010; Lacroix and Leclercq 2004)

Cell lines that showed other amplified bands other than band representing *HLXB9* at 359bp were MDA-MB-157 and MDA-MB-157 stably transfected with breast tumour kinase (BRK). These additional bands suggest there might be a different isoform of *HLXB9* being amplified or may represent artefacts from the PCR reaction. However, further investigations to identify these additional bands were unnecessary in this study as these experiments were designed to confirm *HLXB9* expression.

HLXB9 was also expressed by MCF 10A derivatives

MCF 10A and transformed cell lines derived from MCF 10A were used as a model to study the progressiveness of cancer development. In this study, MCF 10A and two of the three cell lines derived from MCF 10A; MCF 10A-T1K and MCF 10A-CA1h were analysed for *HLXB9* expression. MCF 10A-T1K is a pre-malignant cell line obtained by transforming the MCF 10A cell line with HRAS (Kadota et al. 2010, Wendt and Shiemann 2009). It is the first stage of the MCF 10A cancer progression model. MCF 10A-CA1h, a well differentiated metastatic tumour cell line derived from MCF 10A-T1K, is the next stage in the cancer progression model (Kadota et al. 2010, Wendt and Shiemann 2009). Unfortunately, the third stage in this model was not available therefore was not analysed as part of this study.

This breast cancer progression model is ideal for investigating the expression of *HLXB9* as cells become more tumorigenic. MCF 10A and its derivatives were found to express *HLXB9* by RT-PCR (figure 6.1c). There was also no observable difference in the quantity of *HLXB9* amplified indicated by the intensity of the bands obtained in the agarose gel

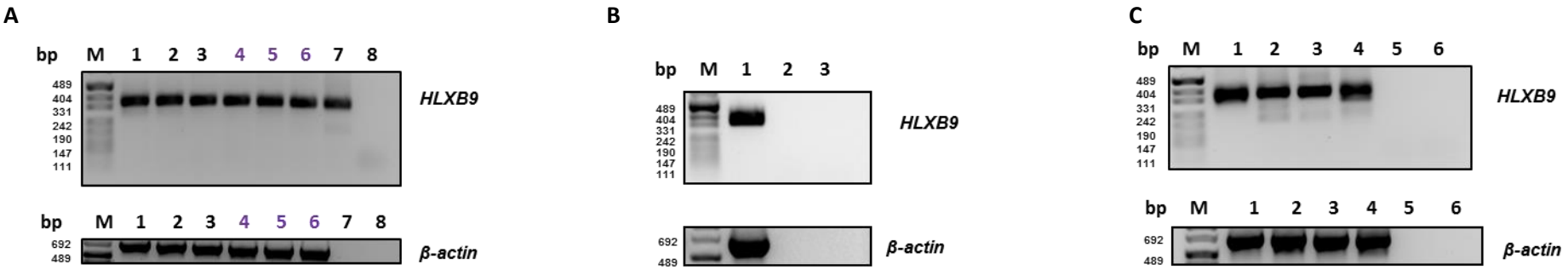

Figure 6.1: Agarose gel electrophoresis analysis of RT-PCR products in control (non-cancer) samples and MCF 10A cancer derivatives. RT-PCR products were electrophoresed in 1.2% Agarose gel at 80 Volts. Amplified products at 660bp obtained using primers for *β-actin* was used as both quality and quantity control. The *β-actin* PCR products in each sample was analysed before *HLXB9* RT-PCR was carried out to make sure cDNA template from samples did not show signs of degradation or contamination. The *β-actin* bands for each sample were also used to evaluate if the same starting amount of cDNA was used in the PCR reaction.

***Figure 6.1A: Agarose gel electrophoresis analysis of RT-PCR products showed HLXB9 was expressed by HMEC 184D cells at different passages* (n=3)**

Lanes: M= peqGOLD DNA SIZER XI, Lane 1= HMEC 184D passage 7, Lane 2= HMEC 184D passage 10, Lane 3= HMEC 184D passage 14, Lane 4= HMEC 184D passage 14 after treatment with NiCl2, Lane 5= HMEC 184D passage 19 after treatment with NiCl2, Lane 6=184D passage 25 after treatment with NiCl2, Lane 7= (Top) positive control (K562 cell line)/ (Bottom) negative control (negative cDNA), Lane 8= negative control (PCR mastermix).

(Top) Gel of RT-PCR reaction using *HLXB9* primers showing the amplified product at 359bp. Lane labelled with purple numbers represent HMEC 184D samples after treatment with NiCl2 to help cells bypass p16 mediated cell senescence. There was no observable difference in the quantity of *HLXB9* amplified indicated by the intensity of the bands obtained before and after treatment with NiCL2.

(Bottom) RT-PCR control gel showing the quality of cDNA samples using *β-actin* primers. The amplified products at 660bp showed the cDNA used in experiments were of good quality as there were no signs of degradation or contamination. The bands for each sample also show that the same starting amount of cDNA was used in the PCR reaction.

***Figure 6.1B: Agarose gel electrophoresis of RT-PCR product showed HLXB9 was expressed by HMEC 240L cells* (n=3).**

Lanes: M= peqGOLD DNA SIZER XI, Lane 1= HMEC, Lane 2= negative control (negative cDNA), Lane 3= negative control (PCR mastermix)

(Top) Agarose gel of RT-PCR reaction using *HLXB9* primers showing the amplified product at 359bp.

(Bottom) RT-PCR control gel showing the quality of cDNA sample using *β-actin* primers. The amplified products at 660bp showed the cDNA used in experiment was of good quality as there were no signs of degradation or contamination

***Figure 6.1C: Agarose gel electrophoresis of RT-PCR products showed HLXB9 was expressed by MCF 10A and its cancer derivatives: MCF 10A-TK1 and MCF 10A Ca1h* (n=3).**

Lanes: M= peqGOLD DNA SIZER XI, Lane 1= MCF 10A Lane 2= MC 10A-TK1, Lane 3= MCF 10A-Ca1h, Lane 4= positive control (K562 cell line), Lane 5=negative control (negative cDNA), Lane 6= negative control (PCR mastermix)

(Top) Gel of RT-PCR reaction using HLXB9 primers showing the amplified product at 359bp. There was no observable difference in the quantity of *HLXB9* amplified indicated by the intensity of the bands obtained

(Bottom) RT-PCR control gel showing the quality of cDNA samples using *β-actin* primers. The amplified products at 660bp showed the cDNA used in experiments were of good quality as there were no signs of degradation or contamination. The bands for each sample also showed that the same starting amount of cDNA was used in the PCR reaction.

Figure 6.2: Agarose gel electrophoresis analysis of RT-PCR products in breast cancer cell lines. RT-PCR products were electrophoresed in 1.2% Agarose gel at 80 Volts. Amplified products at 660bp obtained using primers for β-actin was used as both quality and quantity control.

***Figure 6.2A: Agarose gel electrophoresis analysis of RT-PCR products showed HLXB9 was expressed by the breast cancer cell lines investigated*.**

Lanes: M= peqGOLD DNA SIZER XI, Lane 1= Gi-101, Lane 2= MCF 10A, Lane 3= MDA-MB-361, Lane 4= MDA-MB-157 transfected with a control vector, Lane 5= MDA-MB-157 transfected with BRK, Lane 6= T47D, Lane 7= SKBR 3, Lane 8= positive control K562 (top image)/ control negative cDNA (bottom image), Lane 9=negative control PCR mastermix

(Top) Gel of RT-PCR reaction using *HLXB9* primers showing the amplified product at 359bp. Cell lines that showed other amplified bands other than band representing *HLXB9* at 359bp were MDA-MB-157 and MDA-MB-157 stably transfected with breast cancer kinase. These additional bands might represent the detection of a different isoform of *HLXB9* or artefacts from the PCR experiments.

(Bottom) RT-PCR control gel showing the quality of cDNA samples using *β-actin* primers. The amplified products at 660bp showed the cDNA used in experiments were of good quality as there were no signs of degradation or contamination. The band for each sample also showed the same starting amount of cDNA was used in the PCR reaction.

***Figure 6.2B: Agarose gel electrophoresis analysis of RT-PCR products showed HLXB9 was expressed by non- cancer cells/cell lines and breast cancer cell lines further analysed by qRT-PCR* (n=3)**

Lanes: M= peqGOLD DNA SIZER XI, Lane 1=HMEC 184D passage 7, Lane 2=HMEC 240L Lane 3=T47D, lane 4=MDA-MB-231, Lane 5=BT474, Lane 6= control mastermix

(Top) Gel of RT-PCR reaction using *HLXB9* primers showing the amplified product at 359bp. There was no observable difference in the quantity of *HLXB9* amplified indicated by the intensity of the bands obtained between non cancer cells and breast cancer cell lines investigated.

(Bottom) RT-PCR control gel showing the quality of cDNA samples using *β-actin* primers. The amplified products at 660bp showed the cDNA used in experiments were of good quality as there were no signs of degradation or contamination. The band for each sample also showed the same starting amount of cDNA was used in the PCR reaction.

II. Sequencing of RT-PCR products

The next step in the investigation of *HLXB9* expression in breast cancer was to determine if there was a differential expression between normal breast cells and breast cancer cell lines by quantitative RT-PCR (qRT-PCR). Therefore, cDNA from non-cancer breast cells/cell lines and breast cancer cell lines were sent to our collaborator, Professor Saccone.

The antisense HLXB9 mRNA- HLXB9-AS2 was detected in the cell lines analysed

However, as the primers had only been used in qRT-PCR in neuronal cell lines, RT-PCR experiments using the primers were carried out to determine the specificity of the primers for amplifying *HLXB9* in breast cells. Upon agarose gel electrophoresis analysis of the RT-PCR products, an expected band at 98bp representing *HLXB9* transcript variant one and two was observed. An additional unexpected band above 800bp was also detected in the MCF 10A, T47D, MDA-MB-231, and BT474 cell lines.

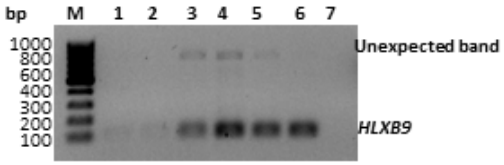

Figure 6.3: HLXB9 transcript variants 1 and 2 at 98bp and an additional band above 800bp in breast cancer cell lines analysed. An additional band was observed in MCF 10A, T47D, MDA-MB-231, and BT474 cell lines. Lanes: M= Peqlab 100bp ladder, Lane 1= HMEC 184D passage 7, Lane 2= HMEC 240L, Lane 3= MCF 10A, Lane 4= T47D, Lane 5= MDA-MB-231, Lane 6= BT474 and Lane 7= negative control, PCR mastermix.

In order to determine the identity of the unexpected band, the primer sequences were analysed through primer blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). The results from primer blast indicated the possibility of unspecific detection of *GNL3L* using the primers. Therefore, the PCR product from the T47D cell line was sent to Beckman Coulter Genomics Inc. UK for sequencing. The resulting sequence was analysed using nucleotide blast search (https://blast.ncbi.nlm.nih.gov). The result obtained from the sequencing reaction indicated the unexpected product above 800bp was most likely an anti-sense *HLXB9* RNA encoded by *HLXB9-AS2*.

III. Quantification of *HLXB9* expression in non-cancer mammary cells and breast cancer cells

Professor Saccone (University of Catania, Italy) provided the results described in this section. Sybr green Real Time PCR experiments using primers for *HLXB9* and *Tubulin* (control gene) were carried out in non-cancer and cancer breast cell lines. The relative concentration of *HLXB9* in the samples of interest to the control gene was obtained from the $2^{-\Delta\Delta Ct}$ formula (fold change between *HLXB9* expression to control gene). GDM1 and SK-N-BE cell lines were used as positive control for *HLXB9* expression. The GDM1 cell line is a leukaemia derived cell line known to overexpress *HLXB9* (Nagel et al. 2005). The SK-N-BE cell line is a neuroblastoma cell line that expresses *HLXB9* during proliferation, adding retinoic acid (RA) to this cell line increases *HLXB9* expression to induce neural differentiation. The increase in *HLXB9* expression is particularly significant five days after the addition of RA when *HLXB9* expression increases fivefold (Leotta et al. 2014).

Non-cancer primary HMEC cells showed a variable level of HLXB9 expression.

There was variability in the expression of *HLXB9* in the two samples of HMEC cells analysed. HMEC 240L cells showed a lower level of *HLXB9* expression compared to HMEC 184D cells (figure 6.4).

The expression of *HLXB9* in HMEC 184D cells was sampled at three time points in culture. These were: an early time point in culture (passage 7), mid-point in culture (passage 10) and a late time point in culture after treatment with $NiCl_2$, (passage 25). *HLXB9* was found to increase at passage 10 (mid-point) and decreases at passage 25 (late passage) to almost the same level at passage 7. Interestingly at passage 10, *HLXB9* expression in HMEC 184D cells was almost comparable to the GDM1 cell line. The GDM1 cell line is a leukaemia cell line known for *HLXB9* overexpression. Therefore *HLXB9* expression was found to increase as mammary cells become more proliferative and then reduce as the cells lose their proliferative ability. This indicates a role for *HLXB9* associated with proliferation in mammary cells.

The variability of *HLXB9* expression observed between the different HMECs samples studied (240L and 184D) might be due to of a number of reasons. One of these might be because the sample from the HMEC 240L cell line was taken at a late passage. Results from the HMEC 184D cell line suggests mammary cells lose *HLXB9* expression when their proliferative potential reduces at late passage. Other reasons include a variation in *HLXB9* expression between healthy individuals or a difference in patient's characteristics. The

HMEC 184D cells are derived from a patient undergoing reduction mammoplasty at 21 years old and the 240L derived from a patient undergoing reduction mammoplasty at 19 years old (Garbe et al. 2014).

HLXB9 was downregulated as cells became tumorigenic (results from MCF 10A and its transformed derivatives)

The expression of *HLXB9* was measured in the MCF 10A cell line and its transformed derivatives (figure 6.4). The MCF 10A cell line and its derivatives are used as a model to study cancer development as the cells go from non-tumorigenic MCF 10A to MCF 10A-TK1, a premalignant cell line to MCF 10A- CA1h, a well differentiated metastatic tumour cell line. *HLXB9* expression was highest in the non-tumorigenic MCF 10A cell line, whilst *HLXB9* expression reduced when the MCF 10A cell line becomes transformed (MCF 10A- TK1 cell line). *HLXB9* expression was at a similar level between the pre malignant MCF 10A-TK1 cell line and the malignant MCF 10A-CA1h cell line. This suggests a downregulation of *HLXB9* occurs early on in the cancer development, and this reduced expression of *HLXB9* persists as cells become more tumorigenic. These results indicate a tumour suppressive function for *HLXB9* in healthy mammary cells.

HLXB9 was downregulated in breast cancer cell lines

HLXB9 expression was measured in three breast cancer cell lines, T47D, BT474 and MDA-MB-231. There was a decrease in expression of *HLXB9* in these cell lines compared to two control (non-cancer) cells analysed (HMEC 184D cell and MCF 10A cell lines). However, a similar level of expression was observed when comparing HMEC 240L cells to the three breast cancer cell lines. This might be because the HMEC 240L cell were sampled at late passage when these cells are not actively proliferating (see figure 6.4). Comparison of the three breast cancer cell lines showed BT474, the luminal B type cell line had the least amount of *HLXB9* whilst T47D (Luminal A) and MDA-MB-231 cell line (Basal) had a similar level of expression of *HLXB9*. The reduced expression of *HLXB9* in the breast cancer cell lines indicates a loss of tumour suppressive function of *HLXB9* in cancer. This finding although contradictory to results presented in Chapter 5 was not surprising as other studies have also found downregulation of *HLXB9* in breast cancer (Tommasi et al. 2009; Lian et al. 2012; Siletz et al. 2013b; Siletz et al. 2013a).

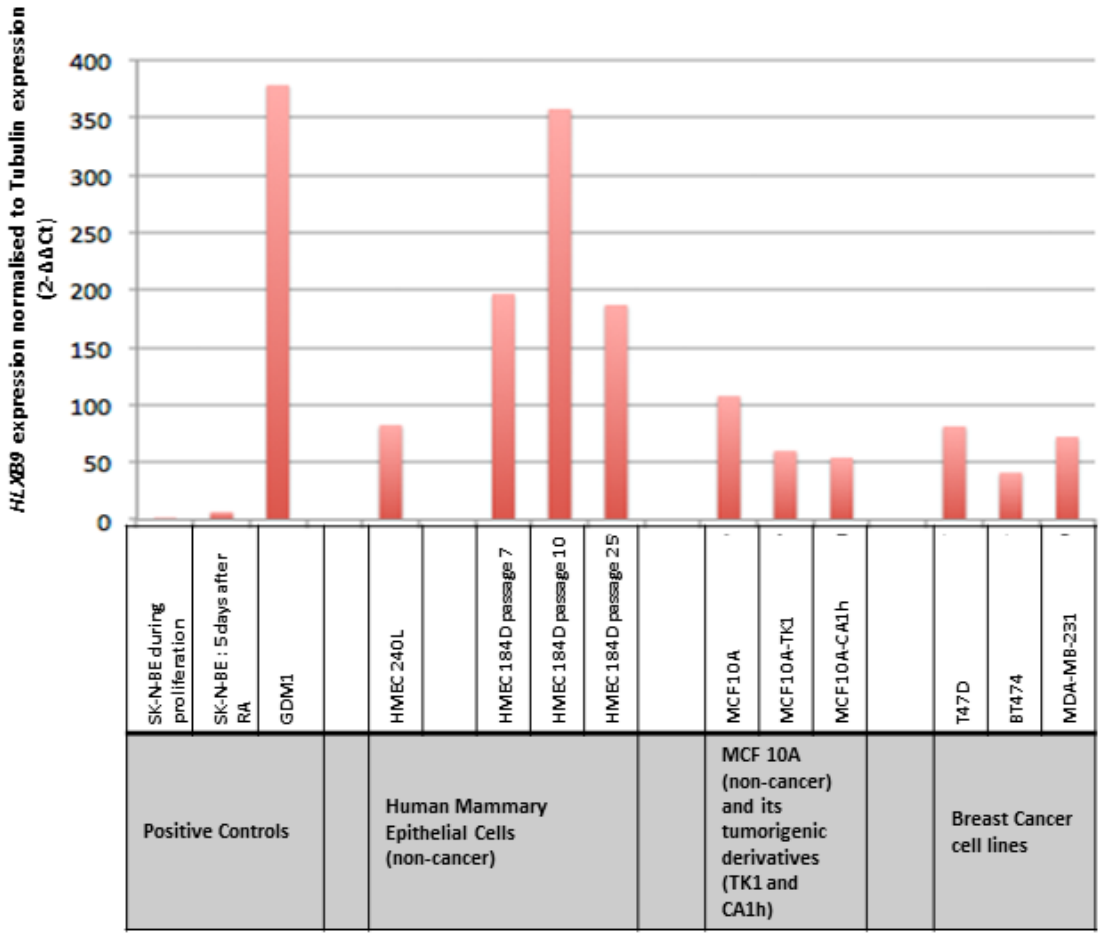

Figure 6.4: Histograms of *HLXB9* expression in controls and breast cancer cell (n=2). There was a reduction in the level of *HLXB9* in the breast cancer cell lines compared to two of the control non cancer cells/cell line (HMEC 184D and MCF 10A). However, a similar level of *HLXB9* expression was observed in the HMEC 240L cells compared to the breast cancer cell lines. The SK-N-BE and the GDM1 cell lines are used as positive controls. GDM1 is a leukaemia cell line known for very high levels of *HLXB9* expression. The SK-N-BE cell line is a neuronal cell line that expresses *HLXB9* during proliferation; upon stimulation with retinoic acid (RA) to induce differentiation, this cell line increases its *HLXB9* expression fivefold. Bars represent of $2^{-\Delta\Delta Ct}$ values of expression level normalised to *TBP* expression.

IV. HLXB9 expression and cellular localisation in breast cancer cell lines by indirect immunofluorescence (IF)

HLXB9 mRNA was found downregulated in breast cancer, Immunofluorescence (IF) experiments to detect the protein were carried out. Antibodies for HLXB9, β-actin and tubulin were used to identify and determine the cellular localisation of the HLXB9 protein in non-cancer MCF 10A cell lines and three breast cancer cell lines: T47D, BT474 and MDA-MB-231 (figure 6.5). It was impossible to determine the specificity of the HLXB9 antibody used, as no negative controls for HLXB9 protein expression were readily available. Additionally, due to time and resource constraints, knocking down the expression of HLXB9 by siRNA to determine the specificity of the antibody was also not possible.

HLXB9 was downregulated in breast cancer cell lines

The MCF 10A cell line (control non-tumorigenic cell line) had the greatest staining intensity (expression) of HLXB9. HLXB9 was present uniformly in the cell nucleus and cytoplasm in these cells.

The T47D (Luminal A) cell line also expresses HLXB9, however compared to the MCF 10A cell line, the staining intensity was reduced and there was a presence of protein aggregates (foci) in the nucleus. The MDA-MB-231 (Basal) cell line also expresses HLXB9, however compared to both control cell lines and the T47D, there was a reduction in the staining intensity with some cell nuclei having no staining for HLXB9. There were also foci of protein aggregates in nuclei when staining was observed. The BT474 cell line had the least amount of HLXB9 in the cytoplasm and the nuclei; HLXB9 was present as foci in the nucleus with some staining in the cytoplasm.

These results indicate a correlation between the levels of HLXB9 mRNA and protein in these cells. The results also showed a downregulation of HLXB9 protein suggesting a tumour suppressive function for HLXB9 in non-cancer mammary cell lines. Additionally, the organisation of HLXB9 as foci in the nucleus in the breast cancer cell lines (not observed in control) might be indicative of perturbed activity of HLXB9 in these cells.

The results obtained, although contrary to results obtained in chapter 5 were not surprising because a previous study had identified downregulation of HLXB9 in breast cancer (Neufing et al. 2003). The downregulation of HLXB9 observed might be due to variations within subtypes of breast cancers and might indicate that the spectrum of cell lines investigated was not comprehensive.

Unfortunately, due to time constraints, experiments to accurately quantify HLXB9 expression using a method such as Enzyme-linked Immunosorbent assay (ELISA) were not possible.

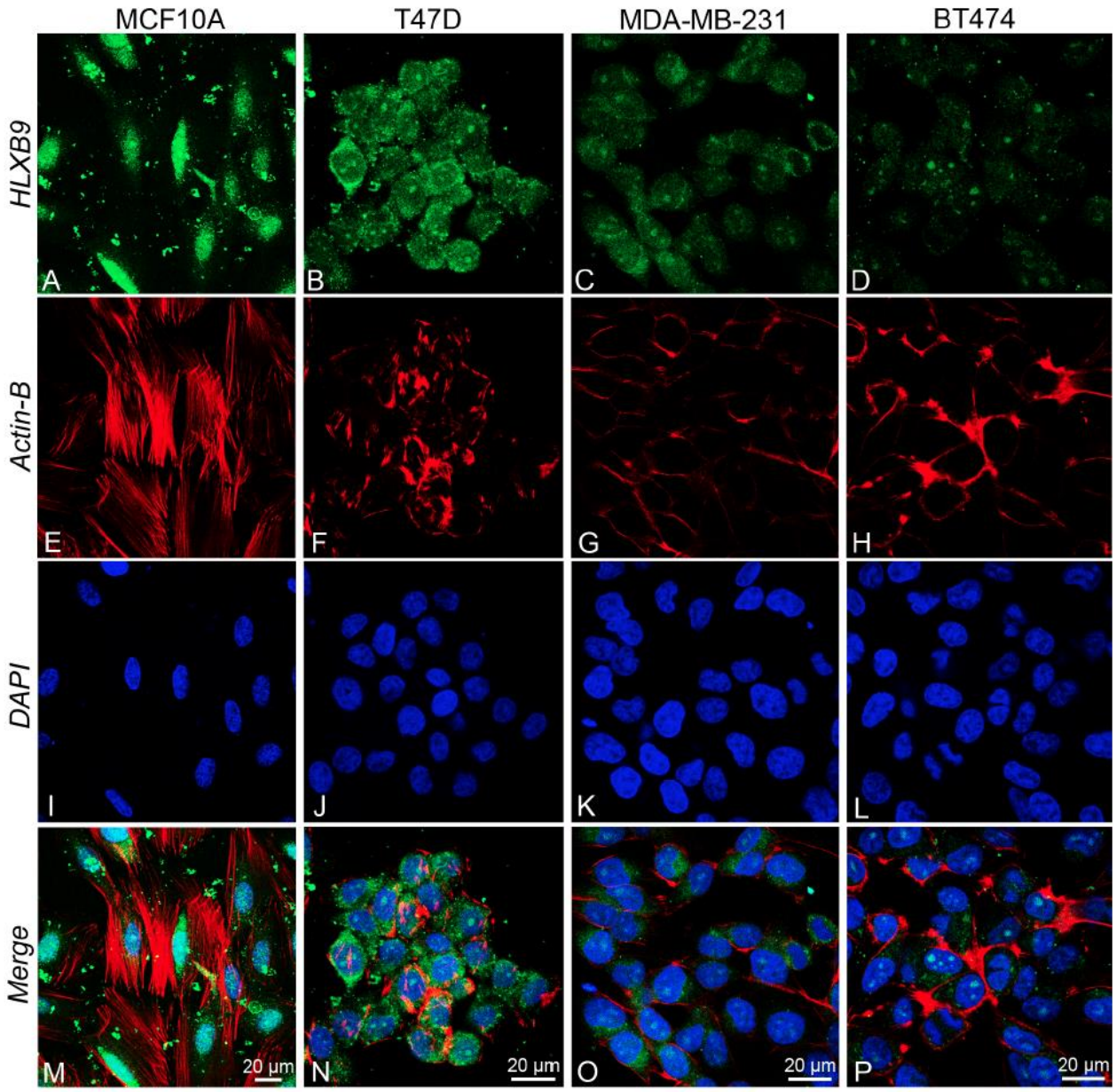

Figure 6.5: HLXB9 was downregulated in the breast cancer cell lines analysed. The MCF 10A cell line had the most HLXB9 expression (staining intensity). The MCF 10A also had a uniform expression of HLXB9 in the nucleus and the cytoplasm. Compared to the MCF 10A cell lines, the T47D had a reduced expression of HLXB9. HLXB9 localisation was also different in the T47D cell lines; HLXB9 was present as foci in the nucleus and surrounding the cytoplasm. The MDA-MB-231 cell line also had a reduced expression of HLXB9 compared to the control cell lines; HLXB9 was also present as foci in these cells. Finally, the BT474 cell had the least amount of HLXB9 compared to control and other cell lines investigated. HLXB9 was also present as foci in these cells. **βeta-Actin in CY3 (red), HLXB9 in FITC (green) and nuclei counterstained with DAPI (blue).**

V. Cytogenetic characterisation of non-cancer breast cells and breast cancer cell lines by FISH and/or G-banding

Breast cancer cell lines have been shown to harbour extensive chromosomal abnormalities that are known to evolve in culture (Rondón-Lagos et al. 2014; Watson et al. 2004). Therefore an attempt was made to characterise the cell lines used in this study with a particular focus on the number of *HLXB9* alleles present in each cell line. This was performed in conjunction with the nuclear localisation of the *HLXB9* gene in these cell lines.

Non-cancer cell lines had two copies of HLXB9 gene

The non-cancer cells/cell lines: HMEC 184D, HMEC 240L and MCF 10A used in this study were analysed by FISH and had two copies of the *HLXB9* gene (as shown in figure 6.6).

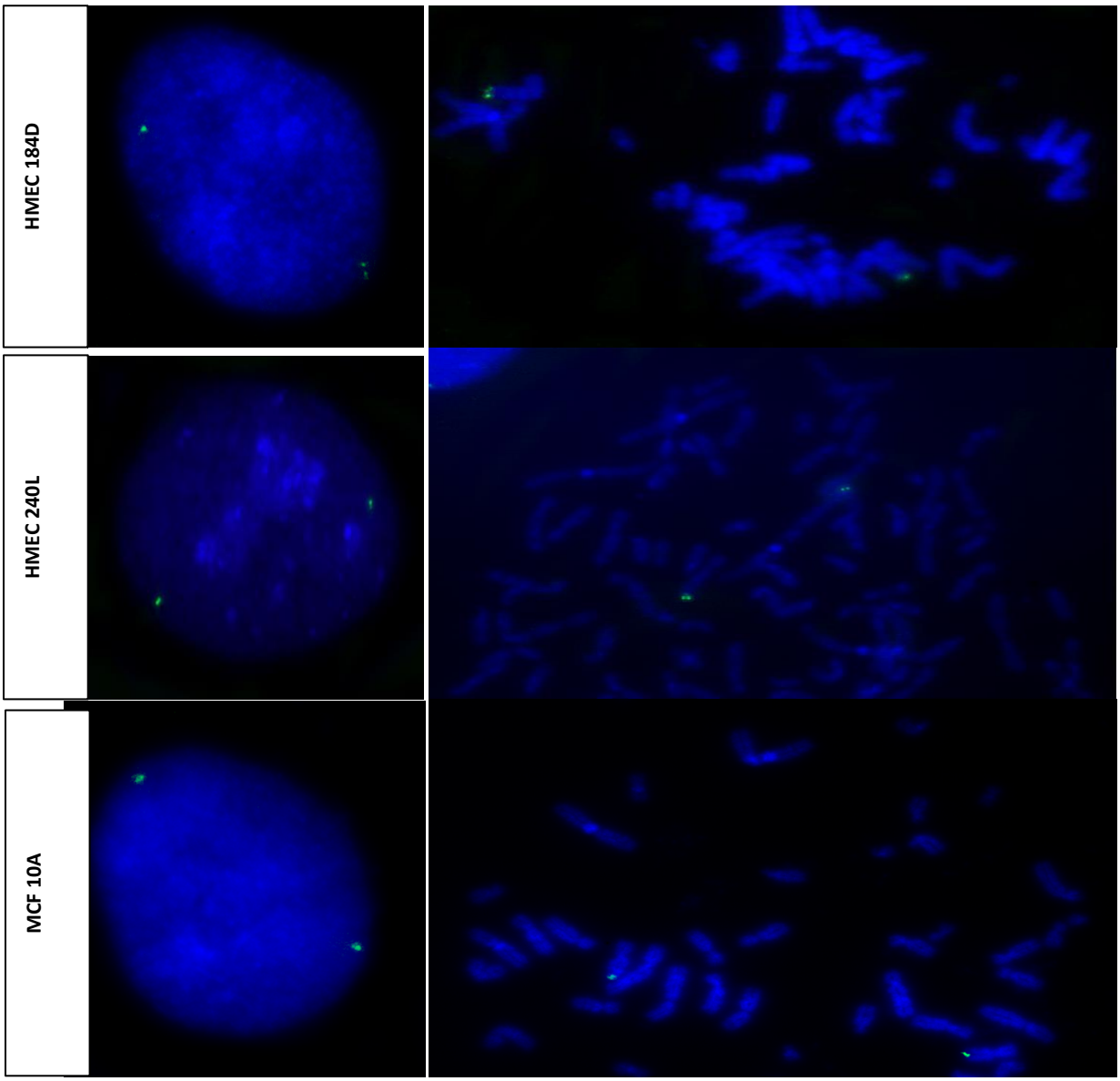

Figure 6.6: Fluorescence in situ hybridisation patterns observed in the metaphases and nuclei of non-cancer HMEC 184D, HMEC 240L and MCF 10A cells. Left: *HLXB9* in green (FITC labelled), and nucleus counterstained with DAPI. Right: *HLXB9* in green (FITC labelled), Chromosome counterstained with DAPI.

T47D cells also had two copies of HLXB9

The T47D cell line was analysed by G- banding and found to have a near triploid number of chromosomes (n= 32, modal number of 63 chromosomes as shown in figure 6.7B). Additional information on the type of chromosome abnormality could not be ascertained from the G-banded images. However, due to time and cost constraints, multiplex fluorescence in situ hybridisation (M-FISH) or spectral karyotype (SKY) could not be carried out on this and other cell lines used. A range of 48-65 chromosomes was observed. T47D similar to the control non-cancer cells used in the study had two *HLXB9* alleles by FISH (figure 6.7A).

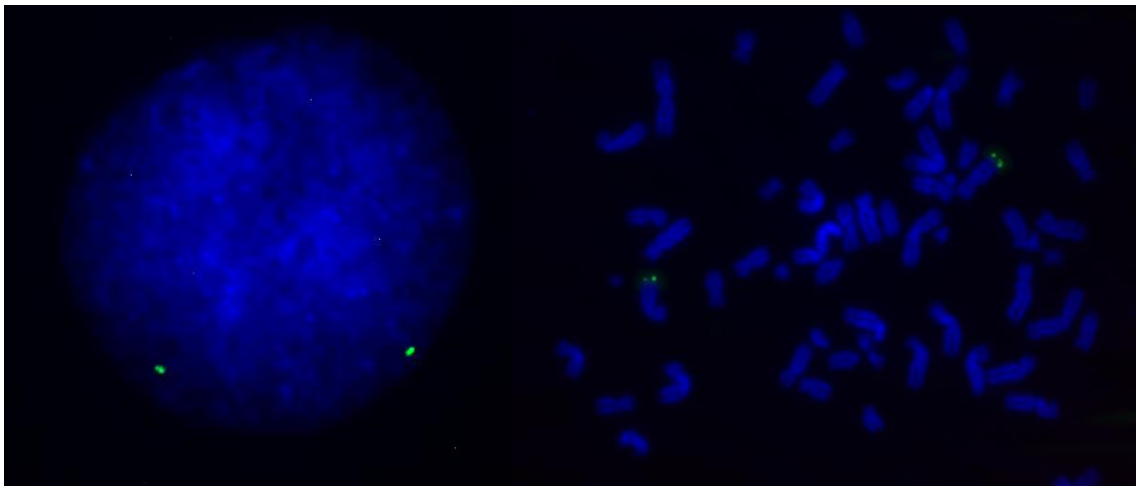

Figure 6.7A: Fluorescence in situ hybridisation patterns observed in the T47D cell line. (Left) *HLXB9* in green (FITC labelled), and nucleus counterstained with DAPI. (Right) Metaphase spread: chromosome counterstained in DAPI and *HLXB9* in green (FITC labelled).

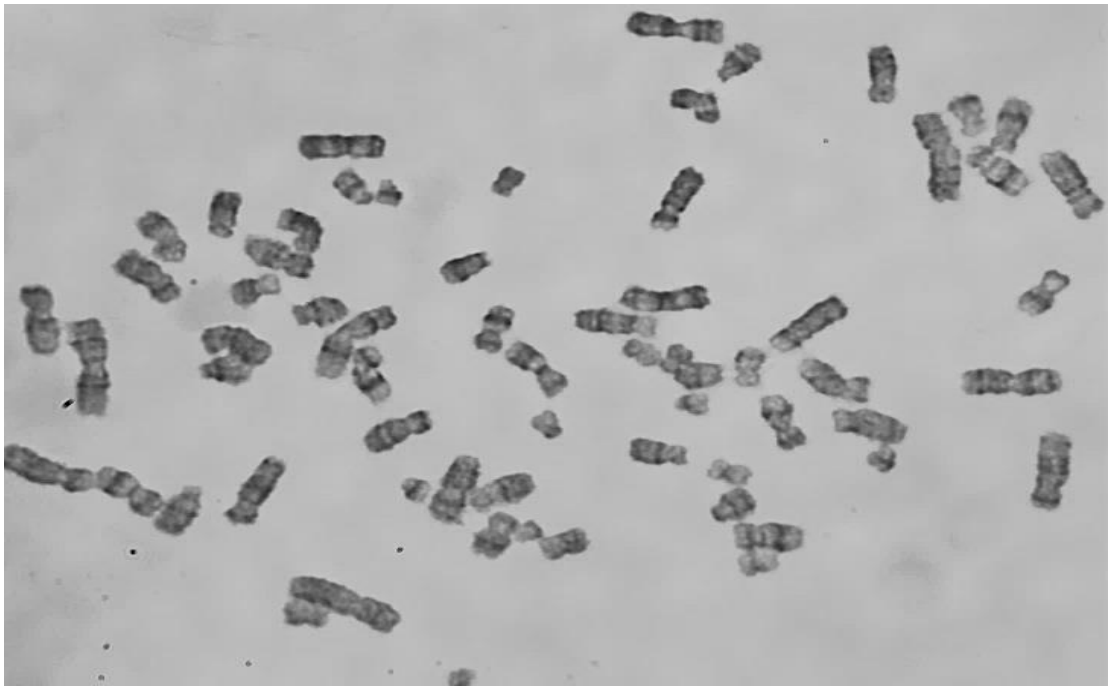

Figure 6.7B: A representative image showing G banded T47D chromosome spread (n=63).

BT474 cells had a variable number of *HLXB9* alleles (5-7n)

The BT474 cell line was analysed by G banding (figure 6.8E). This cell line had a tetraploid number of chromosomes; a modal number of 92 chromosomes was observed (n=15) between a range of 49-101 chromosomes. There was also variability in the number of *HLXB9* alleles in the BT474 nucleus from five to seven alleles (figure 6.8A-D); five alleles (19.2% of nuclei), six alleles (37.6% of nuclei) and seven alleles (43.2% of nuclei) by FISH.

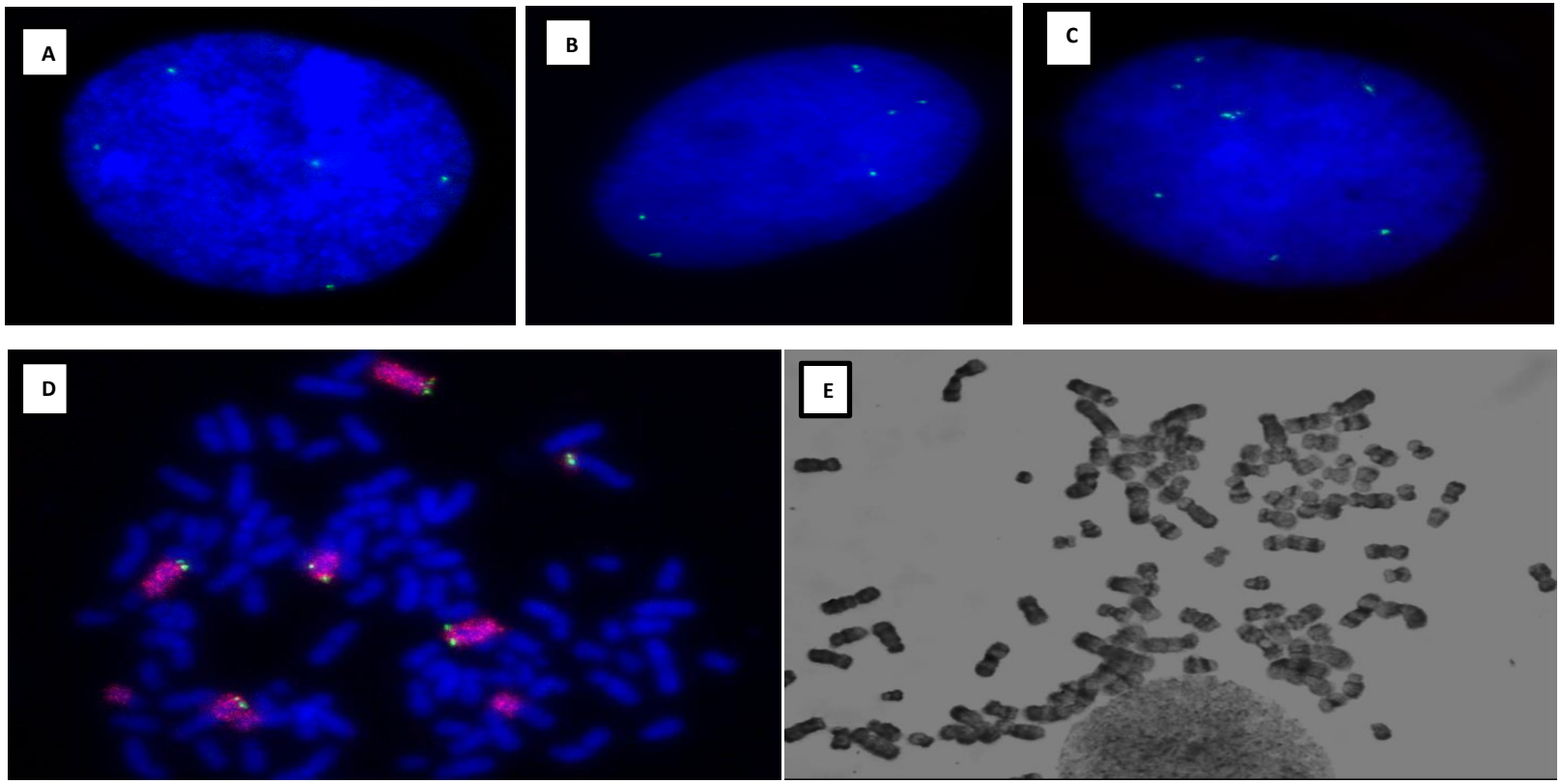

Figure 6.8: Fluorescence in situ hybridisation patterns observed in the metaphase and nuclei *HLXB9* in green (FITC labelled), chromosome and nuclei counterstained with DAPI and representative g banding chromosome spread of BT474 cell line. (A) BT474 nucleus with 5 *HLXB9* alleles. (B) BT474 nucleus with 6 *HLXB9* alleles. (C) BT474 nucleus with 7 *HLXB9* alleles. (D) Partial metaphase spreads of cells with 6 *HLXB9* alleles, chromosome 7 in red. (E) Representative (partial) metaphase spread by G banding.

MDA-MB-231 cells had a variable number of HLXB9 alleles (2-5n)

The MDA-MB-231 cell line was analysed solely by FISH. The cell line had a modal number of 73 chromosomes, a range of chromosomes 65-221 was observed (n=10) (Results used with permission from Dr Susan Wagland). There was also variability in the number of *HLXB9* alleles in the MDA-MB-231 cell line, this ranged from two copies (6.3 percent of cells), three copies (51 percent of cells), four copies (15.8 percent of cells) to five copies (26.8 percent of cells) (figure 6.9A-F).

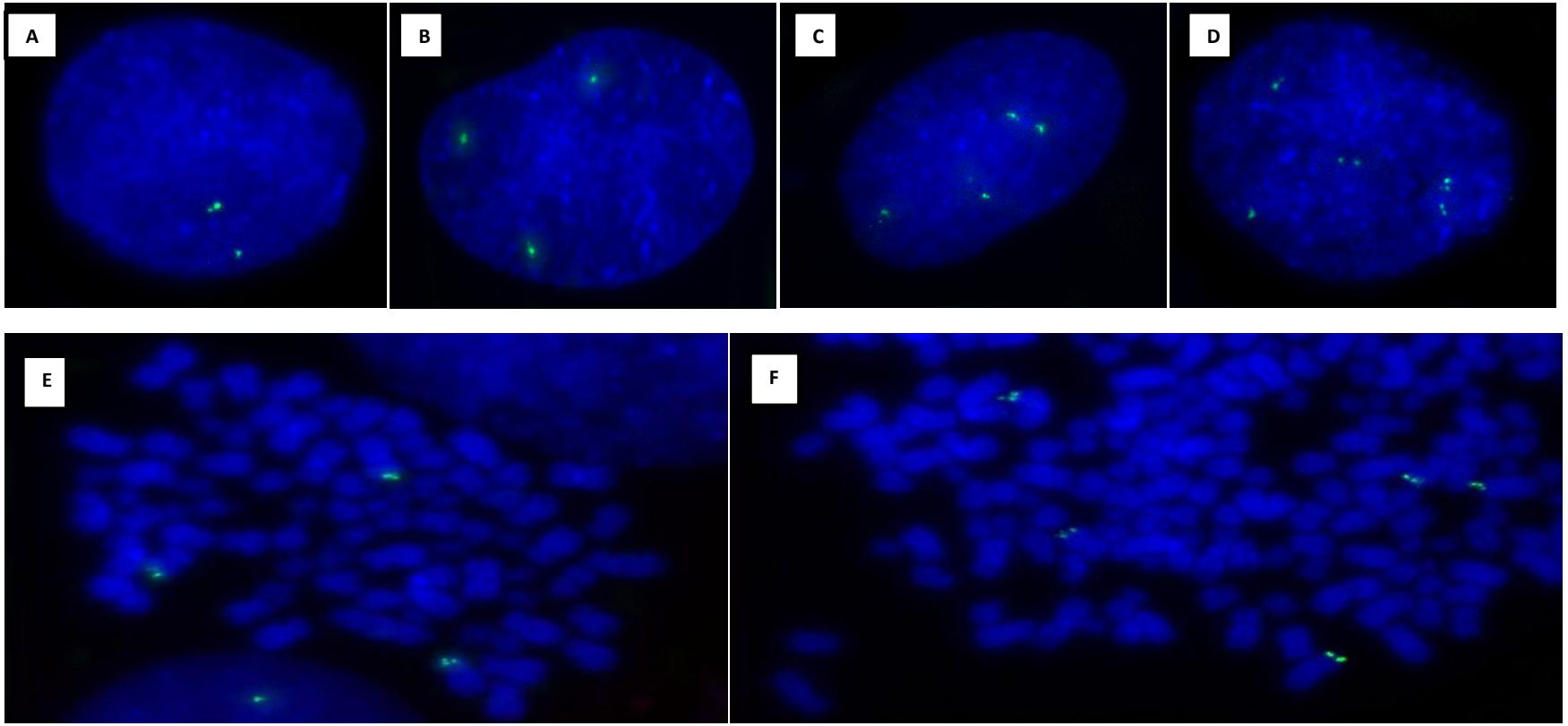

Figure 6.9: Fluorescence in situ hybridisation patterns observed in the metaphase and nuclei of MDA-MB-231 cell line *HLXB9* in green (FITC labelled), and Nuclei/Chromosome counterstained with DAPI. (A) MDA-MB-231 nucleus with 2 *HLXB9* alleles. (B) MDA-MB-231 nucleus with 3 *HLXB9* alleles. (C) MDA-MB-231 nucleus with 4 *HLXB9* alleles. (D) MDA-MB-231 nucleus with 5 *HLXB9* alleles. (E) Metaphase spread showing a cell with three copies of *HLXB9*. (F) Partial Metaphase spread showing a cell with five copies of *HLXB9*.

VI. Nuclear localisation of the *HLXB9* gene in non-cancer breast cells and breast cancer cells

As *HLXB9* expression has previously been linked to its nuclear localisation, the nuclear localisation of *HLXB9* in the non-cancer cells/cell lines and breast cancer cell lines studied were analysed. The aim of these experiments was to determine if a change in the nuclear localisation of *HLXB9* occurs in breast cancer. The results from the analysis are presented below and are shown in figure 6.10.

A. Non-cancer breast cells/ cell lines

HLXB9 was localised in the nuclear periphery of non-cancer primary human mammary epithelial cells (HMEC)

The *HLXB9* gene was localised in the nuclear periphery of both HMEC 184D cells (n=189) and HMEC 240L cells (n=191). There was no significant difference between the localisation of *HLXB9* in both non-cancer HMEC cells studied using an unpaired student's t-test. This result indicates the nuclear localisation of *HLXB9* was conserved between healthy individuals.

HLXB9 was also localised in the nuclear periphery of non-tumorigenic MCF 10A cells

HLXB9 was localised in the nuclear periphery in the MCF 10A cell line (n=190). There was no significant difference in the nuclear localisation of *HLXB9* in the MCF 10A cell line compared to the control HMEC cells studied (result of a one-way ANOVA and Bonferroni post hoc test). The result indicated that *HLXB9* nuclear localisation is conserved between healthy individuals.

B. Breast cancer cell lines

HLXB9 had a peripheral-intermediate nuclear localisation in T47D (Luminal A) cells

HLXB9 had a peripheral to intermediate localisation in 126 nuclei analysed. Although there was a shift away of the *HLXB9* gene from the periphery in this cell line, this shift was only significant when compared to one out of three control cell lines (figure 6.10). There was no significant difference in the localisation of *HLXB9* in the T47D cell line compared to both HMEC cells analysed. However, there was a significant decrease in the proportion of gene signals localised in the nuclear periphery in the T47D cell line compared to the MCF 10A cell line (P=0.018), (as shown in table 6.2).

T47D (n=126) vs	Proportion of Alleles in the Nuclear Periphery	Proportion of Alleles in the Nuclear Intermediate	Proportion of Alleles in the Nuclear Interior
HMEC 184D (n=189)	n.s.d	n.s.d	n.s.d
HMEC 240L (n=191)	n.s.d	n.s.d	n.s.d
MCF 10A (n=190)	Reduction in proportion of *HLXB9* gene signals in the nuclear periphery (*P* =0.018)	n.s.d	n.s.d

Table 6.2: The nuclear localisation of HLXB9 in the T47D cell line compared to the two non-cancer HMEC cells and MCF 10A cell line analysed using a one-way ANOVA and Bonferroni post hoc test. There was no significant difference (n.s.d) in the localisation of the *HLXB9* gene in the nucleus of T47D cell line compared to the HMEC 184D and 240L cells. There was a very significant reduction in the proportion of genes localised in the peripheral shell in the nucleus (*P* =0.018) of the T47D cell line compared to the MCF 10A cell line.

HLXB9 had an intermediate nuclear localisation in BT474 cells (Luminal B) cells

HLXB9 had an intermediate localisation in BT474 nucleus (n=257). In this cell line, there was a significant reduction in the proportion of alleles in the nuclear periphery compared to the non-cancer cells analysed. There was also a significant increase in the proportion of alleles in the nuclear interior compared to two out of the three non-cancer cells analysed (table 6.3). This indicates a change from the normal peripheral position of the *HLXB9* gene observed in the non-cancer nuclei to a more interior position in the BT474 cell nuclei.

BT474 (n=257) vs	Proportion of Alleles in the Nuclear Periphery	Proportion of Alleles in the Nuclear Intermediate	Proportion of Alleles in the Nuclear Interior
HMEC 184D (n=189)	Reduction in proportion of *HLXB9* gene signals in the nuclear periphery (*P* = 0.03)	No significant difference (n.s.d)	n.s.d
HMEC 240L (n=191)	Reduction in proportion of *HLXB9* gene signals in the nuclear periphery (*P* =0.001)	n.s.d	Increase in proportion of *HLXB9* gene signals in the nuclear interior (*P* < 0.0001)
MCF 10A (n=190)	Reduction in proportion of *HLXB9* gene signals in the nuclear periphery (*P* < 0.0001)	n.s.d	Increase in proportion of *HLXB9* gene signals in the nuclear interior (*P* = 0.005)

Table 6.3: The nuclear localisation of *HLXB9* in the BT474 cell line compared to the two non-cancer HMEC cells and MCF 10A cell line analysed by a one-way ANOVA and Bonferroni post hoc test. There was a significant decrease in the proportion of gene signals observed in the nuclear periphery in the BT474 cell line compared to the non-cancer cell lines studied. There was also a significant increase in the proportion of gene signals in the BT474 cell line nuclear interior compared to two out of the three non-cancer cells analysed. This indicates a change in the localisation of *HLXB9* from the nuclear periphery into the nuclear interior in the BT474 cell line.

HLXB9 had an intermediate nuclear localisation in MDA-MB-231 (Basal) cells

HLXB9 also had an intermediate position in the nucleus of MDA-MB-231 cells (n=272). In this cell line, there was a significant reduction in the proportion of *HLXB9* alleles in the nuclear periphery in the MDA-MB-231 cell line when compared to all non-cancer cells studied. There was also an increase in the

proportion of gene signals in the nuclear interior compared to HMEC 240L and MCF 10A only. This indicates a change from the normal peripheral position of the *HLXB9* gene observed in the non-cancer nuclei to a more interior position in the MDA-MB-231 nuclei.

MDA-MB-231 cells (n=272) vs	Proportion of Alleles in the Nuclear Periphery	Proportion of Alleles in the Nuclear Intermediate	Proportion of Alleles in the Nuclear Interior
HMEC 184D (n=189)	Reduction in proportion of *HLXB9* gene signals in the nuclear periphery (P =0.030)	n.s.d	n.s.d
HMEC 240L (n=191)	Reduction in proportion of *HLXB9* gene signals in the nuclear periphery (P =0.009)	n.s.d	Increase in proportion of *HLXB9* gene signals in the nuclear interior (P =0.007)
MCF 10A (n=190)	Reduction in proportion of *HLXB9* gene signals in the nuclear periphery (P <0.0001)	n.s.d	Increase in proportion of *HLXB9* gene signals in the nuclear interior (P =0.050)

Table 6.4: Comparison of the nuclear localisation of *HLXB9* in MDA-MB-231 cell line to the two non-cancer HMEC cells and MCF 10A cell line analysed by a one-way ANOVA and Bonferroni post hoc test. There was a significant decrease in the proportion of gene signals observed in the nuclear periphery in the MDA-MB-231 cell line compared to the non-cancer cell lines studied. There was also a significant increase in the proportion of gene signals in the MDA-MB-231 cell line nuclear interior compared to HMEC 240L and MCF 10A only. This indicates a change in the localisation of *HLXB9* from the nuclear periphery into the nuclear interior in the MDA-MB-231 cell line.

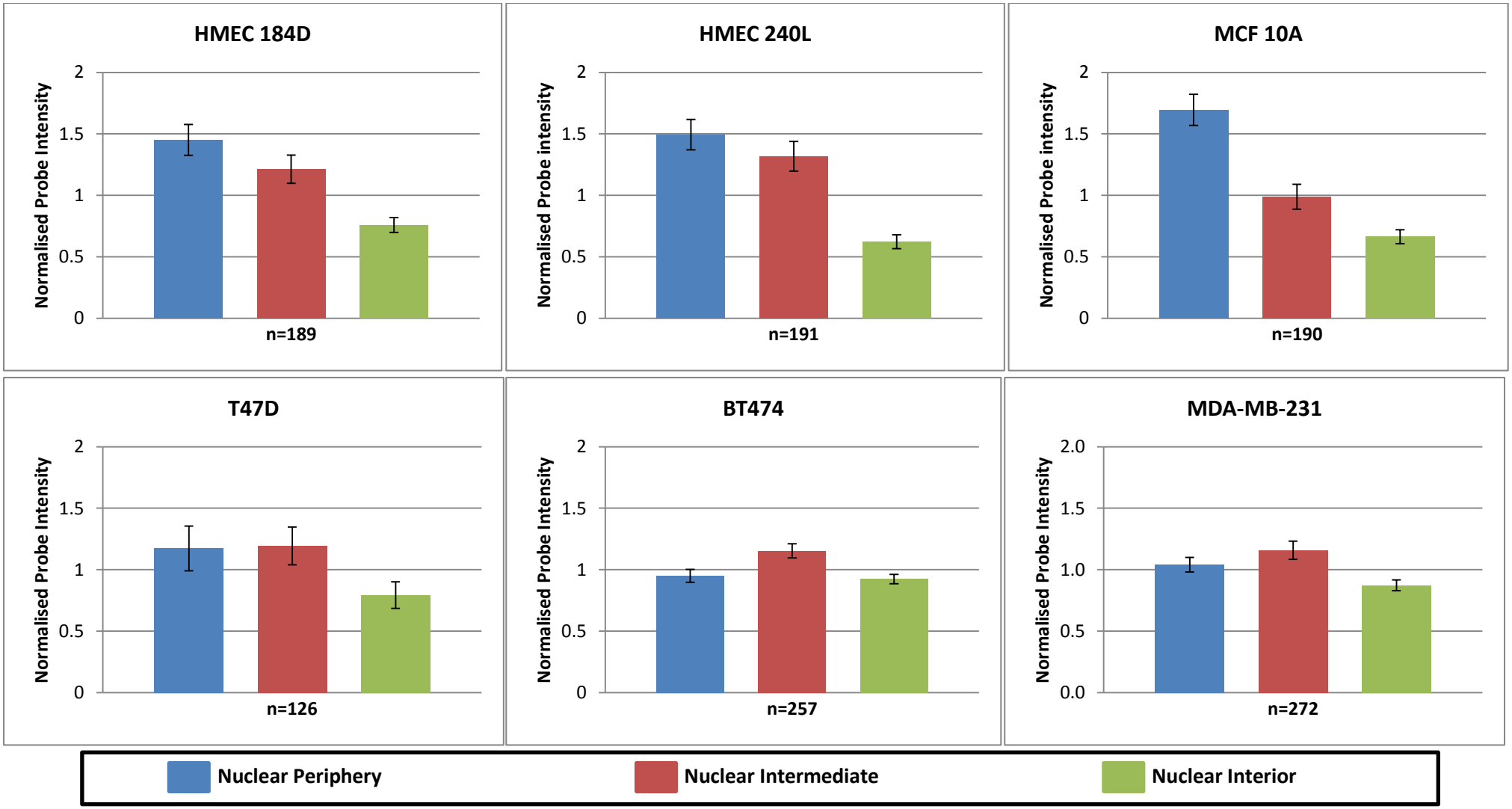

Figure 6.10: Nuclear localisation of HLXB9 gene in non-cancer breast cells/cell lines and Breast cancer cell lines studied. Controls: *HLXB9* has a peripheral position in the control cells used HMEC184D, HMEC 240L and MCF 10A. Breast Cancer Cell lines: T47D: *HLXB9* has a peripheral-intermediate position in this cell line, BT474 and MDA-MB-231: *HLXB9* has an intermediate position in these cells. 2D FISH images were analysed by erosion script analysis to determine the nuclear localisation of *HLXB9*. The histogram displays the average probe intensity (normalised by DAPI staining) on the y-axis and the number of nuclei studied on the x-axis. The standard error bars represent the standard error of the mean for each nuclear position.

VII. The effect of *HLXB9* expression on breast cancer patients' survival

As *HLXB9* potentially acts as a tumour suppressor in breast cancer, the prognostic significance of aberrant *HLXB9* expression was evaluated in breast cancer patients using online databases: KM plotter and *BreastMark* (Gyorffy et al. 2013; Madden et al. 2013). In both databases, the patient cohort was divided into two groups by median gene expression. Patients with lower gene expression than the median were classified as showing low gene expression whilst patients with greater gene expression than the median were classified as showing high gene expression. A survival curve showing the patients' outcome over time by level of gene expression was generated by the databases (as shown in figure 6.11). The difference in patients' survival between the two groups was assigned a prognostic value (hazard ratio with 95% confidence intervals). Hazard ratio less than 1 indicated a positive impact on prognosis whilst hazard ratios greater than 1 indicated a negative impact on prognosis. The *P* value of the hazard ratio was also calculated. *P* values equal or less than 0.05 indicated the hazard ratio observed was statistically significant.

Patients' outcome were categorised into:

- overall survival (OS), which is the time from either diagnosis or treatment that patients are alive;
- relapse free survival (RFS) or disease free survival (DFS), which is the time after primary treatment that the patient survives without any sign and symptom of cancer returning in the original tumour site;
- disease free metastatic survival (DFMS) or distant disease free survival (DDFS), which is the time after treatment or diagnosis that patients are alive without metastasis to secondary tumour site(s) and
- post progression survival (PPS) which is the time after relapse or metastasis that patients are alive.

Two multi-gene expression tests are now used clinically to determine the probability of disease reoccurrence in patients in low-grade tumours without lymph node involvement. (Madden et al. 2013; Marrone et al. 2015). These tests are Oncotype DX, a 21-gene panel test (5 control genes and 16 oncogenes) and MammaPrint a 70-gene panel test. In the UK, Oncotype DX is approved by NICE for determining if patients should receive chemotherapy after surgery in ER positive, HER 2 negative, lymph node negative and early grade breast tumours (https://www.nice.org.uk/guidance/dg10). MammaPrint

is only approved for research purposes in the UK, although it is used in the US in patients under 61 years with grade 1or 2 tumours without lymph node involvement (Marrone et al. 2015).

The 16 oncogenes in the oncotype DX panel is composed of five proliferation genes (*KI67, STK15, Survivin, CCNB1* and *MYBL2*), two invasion genes (*MMP11* and *CTSL2*), two HER2 associated genes (*GRB7* and *HER2)*, ER defining genes (*ER, PGR, BCL2* and *SCUBE2*) and three other genes (*GSTM1, CD68, BAG1*). The expression of these genes are normalised to the control genes and then scored. Proliferation genes have the highest impact on the overall score. In MammaPrint, 16 genes are associated with good prognosis whilst 45 genes are associated with poor prognosis (results could only be obtained on 43 out of 45 genes in *BreastMark*).

Therefore, when *HLXB9* has a statistically significant hazard ratio on patients' survival in both KM plotter and *BreastMark*, the hazard ratio obtained will be compared to the panel of genes utilised in both tests. This will help determine if *HLXB9* expression has better prognostic significance than some of the genes currently used and if it can be added to existing tests to improve their efficacy.

High HLXB9 expression had a deleterious effect on breast cancer patients' survival (RFS and OS)

In this section, using both databases, the effect of *HLXB9* expression was examined in breast cancer patients irrespective of tumour molecular subtype, tumour grade and lymph node involvement (see table 6.5 and figure 6.11 for hazard ratios, *P* values and survival curves generated by both databases).

In both databases, high expression of *HLXB9* was found to significantly reduce the probability of overall survival (KM plotter: HR=1.52, *P*=0.0001; *BreastMark*: HR=1.42, p=5.1E-05). These results indicate breast cancer patients with high *HLXB9* expression showed reduced survival time compared to patients with low *HLXB9* expression. High expression of *HLXB9* was also found to have a deleterious effect on patients' survival when looking at DFS in both databases (KM plotter: HR=1.41, *P*=9.1E-07; *BreastMark*: HR=1.20, *P*=0.0021). These results suggest that high expression of *HLXB9* significantly reduced the time breast cancer patients survived without relapse in the original site of the tumour.

In KM plotter, high expression of *HLXB9* was also found to significantly reduce breast cancer patient DMFS (HR=1.42, *P*=3.5E-04). However this effect was not significant in *BreastMark* (HR=1.16; *P*=0.079).

As the *BreastMark* database had a larger sample size, these results suggest that high expression of *HLXB9* probably has limited prognostic significance on breast cancer patients' metastasis (DMFS).

Data on PPS of breast cancer patients was only available in KM plotter. When the survival curve for PPS was analysed, overexpression of *HLXB9* was deterimental to patient survival, HR: 1.32, *P*=0.024.

Patients' Outcome	KM plotter			*BreastMark*		
	n Number	Hazard Ratio	*P* value	n Number	Hazard Ratio	*P* value
Disease Free Survival (DFS/RFS)	3951	1.41	9.1E-10	2592	1.20	0.0021
Distant Metastasis Free Survival (DMFS)	1746	1.42	3.5E-04	2223	1.16	0.079
Post Progression Survival (PPS)	414	1.32	0.024	N/A	N/A	N/A
Overall Survival (OS)	1402	1.52	0.0001	2091	1.42	5.1E-05

Table 6.5: High *HLXB9* expression is a negative prognostic indicator in breast cancer. In both databases, high expression of *HLXB9* reduces the probability of disease free survival and overall survival. High expression of *HLXB9* significantly reduces the probability of distant disease free survival of patients in KM plotter but not in *BreastMark* (Gyorffy et al. 2013; Madden et al. 2013). Highlighted results have hazard ratios that were statistically significant.

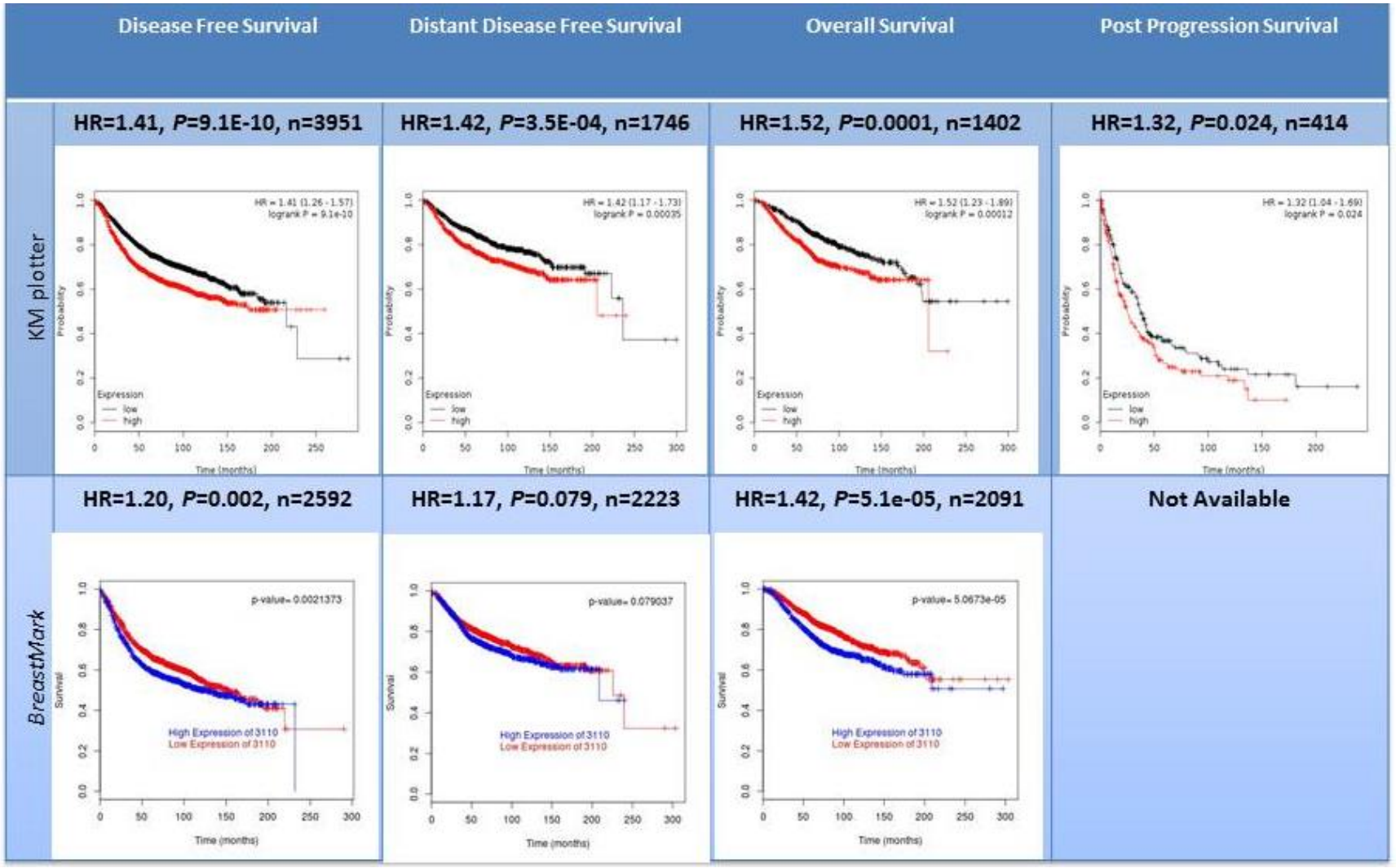

***Figure 6.11: HLXB9 overexpression is a negative prognostic indicator in breast cancer. Breast cancer survival curves from KM plotter and BreastMark* (Gyorffy et al. 2013; Madden et al. 2013).** In both databases, overexpression of *HLXB9* reduces the probability of disease free survival and overall survival. Over expression of *HLXB9* also significantly reduces the probability of distant metastasis free survival of patients in KM plotter only. This effect on patient survival was not observed in *BreastMark* as *BreastMark* analysed a larger number of sample for DMFS, high expression of *HLXB9* probably has limited impact on DDFS survival of breast patients.

The impact of HLXB9 expression on patients' relapse free and overall survival compared to genes used in Oncotype DX and MammaPrint

The impact of overexpression of *HLXB9* on breast cancer patients' relapse free and overall survival was found to be significant in both KM plotter and *BreastMark*. Therefore, the hazard ratio obtained was compared to the panel of genes used in both the Onco*type* DX and MammaPrint gene expression test.

Onco*type* DX

From KM plotter, *HLXB9* had the 5th most significant impact on breast cancer patients relapse free survival when compared to the 16 oncogenes used in the Onco*type* DX test (performing better than the invasion genes *MMP11* and *CSTL2* in the panel). In *BreastMark*, *HLXB9* ranked 9th in prognostic significance when compared to the same gene panel. Genes that had a better hazard ratio than *HLXB9* in the Onco*type* DX panel were genes associated with proliferation (*KI67, STK15, Survivin, CCNB1 and MYBL2*) and genes associated with invasion (*MMP11* and *CTSL2*) as shown in table 6.6. These results indicate that *HLXB9* has more prognostic significance than approximately half of the genes used in the panel.

Entrez ID	Gene Symbol	*BreastMark*			KM plotter		
		HR	*P*	(n)	HR	*P*	(n)
3110	*HLXB9*	1.2	2.1E-03	2592	1.4	9.1E-10	3951
Onco*type* DX gene panel							
4288	*KI67*	1.4	1.9E-09	2652	1.4	7.2E-09	3951
6790	*STK15*	1.5	2.5E-12	2652	1.9	1.0E-16	3951
332	*Survivin*	1.6	1.6E-15	2652	1.7	1.0E-16	3951
891	*CCNB1*	1.5	2.7E-10	2248	2.1	1.0E-16	1764
4605	*MYBL2*	1.5	7.0E-12	2652	1.8	1.0E-16	3951
4320	*MMP11*	1.3	1.7E-06	2592	1.2	6.2E-03	3951
1515	*CTSL2*	1.5	4.7E-10	2592	1.4	1.8E-09	3951
2886	*GRB7*	1.2	1.9E-04	2652	0.9	6.9E-03	3951
2064	*HER2*	0.9	2.2E-01	2592	0.9	2.5E-01	1764
2099	*ER*	0.9	2.1E-02	2592	0.7	2.7E-14	3951
5241	*PGR*	0.7	3.2E-08	2652	0.5	1.2E-15	1764
596	*BCL2*	0.8	4.5E-04	2592	0.6	1.0E-16	3951
57758	*SCUBE2*	0.7	1.1E-04	1759	0.6	1.0E-16	3951
2944	*GSTM1*	0.8	4.2E-03	1753	0.6	2.2E-16	3951
968	*CD68*	1.1	3.4E-02	2652	0.8	1.8E-03	3951
573	*BAG1*	0.9	3.1E-01	2652	0.8	6.7E-03	1764

Table 6.6: Overexpression of *HLXB9* is a better prognsotic indicator of relapse in breast cancer patients than a number of Onco*type* DX genes. Results from analysing relapse free survival in KM plotter and *BreastMark*.

The impact of *HLXB9* overexpression on overall survival was also compared to genes in the Onco*type* DX gene panel. *HLXB9* had the 5th most significant impact on patient overall survival when compared to the 16 oncogenes in the panel (KM plotter, 8th in *BreastMark* as shown in table 8.1 in appendix). These finding indicates that *HLXB9* has more prognostic significance than a number of other genes in the panel.

Therefore *HLXB9* overexpression is a better indicator of disease progression and outcome in breast cancer patients than those genes.

MammaPrint

As expected, genes associated with good prognosis had a hazard ratio of less than or equal to 1 whilst most genes associated with a poor prognosis had a hazard ratio greater than 1 when their impact on breast cancer patients' RFS was compared in both KM plotter and *BreastMark*. When *HLXB9* was added to the 43 genes associated with a poor prognosis, overexpression of *HLXB9* was the 17th most significant indicator of poor prognosis in KM plotter (25th in *BreastMark*, as shown in Table 6.7).

Entrez ID	Gene Symbol	*BreastMark*			KM plotter		
		HR	*P*	(n)	HR	*P*	(n)
3110	*HLXB9*	1.2	2.1E-03	2592	1.4	9.1E-10	3951
Genes associated with good prognosis in MammaPrint							
8659	*ALDH4*	1.03	6.0E-01	2497	1	9.9E-01	3951
8817	*FGF18*	0.93	2.6E-01	2652	0.64	1.8E-08	1764
27113	*BBC3*	0.97	6.0E-01	2297	0.79	2.6E-05	3951
57593	*KIAA1442*	0.70	6.5E-02	157	0.57	2.4E-12	3951
57758	*CEGP1*	0.75	1.1E-04	1759	0.59	1.0E-16	3951
146923	*RUNDC1*	0.76	3.3E-04	1238	0.51	1.0E-16	1764
8840	*WISP1*	0.93	2.4E-01	2643	0.73	6.0E-05	1764
2947	*GSTM3*	0.85	7.4E-03	2652	0.81	7.3E-03	1764
151126	*ZNF533*	0.84	2.7E-02	1229	0.58	9.0E-12	1764
146760	*RTN4RL1*	0.91	2.7E-01	1023	0.72	3.4E-05	3951
10455	*PECI*	0.85	9.3E-03	2421	0.87	1.5E-02	3951
7043	*TGFB3*	0.87	2.0E-02	2592	0.69	1.5E-11	3951
55351	*HSA250839*	0.78	8.8E-05	2516	0.7	6.9E-11	3951
58475	*CFFM4*	0.74	5.6E-05	1473	0.68	8.7E-07	1764
163	*AP2B1*	0.99	9.1E-01	2592	0.73	1.8E-08	3951
79132	*LGP2*	0.92	2.2E-01	2281	0.83	1.0E-03	3951
Genes associated with poor prognosis in MammaPrint							
55321	*C20orf46*	1.11	8.0E-02	2576	0.86	5.2E-03	3951
11082	*ESM1*	1.30	1.3E-05	2544	1.16	8.4E-03	3951
9134	*CCNE2*	1.50	1.4E-10	2357	1.86	1.0E-16	3951
54583	*EGLN1*	1.29	5.1E-05	2272	1.28	1.7E-03	1764
1058	*CENPA*	1.54	6.6E-13	2652	1.7	1.0E-16	3951
9055	*PRC1*	1.61	4.1E-15	2576	2	1.0E-16	3951
445815	*AKAP2*	0.95	4.4E-01	1899	0.86	8.8E-03	3951
10874	*NMU*	1.52	1.9E-12	2652	1.64	1.0E-16	3951
3488	*IGFBP5*	1.13	4.8E-02	2592	1.1	2.2E-01	1764
10531	*MP1*	1.10	1.9E-01	2004	1.32	5.1E-04	1764
57110	*LOC57110*	1.36	5.2E-07	2516	1.54	8.2E-15	3951
8577	*TMEFF1*	1.16	1.6E-02	2437	1.15	1.1E-02	3951
4175	*MCM6*	1.56	8.8E-14	2652	1.36	9.8E-05	1764
643008	LOC643008	NA	NA	NA	1.02	8.0E-01	3951
83879	*CDCA7*	1.26	3.2E-03	1238	1.57	1.0E-08	3951
5984	*RFC4*	1.47	1.4E-10	2652	1.72	1.0E-16	3951
23594	*ORC6L*	1.54	6.3E-13	2576	1.62	1.0E-16	3951
6515	*SLC2A3*	1.08	1.8E-01	2592	1.04	4.7E-01	3951
57211	*DKFZP564D0462*	1.09	1.8E-01	2297	1.32	5.7E-07	3951
79791	*FBXO31*	1.06	3.4E-01	2576	1.13	1.2E-01	1764
1633	*DCK*	1.23	7.8E-04	2592	1.39	2.5E-09	3951

51514	*L2DTL*	1.54	1.4E-12	2516	1.82	6.6E-14	1764
1284	*COL4A2*	1.16	2.0E-02	2297	1.15	1.3E-02	3951
9833	*KIAA0175*	1.69	0.0E+00	2652	1.87	1.0E-16	3951
92140	*MTDH*	1.26	1.1E-04	2592	1.32	3.9E-04	1764
51377	*UCH37*	1.15	2.1E-02	2576	1.41	1.5E-05	1764
51560	*RAB6B*	1.40	7.4E-08	2516	0.89	1.4E-01	1764
160897	*GPR180*	1.24	9.1E-03	1127	0.89	1.6E-01	1764
79888	*FLJ12443*	1.32	1.4E-05	2297	1.3	2.0E-06	3951
8293	*SERF1A*	1.16	4.2E-01	169	0.71	8.8E-10	3951
8476	*PK428*	1.10	1.2E-01	2652	1.2	9.9E-04	3951
10403	*HEC*	1.30	1.1E-05	2652	1.92	1.0E-16	3951
8833	*GMPS*	1.44	6.0E-09	2543	1.9	1.0E-16	3951
1894	*ECT2*	1.39	3.8E-08	2576	1.23	9.7E-03	1764
4318	*MMP9*	1.18	4.9E-03	2652	1.16	6.5E-03	3951
5019	*OXCT*	1.08	1.9E-01	2652	1.12	3.7E-02	3951
2781	*GNAZ*	1.14	3.4E-02	2592	0.91	7.5E-02	3951
2321	*FLT1*	1.02	8.4E-01	1721	0.94	4.0E-01	1764
2131	*EXT1*	1.13	3.9E-02	2652	1.05	4.3E-01	3951
56942	*DC13*	1.53	1.2E-12	2576	1.84	1.0E-16	3951
81624	*DIAPH3*	1.52	4.3E-10	2223	1.8	1.4E-13	1764
169714	*QSOX2*	1.18	3.8E-02	1130	0.87	6.8E-02	1764
286052	LOC286052	NA	NA	NA	1.04	6.5E-01	3951
51203	LOC51203	1.58	4.0E-14	2576	1.92	1.0E-16	3951
85453	*TSPYL5*	1.01	8.2E-01	2288	1.12	1.0E-16	3951

Table 6.7: Overexpression of *HLXB9* is a better prognostic indicator of relapse in breast cancer patients than a number of MammaPrint genes. Results from analysing overall survival in KM plotter and *BreastMark*.

The impact of *HLXB9* overexpression on patients overall survival was also compared to genes in the MammaPrint gene panel. *HLXB9* was the 14th most significant indicator of poor prognosis in KM plotter, 23rd in Breast Mark (as shown in table 8.2 in appendix). These finding indicates that *HLXB9* has more prognostic significance than a number of the other genes in the panel and would be a better indicator of disease progression and overall survival in breast cancer patients.

In summary, these results indicate that high expression of *HLXB9* was detrimental to breast cancer patients' outcome especially in terms of relapse in the original origin of breast tumour and overall survival. The results obtained was contrary to expectations as results from experiments on breast cancer cell lines suggested *HLXB9* functions as a tumour suppressor in breast cancer. Therefore, it was expected that high *HLXB9* expression would have a favourable impact on breast cancer patients' outcome. In comparison to genes from well-established multi gene expression panels Onco*type* DX and MammaPrint, high expression of *HLXB9* was a better indicator of poor prognosis (relapse and overall survival) than a number of the genes currently utilised. Therefore *HLXB9* could be added to the tests to improve their efficacy.

The effect of *HLXB9* expression on breast cancer patients' survival was then analysed by well-known prognostic factors in breast cancer such as tumour grade, lymph node status and tumour intrinsic subtype.

a) HLXB9 expression on breast cancer patients' survival by tumour grade

Histological tumour grade is also used as a prognostic marker in the treatment of breast tumours. It is a measure of tumour aggressiveness. There are three grades of breast tumours. Lower grade tumours are usually associated with better prognosis than higher grade tumours (Rakha et al. 2010; Weigelt et al. 2010). In this section, patients' survival was stratified by tumour grade in order to determine if *HLXB9* expression has a prognostic significance as tumour phenotype becomes more aggressive.

Overexpression of HLXB9 had a negative prognostic significance on relapse free survival in patients diagnosed with grade 2 tumours

HLXB9 expression had no effect on patient's outcome when patients' survival was stratified by tumour grade for grade 1 tumours in both databases. However, high expression of *HLXB9* had prognostic significance in patients diagnosed with grade 2 tumours. Grade 2 breast tumours are usually associated with an intermediate reoccurrence risk but have been shown to include patients with both low and high risk of reoccurrence by gene expression profiling (Sotiriou et al. 2006). In KM plotter, high expression of *HLXB9* reduced the probability of RFS in patients diagnosed with grade 2 tumours in both *BreastMark* and KM plotter HR: 1.61, P=0.001 and HR: 1.31, *P*=0.029 respectively. The DMFS in patients diagnosed with grade 2 breast tumours was also reduced by *HLXB9* overexpression in *BreastMark* only (HR: 1.52, *P*=0.019). In *BreastMark*, high expression of *HLXB9* also had a detrimental effect on the probability of overall survival of patients (HR: 1.67, *P*=0.01). Additionally in *BreastMark* only, overexpression of *HLXB9* was a negative prognostic indicator on RFS and DMFS in patients diagnosed with grade 3 tumours (HR: 1.27, *P*=0.031 and HR: 1.52, *P*=0.019).

The results obtained showed that high expression of *HLXB9* has negative prognostic significance in patients diagnosed with grade two tumours, although the impact on grade 3 tumours was inconclusive.

Tumour Grade	Patients' Outcome	KM plotter			*BreastMark*		
		n	Hazard Ratio	*P* value	n	Hazard Ratio	*P* value
Grade 1	Relapse Free Survival	345	1.09	0.75	202	0.47	0.0973
	Distant Metastasis Free Survival	188	0.67	0.34	234	0.57	0.1964
	Post Progression Survival	34	0.73	0.53	N/A	N/A	N/A
	Overall Survival	161	0.65	0.37	164	0.74	0.6463
Grade 2	Relapse Free Survival	901	1.31	0.029	577	1.61	0.00102
	Distant Metastasis Free Survival	546	1.55	0.014	811	1.29	0.06103
	Post Progression Survival	128	0.88	0.61	N/A	N/A	N/A
	Overall Survival	387	1.2	0.39	440	1.67	0.01273
Grade 3	Relapse Free Survival	903	1.27	0.031	526	1.19	0.2325
	Distant Metastasis Free Survival	458	1.52	0.019	701	0.98	0.8606
	Post Progression Survival	165	1.33	0.14	N/A	N/A	N/A
	Overall Survival	503	1.36	0.066	352	1.177	0.3946

Table 6.8: Overexpression of HLXB9 reduces the survival of breast cancer patients diagnosed with grade 2 tumours. High expression of *HLXB9* reduces the probability of the relapse free survival in patients diagnosed with grade 2 tumours (KM plotter), relapse free survival and overall survival (*BreastMark*) (Gyorffy et al. 2013; Madden et al. 2013). Highlighted results have hazard ratios that were statistically significant.

The prognostic significance of *HLXB9* overexpression on relapse (RFS) in patients diagnosed with grade two tumours was also compared to genes used in Onco*type* DX and MammaPrint gene tests. *HLXB9* was the 4th most significant indicator of poor prognosis in Onco*type* DX (*BreastMark*, 7th in KM plotter as shown in table 8.3 in appendix). *HLXB9* overexpression was a better indicator of poor prognosis than genes associated with invasion in the Onco*type* DX from both databases. When compared to genes that indicate poor prognosis in MammaPrint, overexpression of *HLXB9* was the 12th most significant indicator of poor prognosis (*BreastMark*, 22nd in KM plotter table 8.4 in appendix).

b) HLXB9 expression on breast cancer patients' survival by lymph node status

Lymph node status, which is when tumour cells have spread from the breast into the lymph nodes, is also used as a prognostic marker. The presence of cancer cells in the lymph node is usually associated with a poor prognosis as it indicates the tumour has spread and there is a potential of metastasis to other organs. In this section, patients' outcomes are stratified by lymph node involvement to determine if *HLXB9* expression has a prognostic significant by lymph node status.

HLXB9 overexpression reduced relapse free survival in lymph node positive patients

High expression of *HLXB9* reduced the probability of RFS in lymph node positive patients in both KM plotter (HR: 1.49, *P*=8.1E-05) and *BreastMark* (HR: 1.36, *P*=0.008). This indicates lymph node positive patients with high expression of *HLXB9* are more likely to relapse in the original site of the breast tumour

than patients with low expression of *HLXB9*. Overexpression of *HLXB9* was also found to be detrimental to patients with positive lymph nodes overall survival in KM plotter only (HR: 1.61, *P*=0.017).

In KM plotter, overexpression of *HLXB9* was unfavourable to relapse free survival (HR: 1.35, *P*=0.0005) and distant metastatic free survival (HR: 1.40, *P*=0.016) in lymph node negative patients. In *BreastMark*, the negative correlation of high expression of *HLXB9* to patients' survival seems to continue in terms of relapse free survival (HR: 1.21, *P*=0.076). However, this correlation was not statistically significant possibly due to a smaller sample size in this database (KM plotter n=2020, *BreastMark* n=1155).

Lymph Node Status	Patients' Outcome	KM plotter			*BreastMark*		
		n Number	Hazard Ratio	*P* value	n Number	Hazard Ratio	*P* value
Lymph Node Negative	Relapse Free Survival	2020	1.35	0.00047	1155	1.213	0.07577
	Distant Metastasis Free Survival	988	1.40	0.016	900	1.089	0.5687
	Post Progression Survival	165	1.22	0.37	N/A	N/A	N/A
	Overall Survival	594	1.2	0.34	545	1.338	0.135
Lymph Node Positive	Relapse Free Survival	1133	1.49	8.1E-05	719	1.356	0.0083
	Distant Metastasis Free Survival	382	1.27	0.22	749	1.047	0.7172
	Post Progression Survival	128	1.48	0.083	N/A	N/A	N/A
	Overall Survival	313	1.61	0.017	473	1.314	0.0676

Table 6.9: High expression of HLXB9 reduced the relapse free survival of breast cancer patients with lymph node involvement. High expression of *HLXB9* may also reduce on the relapse free survival and distant metastatic free survival of lymph node negative patients* (*KM Plotter only) (Gyorffy et al. 2013; Madden et al. 2013). Highlighted results have hazard ratios that were statistically significant.

The prognostic significance of *HLXB9* overexpression in lymph node positive patients was compared to genes used in both the Onco*type* DX and MammaPrint gene tests. *HLXB9* was the 6th most significant indicator of poor prognosis when compared to genes used in Onco*type* DX (KM plotter, 8th in *BreastMark*, as shown in table 8.5 in appendix). When compared to genes that indicate poor prognosis in MammaPrint, *HLXB9* was the 17th most significant indicator of prognosis (KM plotter, 19th in *BreastMark*, table 8.6 in appendix).

Currently, both Onco*type* DX and MammaPrint are used to determine prognosis in patients diagnosed with ER positive, lymph node negative tumours. Therefore, *HLXB9* expression as a prognostic indicator was compared to genes used in both test in patients diagnosed with ER positive, lymph node negative tumours. In comparison to genes used in the Onco*type* DX test, overexpression of *HLXB9* was the 7th most significant indicator of poor prognosis (KM plotter, 8th in *BreastMark*, table 8.7 in appendix). In comparison to genes used in the MammaPrint test, *HLXB9* ranked 22nd compared to the 43 genes that

indicate poor prognosis (KM plotter only). *HLXB9* had no significant impact in *BreastMark*. However, this might be due to KM plotter having a larger same size (n=1256) than *BreastMark* (n=875).

c) HLXB9 expression on breast cancer patients' prognosis by tumour intrinsic subtypes

Tumour intrinsic (molecular) subtypes are used as prognostic indicators and to inform treatment decisions in breast cancer. Patients' outcomes were analysed by tumour intrinsic subtypes to determine if *HLXB9* expression has a prognostic significance on particular tumour intrinsic subtypes (refer to table 6.10 for hazard ratios and *P* values, and the appendix A1-A4 for survival curves).

Overexpression of HLXB9 had a deleterious effect on the survival of patients diagnosed with luminal A molecular subtype breast tumours

In both databases, overexpression of *HLXB9* reduced the probability of RFS in patients diagnosed with luminal A tumours (KM plotter: HR=1.23, *P*=0.02; *BreastMark*: HR=1.24, *P*=0.04). In *BreastMark* only, overexpression of *HLXB9* also reduced the probability of DDFS (HR=1.46, *P*=0.03) and OS (HR=1.43, *P*=0.030). These results imply that *HLXB9* overexpression has a negative prognostic significance in patients diagnosed with luminal A subtype breast tumours particularly on RFS of these patients (as shown in figure A1 in appendix and table 6.10).

The prognostic significance of *HLXB9* overexpression in patients diagnosed with luminal A tumours was compared to genes used in both the Onco*type* DX and MammaPrint gene tests. *HLXB9* was the 7th most significant indicator of poor prognosis when compared with genes used in Onco*type* DX (KM plotter, 8th in *BreastMark*, as shown in table 8.9 in appendix). When compared to genes that indicate poor prognosis in MammaPrint, *HLXB9* was the 17th most significant indicator of prognosis (*BreastMark*, 21th in KM plotter, table 8.10 in appendix).

Although the negative prognostic significance of overexpression of *HLXB9* expression on patient survival was observed in the survival curves of patients diagnosed with luminal B subtype tumours, this observation did not reach statistical significance (as shown in figure A2 in appendix and table 6.10).

In addition to this, the trend of negative prognostic significance of *HLXB9* overexpression on survival of patients diagnosed with basal tumours was observed in the survival curves for RFS, DMFS, OS and PPS in both databases (as shown in figure A3 in appendix and table 6.10). However, the aforementioned trend

only reached statistical significance in the analysis of DFS, DMFS and OS in KM plotter (HR: 1:34, p=0.02, HR: 1.7, p=0.04 and HR: 1:85, p=0.02 respectively).

Additionally, the trend of negative prognostic significance of high *HLXB9* expression on survival of patients diagnosed with HER2 tumours was observed in the survival curves for RFS, DMFS and PPS in both databases (as shown in figure A4 in appendix and table 6.6). However, this trend did not reach statistical significance except in the analysis of OS in *BreastMark*. In this database, overexpression of *HLXB9* significantly reduced the probability of overall survival in breast cancer patients diagnosed with HER2 amplified tumours (HR=1.65, p=0.04).

When considering the effect of *HLXB9* expression on patients' survival by tumour intrinsic subtypes, it is worth noting that survival curves for all tumour intrinsic subtypes show a trend of negative correlation of *HLXB9* overexpression to patient survival (figure A1, A2, A3 and A4 in appendix). However, this trend was probably not statistically significant due to the small numbers of samples obtained due to categorising patients by tumour intrinsic subtypes. This is supported by the results obtained for patients diagnosed with luminal A tumours, analyses on these patients reached statistical significance as this tumour subtype had the largest sample size (as shown in table 6.10).

In conclusion, overexpression of *HLXB9* was found to reduce the probability of RFS in breast cancer patients diagnosed with luminal A tumours. Additionally, the possibility that *HLXB9* overexpression was detrimental to breast cancer patients diagnosed with other molecular subtypes could not be ruled out.

Tumour Intrinsic Subtypes	Patients' Outcome	KM plotter			*BreastMark*		
		n Number	Hazard Ratio	*P* value	n Number	Hazard Ratio	*P* value
Luminal A	Relapse Free Survival	1933	1.23	0.017	1035	1.24	0.04041
	Distant Metastasis Free Survival	965	1.27	0.099	723	1.46	0.03176
	Post Progression Survival	179	0.9	0.58	N/A	N/A	N/A
	Overall Survival	611	1.18	0.34	812	1.43	0.02976
Luminal B	Relapse Free Survival	1149	1.22	0.044	617	1.21	0.08502
	Distant Metastasis Free Survival	430	1.21	0.28	904	0.91	0.4404
	Post Progression Survival	134	1.05	0.83	N/A	N/A	N/A
	Overall Survival	433	1.20	0.33	659	1.23	0.1468
Basal	Relapse Free Survival	618	1.34	0.022	394	0.90	0.5033
	Distant Metastasis Free Survival	232	1.7	0.043	341	1.15	0.4924
	Post Progression Survival	64	1.76	0.062	N/A	N/A	N/A
	Overall Survival	241	1.85	0.016	273	0.96	0.8488
Her2 amplified	Relapse Free Survival	251	1.03	0.89	325	1.20	0.2512
	Distant Metastasis Free Survival	119	1.66	0.11	128	1.45	0.2517
	Post Progression Survival	37	1.37	0.53	N/A	N/A	N/A
	Overall Survival	117	1.37	0.34	200	1.65	0.04325

Table 6.10: Overexpression of HLXB9 reduces the relapse free survival of patients diagnosed with luminal A molecular subtype tumours. High expression of *HLXB9* reduces the probability of relapse free survival; distant

metastasis free survival and overall survival on breast cancer patients with diagnosed with luminal A tumours. It also reduces the probability of distance metastasis free survival in basal tumours** and overall survival of HER2 amplified tumours*. **BreastMark* only, **KM plotter (Gyorffy et al. 2013; Madden et al. 2013). Highlighted results have hazard ratios that were statistically significant.

In summary, *HLXB9* overexpression was found to be detrimental to breast cancer patient's survival. *HLXB9* overexpression was also found to be a poor prognostic indicator on survival in breast cancer patients diagnosed with either luminal A, lymph node positive or grade2 tumours.

In comparison to genes from well-established multi gene expression panel tests Onco*type* DX and MammaPrint, high expression of *HLXB9* was a better indicator of poor prognosis (relapse free survival) than a number of the genes currently utilised in patients diagnosed with luminal A, lymph node positive or grade 2 tumours. Therefore *HLXB9* could be added to the tests to improve their efficacy.

6.5. Discussion

HLXB9 was downregulated in breast cancer cell lines

HLXB9 expression was found in all control non-cancer cells and breast cancer cell lines analysed by RT-PCR and qRT-PCR. This result was not surprising as other homeobox genes have been found to be expressed and are known to play various roles in breast development (Chen & Sukumar 2003). The breast is also subject to many growth changes in a woman's lifetime therefore dysregulation of these homeobox genes have unsurprisingly been implicated in breast cancer development and progression (Chen & Sukumar 2003; Yamamoto et al. 2011; Tommasi et al. 2009; Lian et al. 2012).

As *HLXB9* is expressed by normal breast, changes to its expression might be detrimental in breast cancer. To investigate this hypothesis, *HLXB9* expression was compared in non-cancer breast cell and breast cancer cell lines by qRT-PCR. *HLXB9* expression was found in variable levels in the two non-cancer primary human mammary epithelial cells (HMEC) sampled (see figure 6. 5). The level of expression of *HLXB9* was also found to be potentially associated with proliferation in HMEC 184D suggesting that *HLXB9* plays a tumour suppressive role in breast proliferation.

The expression of *HLXB9* in the non-tumorigenic cell line MCF 10A and its tumorigenic derivatives: MCF 10A-TK1 and MCF 10A-Ca1h was also investigated. *HLXB9* expression decreased as the cells became more

tumorigenic, which further suggests *HLXB9* acts as a tumour suppressor in non-cancer breast cells. This hypothesis is supported by other studies that found *HLXB9* was hypermethylated in breast cancer samples compared to normal breast samples (Yamamoto et al. 2011; Tommasi et al. 2009; Lian et al. 2012;). In addition, two studies into transcription factors activity in breast cancer by Siletz and colleagues found downregulation of *HLXB9* contributes to breast cancer progression (Siletz et al. 2013; Siletz et al. 2013). In their first study on Epithelial-Mesenchymal Transition (EMT) in breast cancer, they found *HLXB9* was downregulated in all the models of EMT they studied (Siletz et al.2013). EMT is a process which facilitates cancer progression as breast epithelial cells take on more mesenchymal cell characteristics such as loss of polarity and cell to cell adhesion enabling breast cancer cells migration and invasion. In their other study on breast carcinoma associated fibroblasts (CAF) pro-oncogenic cells that promote proliferation, invasion and angiogenesis, they found compared to normal healthy breast stroma, their model of CAF showed downregulation of *HLXB9* compared to normal breast fibroblast (Siletz et al.2013).

Seven breast cancer cell lines were initially screened for *HLXB9* expression using RT-PCR, and all seven cell lines expressed *HLXB9*. Subsequently, qRT-PCR was carried out on three breast cancer cell lines: T47D (luminal A), BT474 (luminal B), and MDA-MB-231 (Basal). There was a decrease in expression of *HLXB9* in all three breast cancer cell lines analysed compared to the two of the three non-cancer cell/cell lines analysed (HMEC 184D and MCF 10A). This suggests that *HLXB9* expression was downregulated in breast cancer and that *HLXB9* functions as a tumour suppressor in breast cancer.

All breast cancer cell lines investigated also showed expression of the anti-sense non-coding RNA, *HLXB9-AS2*. The function of this anti-sense mRNA has not been identified, however, anti-sense RNA has been implicated in many biological processes, which include DNA methylation, gene inactivation, splicing, RNA stability (Pelechano and Steinmetz 2013). This presence of *HLXB9-AS2* RNA might explain the downregulation of *HLXB9* observed in the breast cancer cell lines investigated.

Although differential levels of gene expression is useful and is a characteristic of cancers, differential mRNA expression does not always correlate to differential protein levels of the genes, therefore HLXB9 protein expression was analysed. HLXB9 protein expression in the breast cancer cell lines was analysed by indirect immunofluorescence (IF). HLXB9 was found to be downregulated in the breast cancer cell lines analysed compared to the MCF 10A cell line (control non-tumorigenic cell line). This finding was in line

with a previous study by Neufing et al. 2003 who found a decrease in the intensity of HLXB9 nuclei staining in tumour cells compared to matched control cells (Neufing et al. 2003). In addition, breast cancer cell lines showed nuclear protein aggregates of HLXB9 suggesting a functional role for HLXB9 in these cells. The nuclear protein aggregates observed in this cell line was similar to the protein aggregates of phosphorylated HLXB9 observed in insulinomas (Desai et al. 2015). In insulinomas, phosphorylated HLXB9 has an anti-apoptotic function, which contributes to cancer development and progression.

There was a direct correlation in the level of *HLXB9* mRNA expression and HLXB9 protein expression in two of the cell lines investigated (MCF 10A and BT474). In these cell lines, the MCF 10A cell line had the highest amount of mRNA expression, which corresponded to high protein expression. The BT474 cell line had the least amount of *HLXB9* mRNA and protein expression. However, the T47D and MDA-MB-231 cell lines had similar levels of *HLXB9* mRNA but different levels of protein expression with the T47D cell lines showing considerably more HLXB9 protein than the MDA-MB-231 cell line. This finding further indicates the possibility of an interaction between HLXB9 and estrogen in this cell line.

In summary, both non-cancer breast cells and the breast cancer cell lines analysed express HLXB9. Some of the cell lines also expressed the anti-sense non-coding RNA. These results indicate a biological function for this gene in breast cell, which is dysregulated in cancer. As there seem to be variability in the expression of HLXB9 between healthy individuals, a better approach might be to compare the level of expression of HLXB9 between matched non-cancer and tumour sample in breast cancer patients. The expression of HLXB9 in patient's samples could then be characterised to evaluate the link between expression and patient prognosis. Downregulation of HLXB9 was also found in the breast cancer cell lines analysed, suggesting a loss of the role of HLXB9 as a tumour suppressor in breast cancer.

Nuclear localisation of HLXB9 in breast cancer

The nuclear localisation of *HLXB9* in non-cancer cells: HMEC 184D, HMEC 240L, MCF 10A cell line, and breast cancer cell lines: T47D, BT474 and MDA-MB-231 was analysed using a 2D FISH assay. In order to increase the reliability of our observations, no fewer than 120 nuclei were analysed per sample.

The non-cancer HMEC cells and the MCF 10A cell line had two copies of the *HLXB9* gene localised in the nuclear periphery in these cells. There was no significant difference in the localisation of *HLXB9* between

these cells suggesting a conservation of gene positioning between healthy individuals. These results were in line with results from other studies that found no difference in gene positioning in healthy tissue of various individuals (Wiech et al. 2005; Meaburn et al. 2009).

The T47D cell line (luminal A) had two copies of the *HLXB9* gene; *HLXB9* had a peripheral intermediate positioning in these cells. There was a reduction of gene signals found in the nuclear periphery in the cell line compared to the controls; however, this difference was significant compared to the MCF 10A cell line only. These results sugget a small population of T47D cells show an altered nuclear localisation of *HLXB9* compared to non-cancer cells.

The BT474 cell line (luminal B) had five to seven copies of the *HLXB9* gene. The *HLXB9* gene had an intermediate positioning in this cell line. Compared to the non-cancer cells analysed, there was a significant decrease in the proportion of gene signals in the nuclear periphery with a concomitant increase in proportion of gene signals in the nuclear interior (HMEC 240L and MCF 10A only). These results indicate a change in the nuclear localisation of *HLXB9* from the nuclear periphery to the nuclear interior in this cell line.

The MDA-MB-231 cell line (basal) showed variability in the number of *HLXB9* alleles with two to five alleles observed. This variability in the number of alleles found is not unusual for this cell line as a study by Watson and colleagues identified three different strains of the MDA-MB-231 cell line grown in culture by their differing karyotype (Watson et al. 2004). *HLXB9* had an intermediate nuclear positioning in this cell line. Compared to the non-cancer cells, there was a decrease in the proportion of gene signal in the nuclear periphery in this cell line compared to the three non-cancer cells and an increase in the proportion of gene signal in the nuclear interior compared to HMEC 240L and MCF 10A only. These results also indicate a change in the nuclear localisation of *HLXB9* from the nuclear periphery to the nuclear interior in this cell line.

As a proportion of the MDA-MB-231 cell nuclei had two alleles, this cell line is an ideal model to determine if a change in the number of alleles from the two copies affects the position of the *HLXB9* gene. Unfortunately only seventeen nuclei with two copies of *HLXB9* were present in our experiment and therefore was not a significantly representative sample size for further analysis.

In summary, these results showed a change in the nuclear localisation of *HLXB9* in the two of the three breast cancer cell lines investigated. A previous study by Meaburn and colleagues showed a disease specific nuclear reorganisation in breast cancer (Meaburn et al. 2009; Meaburn & Misteli 2008). Disease specific nuclear reorganisation of genes has also been shown in prostate cancer (Leshner et al. 2016). However, the possibility the observed change in nuclear localisation of *HLXB9* in the BT474 and MDA-MB-231 cell lines being a result of additional copies of *HLXB9* could not be ruled out in this study.

HLXB9 nuclear localisation and its expression

HLXB9 expression has been shown to be affected by its nuclear localisation. For instance, *HLXB9* expression has been linked to its expression during neuronal cell differentiation and in leukaemia patients (Ballabio et al. 2009, Leotta et al. 2014). Consequently, the nuclear localisation of *HLXB9* was analysed in breast cancer to determine if a change in localisation was associated with gene expression. In the control cells, *HLXB9* was localised in the nuclear periphery and this localisation was associated with active gene transcription (see figures 4 and 11). This finding indicates a breast specific method of regulating *HLXB9* expression in breast cells as previous studies had only shown the silencing of *HLXB9* occurring at the nuclear periphery (Ballabio et al. 2009, Leotta et al. 2014). Although, there was a variation in levels of *HLXB9* expression between our controls samples, there was no significant difference in the localisation of *HLXB9* between these samples. This suggests a conserved localisation of *HLXB9* in healthy individuals. This result also indicates the nuclear localisation of HLXB9 cannot be solely responsible for its regulation.

In the breast cancer cell lines changes in the nuclear localisation of *HLXB9* could not be linked to its expression. This suggests *HLXB9* expression in breast cells is regulated by other mechanisms and is not dependent on nuclear positioning.

Overexpression of HLXB9 is a negative prognostic indicator of breast cancer patient's survival

HLXB9 overexpression was found to have a detrimental impact on breast cancer patients' overall survival and relapse free survival. These findings suggest that high *HLXB9* expression could be useful in stratifying patients for treatment at diagnosis. Stratifying patients in this manner is likely to be beneficial to patients as one of the main aims of breast cancer research is identifying characteristics that could be used to judge

patients' treatment options i.e. to prevent over treatment of patients leading to unwanted side effects or under treatment of patients leading to relapse and metastasis.

Additionally, in comparison to genes from well-established multi gene expression panels Onco*type* DX and MammaPrint, high expression of *HLXB9*, was a better indicator of poor prognosis (relapse and overall survival) than a number of the genes currently utilised. Therefore, *HLXB9* could be added to the tests to improve their efficacy.

Patient outcome was analysed by tumour grade to determine when *HLXB9* expression has a prognostic significance during tumour progression. High expression of *HLXB9* in patients with grade 2 tumours reduced the patients' disease free survival. This is important as grade 2 tumours are characterised by cancer cells that are aggressive, and if left unchecked, these cells can proceed to grade 3 tumours.

High expression of *HLXB9* in lymph node positive patients (results from both databases) showed that high expression of *HLXB9* had an adverse effect on relapse free survival indicating patients with high expression of *HLXB9* are likely to relapse in the original site of the breast tumour.

Patients' outcome was also analysed by tumour intrinsic subtype to determine if *HLXB9* expression had a prognostic significance dependant on tumour subtype. High expression of *HLXB9* in patients diagnosed with luminal A subtype had a detrimental effect on relapse free survival in both databases. In *BreastMark*, this unfavourable effect also extends to distant disease free survival and overall survival in luminal A tumours. This finding suggests that high expression of *HLXB9* might have an effect on estrogen receptors, a well known mediator of breast cancer growth. This result was not surprising as HLXB9 has been shown to interact with menin (MEN 1), a known binding partner of estrogen receptors in insulinomas (Shi et al. 2013). However, HLXB9 must also interact with other proteins to mediate breast cancer as evidenced by its negative prognostic significance in basal and HER2 amplified breast tumours. High expression of *HLXB9* had negative effect of distant metastasis free survival of patients diagnosed with basal tumours (KM plotter only) and overall survival of patients diagnosed with HER2 amplified tumours. Although a negative correlation was observed between high expression of *HLXB9* and prognosis (as shown in the survival curve) in the other tumour intrinsic subtypes, the trend did not reach statistical significance probably due to low sample numbers.

Additionally, in comparison to genes from well-established multi gene expression panel tests Onco*type* DX and MammaPrint, high expression of *HLXB9*, was a better indicator of poor prognosis (relapse free survival) than a number of the genes currently utilised in patients diagnosed with luminal A, lymph node positive or grade 2 tumours. Therefore, *HLXB9* could be added to the tests to improve their efficacy.

As downregulation of HLXB9 was observed in the breast cancer cell lines analysed, it was expected that high expression of *HLXB9* would have a favourable impact on breast cancer patient's prognosis. However, the results obtained suggest that high expression of *HLXB9* was detrimental to patients' survival. In previous research by Shi and colleagues, overexpression of HLXB9 in insulinomas was shown to inhibit the normal apoptotic function of HLXB9 in pancreatic b-cells (Shi et al. 2013; Desai et al. 2014; Desai et al. 2015). This suggests that high expression of *HLXB9* in breast cancer patients might have an anti-apoptotic function, which might be responsible for the poor prognosis observed in patients with high expression of *HLXB9*. It is also worth noting the expression of the anti-sense mRNA *HLXB9-AS2* observed in breast cancer cell lines was not observed in primary mammary epithelial cells HMEC 184D and 240L. The *HLXB9-AS2* mRNA might be responsible for the down regulation of *HLXB9* expression observed in breast cancer cell lines. Therefore, it is worth investigating the prognostic significance of *HLXB9-AS2* gene expression on breast cancer patients' outcome.

Analysing the effect of mRNA expression on patient survival retroactively is a widely used approach in identifying candidate biomarkers as differential genes expression can be analysed in a large number of patients' samples. Although this method of identifying biomarkers shows promising results for *HLXB9*, it is worth noting there are limitations to the methodology. These limitations include the possibility of the introduction of artefacts due to different laboratory protocols. Additionally, the use of median gene expression as a filter to differentiate expression might result in the concealment of significant populations with outlier expression values. From this study, *HLXB9* overexpression was found to be deleterious to patients' outcome suggesting *HLXB9* overexpression is worth further investigation in the laboratory.

6.6. Conclusion

In this study, *HLXB9* overexpression was shown to have a detrimental effect on breast cancer patients' outcomes. Additionally, *HLXB9* overexpression was a better prognostic indicator than a number of genes currently utilised in the Onco*type* DX and MammaPrint gene expression test and could be added to them

improve their efficacy *HLXB9* expression was also found in normal non-cancer breast samples suggesting a tumour suppressive role in normal breast cells. *HLXB9* expression was found to be downregulated in the breast cancer cell lines analysed, which further supports the role of *HLXB9* as a tumour suppressor in breast cancer. *HLXB9* was found localised at the nuclear periphery of non-cancer breast cells where the gene is transcriptionally active. This is a new finding showing *HLXB9* transcriptionally active in the nuclear periphery. This study also found that a change in the nuclear localisation of *HLXB9* cannot be the sole mechanism for regulating *HLXB9* expression. As both downregulation and upregulation of *HLXB9* was analysed in this study, both states of dysregulation are worth further investigation in order to identify transcriptional targets of *HLXB9*. Hopefully this will shed some light into the function of *HLXB9* in both breast cancer development and progression. Although the potential of *HLXB9* as a biomarker in breast cancer is appealing, further work is needed to confirm the real value of this biological parameter in the diagnosis and prognosis of breast cancer.

.

7. General Discussion

One of the main aims of cancer research is the identification and subsequent validation of biomarkers. Clinically, molecular biomarkers in cancer are routinely used to: classify risk for predisposition, diagnosis, monitoring disease status (including remittance and relapse), inform prognosis or as targets for therapy (Henry and Hayes 2012; Burke 2016). In leukaemia, specific chromosomal abnormalities are used as a biomarkers and are the most important indicators of patients' prognosis (Betz and Hess 2010; Ferrara and Schiffer 2013; Estey et al. 2014). For example, deletions in the long arm of chromosomes 5 and 7 are associated with a poor prognosis in both acute myeloid leukaemia (AML) and the myelodyspstic syndromes (MDS) (van der Straaten et al. 2005).

In breast cancer factors such as tumour grade, lymph node involvement, metastasis and tumour molecular subtypes that are used together as biomarkers. The identification of tumour molecular subtypes has been particularly significant in improving patients' outcome in breast cancer as they provided new targets for therapy (Parker et al. 2009; Eroles et al. 2012; Tang et al. 2016). The basal molecular subtype of breast tumours best exemplifies the importance of identifying genetic biomarkers, as there are currently no genetic targets for therapy. Therefore, the standard course of treatment such as surgery, radiotherapy, immunotherapy, and chemotherapy are the only options available.

In this project, *HLXB9*, a homeobox gene aberrantly expressed in a number of cancer types, was investigated to determine if it could be used as biomarker in cancer.

A preliminary finding from this project is the possibility that alteration in *HLXB9* expression is an early event in cancer development. *HLXB9* expression was more frequently occurring in patients diagnosed with pre-leukaemic disorders states such as myeloproliferative disorders (MPD) and myelodysplastic syndrome (MDS). 44% of patients diagnosed with MPD and MDS showed *HLXB9* expression compared to 6% of AML patients (P=0.02). Additionally, downregulation in HLXB9 expression was found to occur early in breast cancer development (results obtained from MCF 10A and its derivatives). The first stage of the MCF 10A cancer progression model, pre malignant - MCF 10A TK1 showed a significant reduction in *HLXB9* expression compared to the non-tumorigenic MCF 10A. This downregulation was then maintained in the malignant model MCF 10A –CA1h. This project was not the first to identify early deregulation of *HLXB9* in cancer. In fact, hypermethylation of *HLXB9* has been previously reported to be an early occurrence in breast cancer (Tommasi et al. 2009; Lian et al. 2012). Another tumour type showing early

HLXB9 deregulation is testicular cancer, where *HLXB9* overexpression is observed in pre-invasive testicular carcinoma in situ (Almstrup et al. 2005; Almstrup et al. 2007; Novotny et al. 2007). Therefore the results obtained in this project, in conjunction with findings from other studies, suggest that deregulation of *HLXB9* might have a potential as a biomarker of pre malignancy.

Other results obtained from this project are discussed in the different chapters of this thesis and can be summarised as follows.

7.1. Detection of t(7;12)(q36;p13) in paediatric leukaemia using dual colour fluorescence in situ hybridisation (FISH)

One of the defining characteristics of a biomarker is that it can be accurately detected. It has been proven that consistent overexpression of *HLXB9* is the consequence of a chromosomal rearrangement namely the t(7;12)(q36;p13) found in some infant leukaemia patients. In the first chapter of this thesis, a new probe set for the detection of the translocation-t(7;12)(q36;q13) was described. The probe set was validated in leukaemia patients using fluorescence *in situ* hybridisation as part of this thesis work. This approach for the detection of this translocation is advantageous over existing approaches because it only requires a fluorescent microscope equipped with three fluorescence filter sets, which is commonly available in laboratories. The accurate detection of this translocation is important for clinicians as its presence confers a very poor prognosis for patients.

7.2. Investigating the expression and the nuclear localisation of *HLXB9* in patients diagnosed with haematological malignancies

The second study of this project continued the investigation into *HLXB9* as a biomarker in leukaemia. In this study, *HLXB9* expression and its nuclear localisation was investigated in a different patient cohort rather than the classic paediatric patients cohort harbouring t(7;12)(q36;p13) known to overexpress *HLXB9*. *HLXB9* expression was investigated in 48 patients diagnosed with various haematological malignancies. 25% of the patients investigated showed *HLXB9* expression. Patients chosen for this study included patients with normal copies of chromosome 7 and patients with chromosome 7 abnormalities (with a particular focus on patients harbouring interstitial deletions of 7q).

Patients with interstitial deletion of chromosome 7 were chosen primarily because of the possibility of a similar mechanism for ectopic expression of *HLXB9* observed in translocation t(7;12) (q36;p13) patients. This is because similar to a translocation event, an interstitial deletion of chromosome 7 results in a disruption of the genomic region proximal to *HLXB9* and the rejoining of that region with another part of the genome. We speculated this event could also result in the repositioning of the *HLXB9* gene in the nucleus of the leukaemia cells harbouring the rearrangement, as this is seen in t(7;12) cases.

In this study, there was no significant difference in *HLXB9* expression in both groups of patients. *HLXB9* was expressed in 37.5% of patients with normal chromosome 7 and 20% of patients with chromosome 7 abnormalities (P=0.290). To date, there has only been one other study that identified the expression of *HLXB9* in a patient not harbouring a chromosome 7 abnormality. In that study, *HLXB9* expression was found in patient with karyotype 45–46,XY,–14,+mar[cp2]/46,XY (pt 10)(Ballabio et al. 2009). Therefore, this study confirms that *HLXB9* expression in leukaemia is not limited to patients harbouring chromosome 7 abnormalities.

Although *HLXB9* was expressed in patients, the functional significance of *HLXB9* expression in these patients remains unknown. Additionally, the prognostic significance of *HLXB9* expression in these patients could not be determined as there was no clinical information on patients' outcome. It would have been interesting to compare the clinical outcome of patients showing *HLXB9* expression in this study to paediatric patients with t(7;12) to determine if a similar level of risk is observed in both groups.

The nuclear localisation of *HLXB9* was also investigated in both patients without chromosome 7 abnormalities expressing *HLXB9* and patients harbouring interstitial deletion of chromosome 7 in order to determine if its nuclear localisation is linked with expression. A majority of patients from both groups showed no difference in nuclear localisation from the control bone marrow cells investigated although there was a variation in gene expression. These results suggest the nuclear localisation of *HLXB9* is not the mechanism of regulation of its expression in patients diagnosed with haematological disorders.

7.3. Investigating the differential expression of *HLXB9* as a potential cancer biomarker by datamining online transcriptomics databases

In this study on investigating *HLXB9* expression as a cancer biomarker, the expression of *HLXB9* was investigated in four publicly available databases. Differential expression of *HLXB9* was identified in cancers not previously associated with its aberrant expression; these were lung cancer, kidney cancer and endometrial cancer, which could provide new avenues for research. The aberrant expression of *HLXB9* in kidney cancer was particularly interesting as it was identified in three subtypes of kidney cancer. *HLXB9* was upregulated in two kidney cancer subtypes. These were kidney clear cell carcinoma (KIRC) and kidney papillary cell carcinomas (KIRP), which are associated with poor prognosis. In contrast, *HLXB9* was downregulated in kidney chromophobe (KICH), a kidney cancer subtype with better prognosis.

Another key finding from this study was the differential expression of *HLXB9* that was observed in cancers from organs/tissue solely or partially derived from the endodermal germ layer such as pancreas, rectum, lung, liver, bladder, stomach and oesophagus. This finding indicates *HLXB9* expression has to be correctly regulated in endodermal tissue derivatives after foetal development to maintain health.

However, this study highlights the limitation of the datamining approach for the identification of new biomarkers. One of the limitations encountered was the small number of samples available or lack of data on particular cancer types. Additionally, although *HLXB9* expression was identified in subpopulations of cancer type, further investigation into the clinical characteristics unifying these patients was not possible. There was also a lack of data on the patients' outcome so the prognostic significance of identified aberrant expression of *HLXB9* could not be determined.

7.4. Investigating *HLXB9* as a biomarker in breast cancer

In the final study on investigating *HLXB9* as a biomarker in breast cancer, HLXB9 expression was found to be downregulated in breast cancer cell lines by qRT-PCR and IF. This finding was in line with previous reports from other studies (Neufing et al. 2003; Tommasi et al. 2009; Lian et al. 2012; Siletz et al. 2013b; Siletz et al. 2013a). From the IF experiments, foci of *HLXB9* protein aggregates were

observed in the nucleus of breast cancer cell lines investigated suggesting a function for *HLXB9* in these cell lines, similar foci of HLXB9 protein aggregates have been previously observed in insulinomas (Desai et al. 2015). In insulinomas, these aggregates have been confirmed to be primarily composed of phosphorylated *HLXB9*, which has an anti-apoptotic function. However further investigation is required to confirm the identity and function of the protein aggregates observed in *HLXB9* in breast cancer.

The nuclear localisation of *HLXB9* in breast cancer cell lines was also investigated. *HLXB9* was localised in the nuclear periphery in control cells, however, the breast cancer cell lines showed a localisation of *HLXB9* away from the nuclear periphery. This change in the nuclear localisation of *HLXB9* was particularly significant in the MDA-MB-231 and BT474 cell line. However, the repositioning of *HLXB9* observed was not associated with gene expression thereby suggesting that *HLXB9* expression was not regulated by a position effect mechanism.

Finally, the prognostic significance of differential expression of *HLXB9* was investigated by datamining breast cancer survival databases. Contrary to expectation, *HLXB9* overexpression was found to be detrimental to patients relapse free and overall survival. *HLXB9* overexpression was also a negative prognostic indicator in patients diagnosed with lymph node positive, grade 2 or luminal A tumours. Additionally, the prognostic significance of *HLXB9* expression was compared to genes from well-established multi gene expression panels Onco*type* DX and MammaPrint. Overexpression of *HLXB9* was a better indicator of poor prognosis (relapse and overall survival) than a number of the genes currently utilised. Therefore, *HLXB9* could be added to the tests to improve their efficacy.

Overexpression of *HLXB9* has been previously been reported to result in an increase in the phosphorylation of HLXB9 in insulinomas (Desai et al.2015). In insulinomas, phosphorylation of HLXB9 has been found to inhibit apoptosis by downregulating the expression of Cblb, a ubiquitin ligase which would normally target C-met, a tyrosine kinase for degradation (Desai et al 2015). Further work is required to determine the functional consequence of *HLXB9* overexpression in breast cancer patients.

In conclusion, these studies show the appeal of *HLXB9* as a biomarker in cancer. However, further investigation is needed to determine the prognostic significance and the clinical characteristics of cancer patients who show aberrant expression of *HLXB9*. Therefore, further studies will be crucial to understand the consequence of aberrant *HLXB9* expression in cancer. Additionally, studies to identify the function of *HLXB9* in cancer cells is vital in order to provide further information on the suitability of *HLXB9* as a biomarker in cancer.

Future Work

The following future work is required to fully understand the role of *HLXB9* expression in both non-cancer states and cancer cells to determine its potential as a cancer biomarker.

1. Chromatin immunoprecipitation (ChIP) experiments should be carried out to identify *HLXB9* targets genes in non-cancer and cancer samples.
2. Experiments using RNA silencing (siRNA) could be conducted to investigate the function of *HLXB9* in breast cancer and leukaemia. By silencing the expression of *HLXB9* in normal breast or blood cells, the function of this gene in normal cells could be determined. These experiments might also help understand how the downregulation of HLXB9 expression might contribute to breast cancer development.
3. Additionally, experiments using phospho-HLXB9 antibodies to determine the composition of the HLXB9 aggregates observed in breast cancer cell lines would help determine the function of HLXB9 in breast cancer cells.
4. In this study, overexpression of *HLXB9* in breast cancer has also been found to be detrimental to patient outcome. Previous studies have also identified HLXB9 overexpression in a subset of leukaemia patients with poor prognosis. Therefore overexpressing *HLXB9* in breast cancer or leukeamia cell lines could also help understand how this overexpression of this gene contributes to cancer development and/or progression. In addition, siRNA experiments to knock down HLXB9 expression in a cell line overexpressing HLXB9 would be useful to determine if a gain of apoptotic function occurs in cell lines.
5. As the suggested experiments are carried out on cell lines, which are known to be able to evolve in vitro introducing artefacts not seen in patients, primary breast cancer patients'

samples could be analysed in order to fully understand the dysregulation of *HLXB9* in breast cancer.

6. Comparing breast cancer chromatin (where *HLXB9* located at the nuclear periphery is transcriptionally active) to leukaemia cells chromatin (where *HLXB9* is located at the periphery is transcriptionally silent) could help with understanding the mechanism of *HLXB9* expression. This could be done by staining for open or closed chromatin markers.
7. Finally, expression analysis of the *HLXB9-AS2* gene in breast cancer cells should be carried out to determine the function of this non-coding RNA. In order to eliminate the possibility of this gene's detection as an artefact of the reverse transcription cDNA synthesis reaction, strand specific RT-PCR or cDNA synthesis in the presence of actinomycin D could be carried out. If the expression of *HLXB9-AS2* is confirmed in breast cancer cell lines, the function of *HLXB9-AS2* in non-cancer breast and breast cancer can then be further investigated. Furthermore, comparing the impact of *HLXB9* and *HLXB9-AS2* gene expression on patient survival would be used to determine if there is a correlation between their expression and patient survival.

All of the proposed future work will provide valuable insight into how *HLXB9* is expressed, the impact of this expression in cancer and determine if this gene could be used as a biomarker in cancer.

Chapter 8. Appendix

Entrez ID	Gene symbol	*BreastMark*			Km plotter		
		HR	***P***	**N**	**HR**	***P***	**N**
3110	*HLXB9*	1.42	5.1E-05	2091	1.52	1.0E-04	1402
Onco*type* DX gene panel							
4288	*KI67*	1.77	5.2E-11	2091	1.46	4.7E-04	1402
6790	*STK15*	1.79	3.1E-11	2091	1.81	5.5E-08	1402
332	*Survivin*	2.02	7.8E-16	2091	1.67	2.4E-06	1402
891	*CCNB1*	1.93	5.6E-12	1796	1.51	9.6E-03	626
4605	*MYBL2*	1.79	1.3E-11	2091	1.98	5.5E-10	1402
4320	*MMP11*	1.35	5.3E-04	2091	1.14	2.4E-01	1402
1515	*CTSL2*	1.58	1.1E-07	2091	1.67	2.9E-06	1402
2886	*GRB7*	1.49	3.0E-06	2091	1.27	3.0E-02	1402
2064	*HER2*	1.01	9.0E-01	2091	0.97	8.3E-01	626
2099	*ER*	0.69	1.9E-05	2091	0.65	8.0E-05	1402
5241	*PGR*	0.61	3.4E-08	2091	0.62	2.7E-03	626
596	*BCL2*	0.64	3.4E-07	2091	0.56	1.0E-07	1402
57758	*SCUBE2*	0.58	5.6E-06	1285	0.58	5.0E-07	1402
2944	*GSTM1*	0.62	7.6E-07	1778	0.60	2.9E-06	1402
968	*CD68*	1.19	4.7E-02	2091	1.00	1.0E+00	1402
573	*BAG1*	0.86	8.2E-02	2091	0.87	3.8E-01	1402

Table 8.1: Overexpression of *HLXB9* is a better indicator of overall survival in breast cancer patients than a number of genes used in the Onco*type* DX gene panel. Results from analysing overall survival (OS) in KM plotter and *BreastMark*.

Entrez ID	Gene Symbol	*BreastMark*			KM plotter		
		HR	***P***	**N**	**HR**	***P***	**N**
3110	*HLXB9*	1.42	5.1E-05	2091	1.52	1.2E-04	1402
Good Prognosis							
8659	*ALDH4*	1.13	1.7E-01	1963	0.91	3.9E-01	1402
8817	*FGF18*	0.85	5.5E-02	2091	0.98	8.9E-01	626
27113	*BBC3*	1.12	2.5E-01	1796	1.04	7.0E-01	1402
57593	*EBF4*	0.56	4.8E-01	48	0.96	8.0E-01	626
57758	*SCUBE2*	0.58	5.6E-06	1285	0.58	5.0E-07	1402
146923	*RUNDC1*	0.80	1.5E-01	584	0.48	5.8E-06	626
8840	*WISP1*	0.80	1.1E-02	2082	0.77	9.6E-02	626
2947	*GSTM3*	0.60	2.8E-09	2091	0.51	2.2E-05	626
151126	*ZNF385B*	0.64	4.3E-03	575	0.48	5.2E-06	626
146760	*RTN4RL1*	1.30	1.4E-01	429	0.68	1.8E-02	626
10455	*PECI*	0.74	1.7E-03	1778	1.19	1.1E-01	1402
7043	*TGFB3*	0.65	4.8E-07	2091	0.69	6.5E-04	1402
55351	*HSA250839*	0.59	2.7E-08	1933	0.69	5.7E-04	1402
58475	*MS4A7*	0.60	7.2E-05	879	0.62	3.1E-03	626
163	*AP2B1*	0.93	3.8E-01	2091	0.82	6.5E-02	1402
79132	*LGP2*	1.04	7.2E-01	1638	0.85	1.5E-01	1402
Poor Prognosis							
55321	*C20ORF46*	1.43	9.6E-05	1933	1.26	3.5E-02	1402
11082	*ESM1*	1.39	1.5E-04	2043	1.41	1.5E-03	1402
9134	*CCNE2*	1.81	4.0E-10	1796	1.83	3.0E-08	1402
54583	*EGLN1*	1.37	1.6E-03	1629	1.22	2.1E-01	626
1058	*CENPA*	1.86	1.4E-12	2091	1.7	1.3E-06	1402
9055	*PRC1*	2.12	5.6E-16	1933	1.97	5.9E-10	1402
445815	*AKAP2*	1.14	1.5E-01	1924	0.78	2.6E-02	1402
10874	*NMU*	1.55	3.6E-07	2091	1.55	5.1E-05	1402
3488	*IGFBP5*	1.08	3.8E-01	2091	1.32	8.7E-02	626
10531	*MP1*	1.27	3.2E-02	1443	0.7	3.0E-02	626
57110	*HRASLS*	1.43	7.3E-05	1933	1.5	1.9E-04	1402
8577	*TMEFF1*	1.15	1.3E-01	1936	1.16	1.7E-01	1402
4175	*MCM6*	1.81	1.3E-11	2091	1.09	5.7E-01	626
643008	*SMIM5*	NA	NA	NA	1.33	7.5E-02	626
83879	*CDCA7*	1.05	7.8E-01	584	1.35	6.3E-02	626
5984	*RFC4*	1.90	2.5E-13	2091	1.45	6.2E-04	1402
23594	*ORC6L*	1.70	4.3E-09	1933	1.75	3.5E-07	1402

6515	*SLC2A3*	1.02	8.6E-01	2091	0.97	7.9E-01	1402
57211	*GPR126*	1.08	4.0E-01	1796	1.44	9.2E-04	1402
79791	*FBXO31*	1.14	1.5E-01	1933	1.76	4.0E-04	626
1633	*DCK*	1.34	7.2E-04	2091	1.3	1.5E-02	1402
51514	*L2DTL*	1.64	5.8E-08	1933	1.21	2.3E-01	626
1284	*COL4A2*	1.23	2.7E-02	1796	1.21	8.6E-02	1402
9833	*KIAA0175*	2.13	0.0E+00	2091	2.04	9.2E-11	1402
92140	*MTDH*	1.55	5.6E-07	2091	1.54	6.8E-03	626
51377	*UCH37*	1.36	7.9E-04	1933	1.13	4.4E-01	626
51560	*RAB6B*	1.73	7.3E-09	1933	1.21	2.4E-01	626
160897	*GPR180*	1.99	3.3E-05	533	1.17	3.3E-01	626
79888	*FLJ12443*	1.59	5.8E-07	1796	1.28	2.5E-02	1402
8293	*SERF1A*	0.94	8.8E-01	183	0.83	9.5E-02	1402
8476	*PK428*	1.10	2.9E-01	2091	1.24	4.9E-02	1402
10403	*HEC*	1.61	4.3E-08	2091	1.56	4.8E-05	1402
8833	*GMPS*	1.96	7.2E-14	1908	1.63	9.4E-06	1402
1894	*ECT2*	1.60	3.0E-07	1933	1.09	6.0E-01	626
4318	*MMP9*	1.06	5.1E-01	2091	1.08	5.0E-01	1402
5019	*OXCT*	1.17	6.7E-02	2091	1.05	6.8E-01	1402
2781	*GNAZ*	1.23	1.8E-02	2091	1.04	7.1E-01	1402
2321	*FLT1*	1.29	4.4E-03	1888	1.14	4.1E-01	626
2131	*EXT1*	1.36	4.1E-04	2091	0.89	3.0E-01	1402
56942	*DC13*	1.92	1.3E-12	1933	1.73	5.0E-07	1402
81624	*DIAPH3*	1.70	2.8E-05	1580	1.53	7.8E-03	626
169714	*QSOX2*	1.66	1.5E-03	536	1.13	4.5E-01	626
286052	*LOC286052*	NA	NA	NA	1.18	3.1E-01	626
51203	*LOC51203*	2.12	5.6E-16	1933	2.14	6.0E-12	1402
85453	*TSPYL5*	1.04	6.8E-01	1787	1.12	3.0E-01	1402

Table 8.2: Overexpression of *HLXB9* is a better indicator of overall survival in breast cancer patients than a number of genes used in the MammaPrint gene panel. Results from analysing overall survival in KM plotter and *BreastMark*.

Entrez Gene ID	Gene symbol	BreastMark			Kmplotter		
		HR	***P***	**N**	**HR**	***P***	**N**
3110	*HLXB9*	1.61	1.0E-03	577	1.31	2.9E-02	901
Onco*type* DX gene panel							
4288	*KI67*	1.40	1.7E-02	616	1.56	2.9E-04	901
6790	*STK15*	1.51	3.1E-03	616	1.55	4.0E-04	901
332	*Survivin*	1.75	5.7E-05	616	1.75	6.2E-06	901
891	*CCNB1*	1.89	4.4E-06	616	1.85	2.0E-02	227
4605	*MYBL2*	1.39	2.0E-02	616	1.55	4.0E-04	901
4320	*MMP11*	1.28	9.3E-02	577	1.01	9.5E-01	901
1515	*CTSL2*	1.46	9.0E-03	577	1.27	4.8E-02	901
2886	*GRB7*	1.89	4.6E-06	616	1.21	1.2E-01	901
2064	*HER2*	1.07	6.6E-01	577	1.63	6.4E-02	227
2099	*ER*	0.83	2.0E-01	577	0.68	1.9E-03	901
5241	*PGR*	0.72	1.8E-02	616	0.45	2.4E-03	227
596	*BCL2*	0.71	1.7E-02	577	0.68	1.7E-03	901
57758	*SCUBE2*	0.58	9.4E-04	433	0.65	5.0E-04	901
2944	*GSTM1*	0.78	1.8E-01	322	0.86	2.3E-01	901
968	*CD68*	1.16	2.8E-01	616	1.11	4.1E-01	901
573	*BAG1*	1.00	9.9E-01	616	0.86	5.5E-01	227

Table 8.3: Overexpression of *HLXB9* is a better indicator of relapse in breast cancer patients in patients diagnosed with grade 2 tumours than a number of genes used in the Onco*type* DX gene panel. Results from analysing overall survival in KMplotter and *BreastMark*.

Entrez Gene ID	Gene symbol	*BreastMark*			Kmplotter		
		HR	***P***	**n**	**HR**	***P***	**n**
3110	*HLXB9*	1.61	1.0E-04	577	1.31	2.9E-02	901
Good Prognosis							
8659	*ALDH4*	1.12	4.7E-01	522	1.03	8.4E-01	901
8817	*FGF18*	1.03	8.6E-01	616	1.16	5.6E-01	227
27113	*BBC3*	1.13	4.1E-01	577	1.14	2.9E-01	901
57593	*KIAA1442*	NA	NA	NA	0.99	9.8E-01	227
57758	*CEGP1*	0.58	9.4E-04	433	0.65	5.0E-04	901
146923	*RUNDC1*	0.71	7.6E-02	338	0.69	1.5E-01	227
8840	*WISP1*	0.80	1.1E-01	614	0.57	3.6E-02	227
2947	*GSTM3*	0.89	4.0E-01	616	0.79	3.7E-01	227
151126	*ZNF533*	0.69	5.6E-02	336	0.64	9.2E-02	227
146760	*RTN4RL1*	1.19	5.2E-01	205	0.79	3.6E-01	227
10455	*PECI*	0.75	6.9E-02	493	0.86	2.1E-01	901
7043	*TGFB3*	0.94	6.5E-01	577	0.78	3.9E-02	901
55351	*HSA250839*	0.73	3.5E-02	548	0.74	1.6E-02	901
58475	*CFFM4*	0.76	2.0E-01	299	0.89	6.5E-01	227
163	*AP2B1*	1.21	1.9E-01	577	0.84	1.4E-01	901
79132	*LGP2*	0.98	8.7E-01	587	1.04	7.3E-01	901
Poor prognosis							
55321	*C20ORF46*	1.08	6.2E-01	587	1.10	4.5E-01	901
11082	*ESM1*	1.35	4.1E-02	564	1.24	7.9E-02	901
9134	*CCNE2*	1.99	6.8E-07	616	1.32	2.5E-02	901
54583	*EGLN1*	1.43	1.3E-02	585	1.75	3.2E-02	227
1058	*CENPA*	1.49	4.7E-03	616	1.59	2.0E-04	901
9055	*PRC1*	1.72	1.3E-04	587	1.53	6.0E-04	901
445815	*AKAP2*	1.03	8.5E-01	414	1.10	4.4E-01	901
10874	*NMU*	1.62	6.0E-04	616	1.42	4.4E-03	901
3488	*IGFBP5*	1.16	3.2E-01	577	1.37	2.3E-01	227
10531	*MP1*	1.29	9.9E-02	501	1.08	7.6E-01	227
57110	*LOC57110*	1.54	3.5E-03	548	1.51	8.0E-04	901
8577	*TMEFF1*	1.02	9.1E-01	483	0.84	1.6E-01	901
4175	*MCM6*	2.01	3.9E-07	616	1.30	3.1E-01	227
643008	*LOC643008*	NA	NA	NA	1.17	5.4E-01	227
83879	*CDCA7*	1.13	5.3E-01	338	1.41	1.9E-01	227
5984	*RFC4*	2.02	3.2E-07	616	1.46	2.2E-03	901
23594	*ORC6L*	1.46	8.1E-03	587	1.44	3.2E-03	901
6515	*SLC2A3*	0.91	5.1E-01	577	0.94	6.1E-01	901
57211	*DKFZP564D0462*	1.15	3.5E-01	577	1.22	1.1E-01	901
79791	*FBXO31*	1.06	6.8E-01	587	1.44	1.6E-01	227
1633	*DCK*	1.22	1.7E-01	577	0.84	1.7E-01	901
51514	*L2DTL*	1.93	7.7E-06	548	1.33	2.7E-01	227
1284	*COL4A2*	1.55	2.9E-03	577	1.31	2.6E-02	901
9833	*KIAA0175*	2.09	1.1E-07	616	1.82	1.2E-06	901
92140	*MTDH*	1.41	1.7E-02	577	1.19	5.1E-01	227
51377	*UCH37*	1.10	5.1E-01	587	1.39	2.0E-01	227
51560	*RAB6B*	1.06	7.2E-01	548	1.23	4.2E-01	227
160897	*GPR180*	1.42	1.1E-01	278	0.81	4.1E-01	227
79888	*FLJ12443*	1.82	3.5E-05	577	1.09	4.8E-01	901
8293	*SERF1A*	0.87	7.9E-01	39	1.00	9.8E-01	901
8476	*PK428*	1.20	2.0E-01	616	1.00	9.8E-01	901
10403	*HEC*	1.27	9.7E-02	616	1.60	1.4E-04	901
8833	*GMPS*	1.69	1.6E-04	616	1.30	3.7E-02	901
1894	*ECT2*	1.72	1.2E-04	587	0.85	5.3E-01	227
4318	*MMP9*	1.08	5.9E-01	616	1.41	5.3E-03	901
5019	*OXCT*	1.41	1.5E-02	616	1.06	6.5E-01	901
2781	*GNAZ*	1.03	8.3E-01	577	1.12	3.6E-01	901
2321	*FLT1*	1.06	7.8E-01	299	1.58	7.8E-02	227
2131	*EXT1*	1.19	2.1E-01	616	0.93	5.3E-01	901
56942	*DC13*	1.52	3.2E-03	587	1.36	1.2E-02	901
81624	*DIAPH3*	1.23	2.3E-01	472	1.46	1.5E-01	227

169714	*QSOX2*	1.56	3.8E-02	286	1.40	2.0E-01	227
286052	*LOC286052*	NA	NA	NA	0.99	9.6E-01	227
51203	*LOC51203*	1.80	3.3E-05	587	1.61	1.1E-04	901
85453	*TSPYL5*	1.05	7.4E-01	575	1.04	7.6E-01	901

Table 8.4: Overexpression of *HLXB9* is a better indicator of relapse in breast cancer patients in patients diagnosed with grade 2 tumours than a number of genes used in the MammaPrint gene panel. Results from analysing overall survival in KM plotter and *BreastMark*.

Entrez ID	Gene symbol	BreastMark			KMplotter		
		HR	*P*	N	HR	*P*	N
3110	*HLXB9*	1.36	8.3E-03	719	1.49	8.10E-05	1133
Onco*type* DX gene panel							
4288	*KI67*	1.47	6.7E-04	744	1.36	2.3E-03	1133
6790	*STK15*	1.52	2.1E-04	744	1.71	1.0E-07	1133
332	*Survivin*	1.74	1.5E-06	744	1.66	4.1E-07	1133
891	*CCNB1*	1.49	1.4E-03	600	1.73	1.8E-05	724
4605	*MYBL2*	1.40	2.9E-03	744	1.65	7.0E-07	1133
4320	*MMP11*	1.50	4.4E-04	719	1.24	3.4E-02	1133
1515	*CTSL2*	1.78	7.4E-07	719	1.64	6.9E-07	1133
2886	*GRB7*	1.15	2.3E-01	744	0.99	9.4E-01	1133
2064	*HER2*	0.99	9.3E-01	719	0.92	5.1E-01	724
2099	*ER*	0.72	4.0E-03	719	0.6	3.6E-07	1133
5241	*PGR*	0.65	1.1E-04	744	0.5	6.6E-08	724
596	*BCL2*	0.87	2.4E-01	719	0.61	7.1E-07	1133
57758	*SCUBE2*	0.54	6.9E-04	370	0.53	1.4E-10	1133
2944	*GSTM1*	0.62	1.3E-03	385	0.65	1.4E-05	1133
968	*CD68*	1.28	3.1E-02	744	1.38	1.3E-03	1133
573	*BAG1*	0.99	9.4E-01	744	0.8	8.5E-02	724

Table 8.5: Overexpression of *HLXB9* is a better indicator of relapse in breast cancer patients diagnosed with lymph node positive tumours than a number of genes used in the Onco*type* DX gene panel. Results from analysing overall survival in KMplotter and *BreastMark*.

Entrez Gene ID	Gene symbol	BreastMark			KMPlotter		
		HR	*P*	N	HR	*P*	N
3110	*HLXB9*	1.36	8.3E-03	719	1.49	8.1E-05	1133
MammaPrint gene signature							
Good Prognosis							
8659	*ALDH4*	1.09	4.9E-01	672	1.06	5.4E-01	1133
8817	*FGF18*	0.99	9.3E-01	744	0.94	6.3E-01	724
27113	*BBC3*	1.10	4.4E-01	575	1.11	3.2E-01	1133
57593	*KIAA1442*	0.72	5.9E-01	32	0.70	5.8E-03	724
57758	*CEGP1*	0.54	6.9E-04	370	0.53	1.4E-10	1133
146923	*RUNDC1*	0.71	1.2E-02	489	0.59	3.6E-05	724
8840	*WISP1*	0.88	2.6E-01	740	0.82	1.2E-01	724
2947	*GSTM3*	0.79	3.8E-02	744	0.77	4.4E-02	724
151126	*ZNF533*	0.78	6.0E-02	485	0.66	1.4E-03	724
146760	*RTN4RL1*	1.13	4.2E-01	392	0.69	3.3E-03	724
10455	*PECI*	0.86	2.2E-01	642	1.12	2.7E-01	1133
7043	*TGFB3*	1.03	7.8E-01	719	0.80	2.3E-02	1133
55351	*HSA250839*	0.71	4.5E-03	689	0.73	1.9E-03	1133
58475	*CFFM4*	0.66	7.6E-04	608	0.63	3.4E-04	724
163	*AP2B1*	1.30	2.3E-02	719	0.77	1.0E-02	1133
79132	*LGP2*	1.05	7.2E-01	570	0.77	7.6E-03	1133
Poor prognosis							
55321	*C20orf46*	1.06	6.4E-01	714	1.32	5.4E-03	1133
11082	*ESM1*	1.39	5.0E-03	687	1.58	5.3E-06	1133
9134	*CCNE2*	1.39	8.4E-03	600	1.82	2.6E-09	1133
54583	*EGLN1*	1.41	7.1E-03	566	1.21	1.4E-01	724

1058	*CENPA*	1.51	3.0E-04	744	1.73	5.8E-08	1133
9055	*PRC1*	1.80	3.6E-07	714	1.91	1.1E-10	1133
445815	*AKAP2*	1.13	3.8E-01	453	0.86	1.3E-01	1133
10874	*NMU*	1.57	7.4E-05	744	1.38	1.3E-03	1133
3488	*IGFBP5*	1.16	1.9E-01	719	1.55	7.7E-04	724
10531	*MP1*	1.19	3.1E-01	425	1.17	2.2E-01	724
57110	*LOC57110*	1.27	4.3E-02	689	1.42	4.5E-04	1133
8577	*TMEFF1*	1.16	2.3E-01	647	1.09	3.7E-01	1133
4175	*MCM6*	1.61	3.2E-05	744	1.40	8.4E-03	724
643008	*LOC643008*	N/A	N/A	N/A	1.01	9.1E-01	724
83879	*CDCA7*	1.25	9.8E-02	489	1.96	1.4E-07	724
5984	*RFC4*	1.62	2.1E-05	744	1.53	2.3E-05	1133
23594	*ORC6L*	1.74	1.9E-06	714	1.60	2.8E-06	1133
6515	*SLC2A3*	1.15	2.2E-01	719	1.03	7.6E-01	1133
57211	*DKFZP564D0462*	1.14	3.0E-01	575	1.48	8.4E-05	1133
79791	*FBXO31*	1.02	8.8E-01	714	1.65	9.7E-05	724
1633	*DCK*	1.17	1.8E-01	719	1.06	5.9E-01	1133
51514	*L2DTL*	1.82	6.8E-07	689	1.58	3.8E-04	724
1284	*COL4A2*	1.25	8.3E-02	575	1.30	8.2E-03	1133
9833	*KIAA0175*	1.76	8.5E-07	744	1.74	3.2E-08	1133
92140	*MTDH*	1.29	2.7E-02	719	1.41	7.0E-03	724
51377	*UCH37*	1.13	3.0E-01	714	1.27	6.5E-02	724
51560	*RAB6B*	1.22	8.4E-02	689	1.45	3.8E-03	724
160897	*GPR180*	1.51	4.0E-03	436	1.36	1.7E-02	724
79888	*FLJ12443*	1.48	2.3E-03	575	1.19	8.8E-02	1133
8293	*SERF1A*	1.19	7.8E-01	25	0.73	1.7E-03	1133
8476	*PK428*	1.18	1.5E-01	744	1.15	1.6E-01	1133
10403	*HEC*	1.28	2.9E-02	744	1.81	3.8E-09	1133
8833	*GMPS*	1.74	1.1E-06	744	1.83	2.1E-09	1133
1894	*ECT2*	1.46	1.0E-03	714	1.31	3.2E-02	724
4318	*MMP9*	1.27	3.4E-02	744	1.21	5.7E-02	1133
5019	*OXCT*	1.14	2.4E-01	744	1.09	3.9E-01	1133
2781	*GNAZ*	1.17	1.8E-01	719	1.24	3.1E-02	1133
2321	*FLT1*	1.18	2.7E-01	358	0.97	8.4E-01	724
2131	*EXT1*	1.18	1.5E-01	744	0.87	1.7E-01	1133
56942	*DC13*	1.49	5.3E-04	714	1.64	7.8E-07	1133
81624	*DIAPH3*	1.52	4.1E-03	539	1.62	1.7E-04	724
169714	*QSOX2*	1.33	4.8E-02	432	1.21	1.4E-01	724
286052	*LOC286052*	N/A	N/A	N/A	1.27	6.2E-02	724
51203	*LOC51203*	1.76	1.4E-06	714	1.81	3.3E-09	1133
85453	*TSPYL5*	0.98	8.5E-01	571	1.14	1.9E-01	1133

Table 8.6: Overexpression of HLXB9 is a better indicator of relapse in breast cancer patients diagnosed with lymph node positive tumours than a number of genes used in the MammaPrint gene panel. Results from analysing overall survival in KM plotter and *BreastMark*.

Entrez Gene ID	Gene symbol	BreastMark			KMPlotter		
		HR	*P*	n	HR	*P*	n
3110	*HLXB9*	1.16	2.5E-01	875	1.27	3.3E-02	1256
Oncotype DX gene panel							
4288	*KI67*	1.68	4.4E-05	902	1.77	4.3E-07	1256
6790	*STK15*	2.32	4.0E-11	902	2.15	1.3E-11	1256
332	*Survivin*	1.96	8.6E-08	902	2.10	7.6E-11	1256
891	*CCNB1*	1.89	3.6E-06	793	1.71	5.0E-02	296
4605	*MYBL2*	1.76	8.0E-06	902	1.93	6.0E-09	1256
4320	*MMP11*	1.55	1.0E-03	875	1.19	1.2E-01	1256
1515	*CTSL2*	1.42	7.1E-03	875	1.55	8.3E-05	1256
2886	*GRB7*	1.26	7.0E-02	902	0.85	1.5E-01	1256
2064	*HER2*	1.03	8.3E-01	875	0.93	7.9E-01	296
2099	*ER*	1.32	5.0E-02	875	1.02	8.9E-01	1256
5241	*PGR*	0.80	8.0E-02	902	1.06	8.4E-01	296

596	*BCL2*	0.75	3.0E-02	875	0.71	2.1E-03	1256
57758	*SCUBE2*	0.71	3.0E-02	628	0.79	3.2E-02	1256
2944	*GSTM1*	0.92	5.6E-01	651	0.69	9.4E-04	1256
968	*CD68*	0.96	7.4E-01	902	0.95	6.6E-01	1256
573	*BAG1*	1.01	9.1E-01	902	1.10	7.2E-01	296

Table 8.7: Overexpression of *HLXB9* is a better indicator of relapse in breast cancer patients diagnosed with ER positive, lymph node negative tumours than a number of genes used in the Onco*type* DX gene panel.
Results from analysing overall survival in KM plotter and *BreastMark*.

Entrez Gene ID	Gene symbol	BreastMark			Kmplotter		
		HR	*P*	N	HR	*P*	N
3110	*HLXB9*	1.27	3.3E-02	1256	1.16	2.5E-01	875
Good Prognosis							
8659	ALDH4	0.87	2.0E-01	1256	0.93	6.0E-01	827
8817	FGF18	0.80	4.1E-01	296	0.79	7.0E-02	902
27113	BBC3	0.82	7.4E-02	1256	0.74	4.0E-02	766
57593	KIAA1442	0.65	1.2E-01	1256	NA	NA	NA
57758	CEGP1	0.79	3.2E-02	1256	0.71	2.5E-02	628
146923	RUNDC1	0.82	4.4E-01	296	0.60	3.4E-02	288
8840	WISP1	0.94	8.1E-01	296	0.80	8.0E-02	898
2947	GSTM3	0.92	7.5E-01	296	0.82	1.1E-01	902
151126	ZNF533	0.77	3.2E-01	296	1.13	6.1E-01	284
146760	RTN4RL1	0.88	6.3E-01	1256	0.76	3.4E-01	186
10455	PECI	0.95	6.5E-01	1256	0.86	2.6E-01	796
7043	TGFB3	0.73	4.7E-03	1256	0.83	1.5E-01	875
55351	HSA250839	0.78	2.9E-02	1256	0.82	1.4E-01	844
58475	CFFM4	0.69	1.6E-01	296	0.65	3.4E-02	370
163	AP2B1	0.90	3.6E-01	1256	0.83	1.6E-01	875
79132	LGP2	0.89	2.8E-01	1256	0.75	4.8E-02	762
Poor Prognosis							
55321	C20orf46	1.16	1.8E-01	1256	1.01	9.5E-01	871
11082	ESM1	1.08	4.9E-01	1256	1.29	5.0E-02	866
9134	CCNE2	1.74	8.6E-07	1256	1.74	6.5E-05	793
54583	EGLN1	1.58	8.2E-02	296	1.39	1.7E-02	758
1058	CENPA	1.83	9.3E-08	1256	1.97	9.4E-08	902
9055	PRC1	2.19	5.7E-12	1256	2.04	1.8E-08	871
445815	AKAP2	0.83	9.4E-02	1256	0.88	3.5E-01	722
10874	NMU	1.66	6.5E-06	1256	1.52	1.0E-03	902
3488	IGFBP5	1.33	2.8E-01	296	1.05	7.4E-01	875
10531	MP1	1.17	5.6E-01	296	1.18	2.5E-01	686
57110	LOC57110	1.32	1.2E-02	1256	1.49	2.2E-03	844
8577	TMEFF1	1.09	4.2E-01	1256	1.30	4.9E-02	800
4175	MCM6	1.61	7.8E-02	296	1.76	6.9E-06	902
643008	LOC643008	0.88	6.3E-01	1256	NA	NA	NA
83879	CDCA7	1.27	3.7E-01	1256	1.33	2.3E-01	288
5984	RFC4	1.91	9.4E-09	1256	1.71	2.2E-05	902
23594	ORC6L	1.41	2.2E-03	1256	1.82	2.5E-06	871
6515	SLC2A3	0.82	7.9E-02	1256	1.01	9.3E-01	875
57211	DKFZP56D0462	0.92	4.4E-01	1256	0.87	3.3E-01	766
79791	FBXO31	0.97	9.1E-01	296	0.85	2.3E-01	871
1633	DCK	1.17	1.7E-01	1256	1.55	8.0E-04	875
51514	L2DTL	1.69	5.5E-02	296	1.97	2.0E-07	844
1284	COL4A2	1.10	3.9E-01	1256	1.12	4.4E-01	766
9833	KIAA0175	2.19	6.1E-12	1256	1.84	1.2E-06	902
92140	MTDH	1.06	8.3E-01	296	1.22	1.3E-01	875
51377	UCH37	1.52	1.2E-01	296	1.16	2.5E-01	871
51560	RAB6B	0.84	5.2E-01	296	0.81	1.5E-01	844

160897	GPR180	1.03	9.2E-01	296	1.20	4.9E-01	248
79888	FLJ12443	1.30	1.9E-02	1256	1.18	2.5E-01	766
8293	SERF1A	0.94	6.1E-01	1256	1.66	3.9E-01	27
8476	PK428	1.09	4.4E-01	1256	1.26	7.3E-02	902
10403	HEC	2.07	1.3E-10	1256	1.40	8.3E-03	902
8833	GMPS	1.90	2.4E-08	1256	1.40	8.8E-03	902
1894	ECT2	1.25	3.9E-01	296	1.72	2.1E-05	871
4318	MMP9	1.38	4.4E-03	1256	1.23	1.1E-01	902
5019	OXCT	1.07	5.2E-01	1256	1.00	9.8E-01	902
2781	GNAZ	1.38	4.1E-03	1256	1.09	5.0E-01	875
2321	FLT1	1.07	7.9E-01	296	0.92	5.8E-01	646
2131	EXT1	0.98	8.6E-01	1256	1.16	2.6E-01	902
56942	DC13	1.70	2.6E-06	1256	1.81	2.9E-06	871
81624	DIAPH3	1.90	1.7E-02	296	1.09	5.3E-01	764
169714	QSOX2	1.31	3.0E-01	296	1.54	9.6E-02	252
286052	LOC286052	1.06	8.3E-01	1256	NA	NA	NA
51203	LOC51203	1.94	4.1E-09	1256	1.92	2.2E-07	871
85453	TSPYL5	0.99	9.1E-01	1256	0.93	6.0E-01	762

Table 8.8: Overexpression of *HLXB9* is a better indicator of relapse in breast cancer patients diagnosed with ER positive, lymph node negative tumours than a number of genes used in the MammaPrint gene panel. Results from analysing overall survival in KM plotter and *BreastMark*.

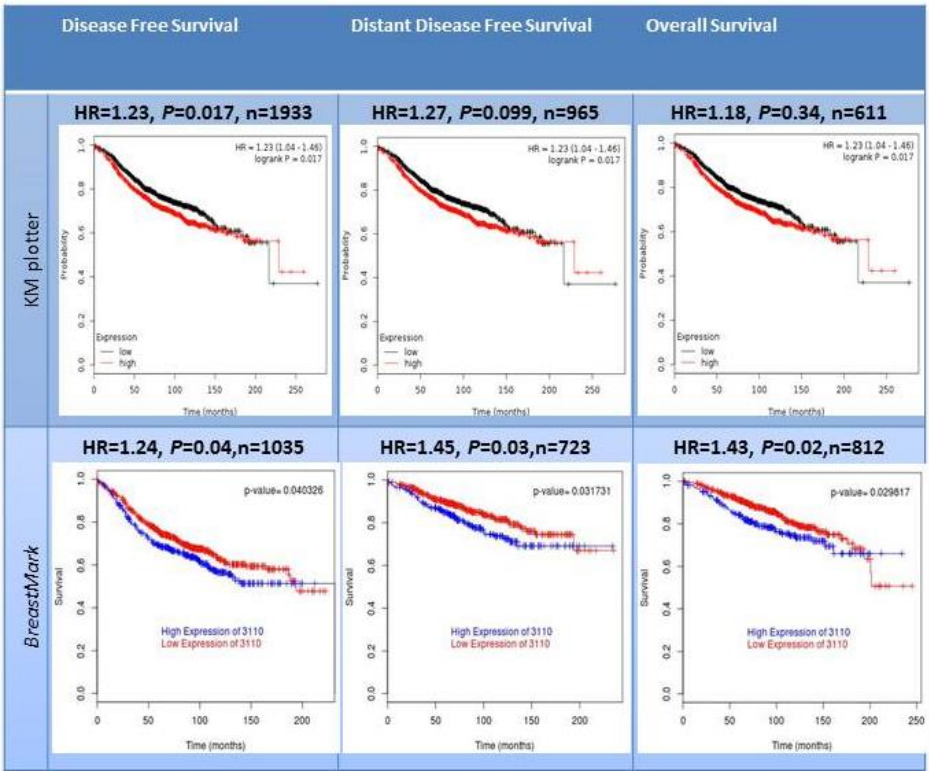

Figure A1: Overexpression of HLXB9 reduces luminal A patients' survival. High expression of *HLXB9* reduces the probability of relapse free survival *(KM plotter and BreastMark),* distant metastasis free survival and overall survival (*BreastMark only)* on breast cancer patients with diagnosed with luminal A breast tumours.

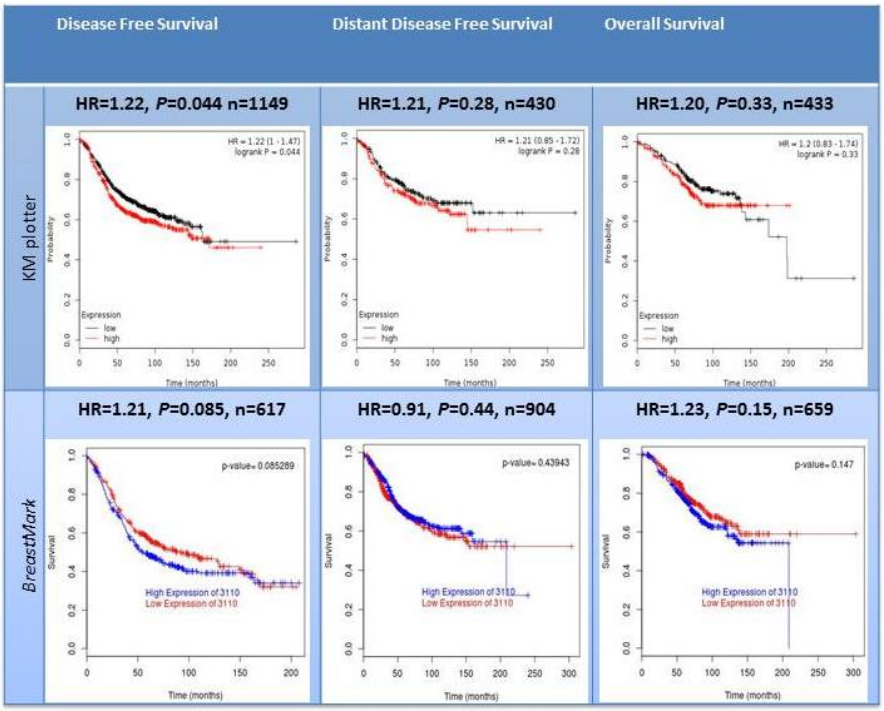

Figure A2: Overexpression of *HLXB9* expression reduces relapse free survival in breast cancer patients diagnosed with luminal B tumors (KM plotter only).

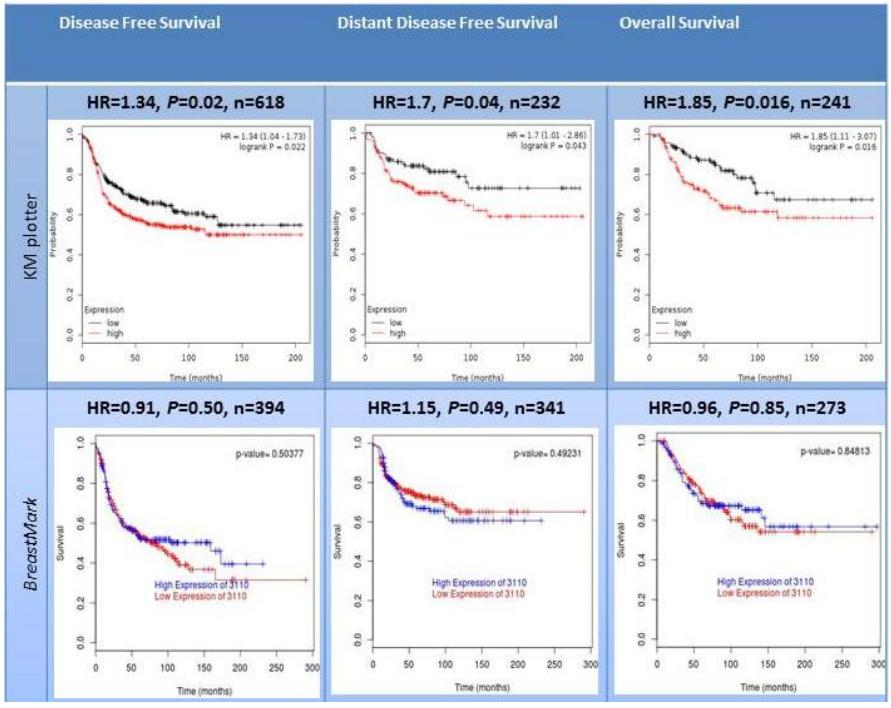

Figure A3: Overexpression of HLXB9 reduces the probability of disease free, distance metastasis free and overall survival in patients diagnosed with basal tumours *(KM plotter only).*

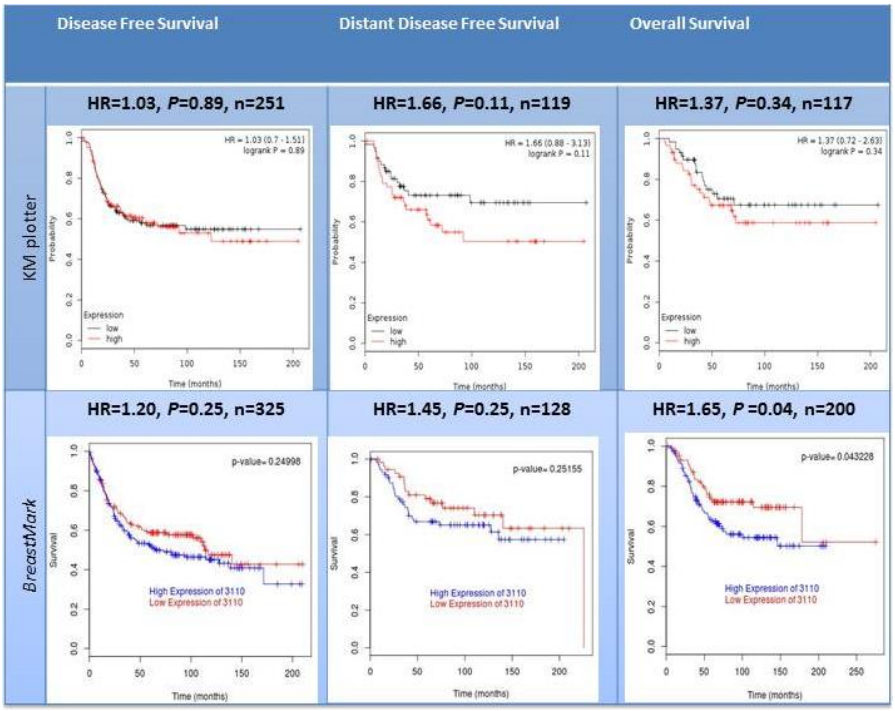

***Figure A4: Overexpression of HLXB9 reduces the* overall survival of breast cancer patients diagnosed with HER2 amplified tumours (BreastMark only)**

Entrez ID	Gene symbol	BreastMark			Kmplotter		
		HR	***P***	**N**	**HR**	***P***	**N**
3110	*HLXB9*	1.24	4.0E-02	1035	1.23	1.7E-02	1933
Onco*type* DX gene panel							
4288	*KI67*	1.32	7.9E-03	1063	1.5	3.9E-06	1933
6790	*STK15*	1.70	3.4E-07	1063	2.05	5.6E-16	1933
332	*Survivin*	1.68	5.9E-07	1063	1.76	1.8E-10	1933
891	*CCNB1*	1.49	4.6E-04	904	2.08	1.1E-08	841
4605	*MYBL2*	1.63	4.0E-06	1063	1.66	7.4E-09	1933
4320	*MMP11*	1.13	2.3E-01	1035	1.13	1.5E-01	1933
1515	*CTSL2*	1.59	1.6E-05	1035	1.41	6.9E-05	1933
2886	*GRB7*	1.34	4.8E-03	1063	0.71	8.7E-05	1933
2064	*HER2*	0.97	7.9E-01	1035	0.84	1.7E-01	841
2099	*ER*	1.17	2.3E-01	1035	0.83	3.4E-02	1933
5241	*PGR*	0.92	4.1E-01	1063	0.54	9.4E-07	841
596	*BCL2*	0.89	2.8E-01	1035	0.7	5.2E-05	1933
57758	*SCUBE2*	1.05	7.6E-01	712	0.58	4.8E-10	1933
2944	*GSTM1*	0.79	7.1E-02	738	0.62	4.0E-08	1933
968	*CD68*	1.15	1.9E-01	1063	0.87	1.0E-01	1933
573	*BAG1*	1.05	6.4E-01	1063	0.77	3.7E-02	841

Table 8.9: Overexpression of *HLXB9* is a better indicator of relapse in breast cancer patients in patients diagnosed with luminal A tumours than a number of genes used in the Onco*type* DX gene panel. Results from analysing overall survival in KMplotter and *BreastMark*.

Entrez Gene ID	Gene symbol	BreastMark			Kmplotter		
		HR	*P*	N	HR	*P*	N
3110	*HLXB9*	1.24	4.0E-02	1035	1.23	1.7E-02	1933
Good Prognosis							
8659	ALDH4	1.10	3.5E-01	1015	0.94	4.5E-01	1933
8817	FGF18	0.91	3.5E-01	1063	0.58	2.7E-05	841
27113	BBC3	1.27	3.3E-02	901	0.89	1.6E-01	1933
57593	KIAA1442	0.95	8.9E-01	46	0.6	5.8E-05	841
57758	CEGP1	1.05	7.6E-01	712	0.58	4.8E-10	1933
146923	RUNDC1	1.16	4.0E-01	451	0.57	8.6E-06	841
8840	WISP1	0.93	5.0E-01	1061	0.7	4.0E-03	841
2947	GSTM3	0.87	1.7E-01	1063	0.78	4.4E-02	841
151126	ZNF533	1.04	8.1E-01	449	0.69	3.2E-03	841
146760	RTN4RL1	1.02	9.2E-01	375	0.69	3.3E-03	841
10455	PECI	0.81	5.4E-02	983	0.83	3.1E-02	1933
7043	TGFB3	0.90	3.3E-01	1035	0.63	1.3E-07	1933
55351	HSA250839	0.84	1.0E-01	1003	0.68	1.4E-05	1933
58475	CFFM4	0.76	3.6E-02	557	0.68	2.6E-03	841
163	AP2B1	1.00	9.9E-01	1035	0.81	1.4E-02	1933
79132	LGP2	1.17	1.6E-01	897	1.00	9.6E-01	1933
Poor Prognosis							
55321	C20orf46	1.09	4.5E-01	1031	0.89	2.0E-01	1933
11082	ESM1	1.04	7.0E-01	1014	1.02	8.4E-01	1933
9134	CCNE2	1.58	2.9E-05	929	1.94	4.3E-14	1933
54583	EGLN1	1.08	5.3E-01	895	1.16	2.3E-01	841
1058	CENPA	1.49	1.7E-04	1063	1.56	3.4E-07	1933
9055	PRC1	1.81	1.6E-08	1031	2.28	<1E-16	1933
445815	AKAP2	1.11	4.1E-01	784	0.9	2.0E-01	1933
10874	NMU	1.45	4.3E-04	1063	1.63	2.2E-08	1933
3488	IGFBP5	1.30	1.3E-02	1035	1.26	6.3E-02	841
10531	MP1	1.06	6.5E-01	797	1.42	4.8E-03	841
57110	LOC57110	1.09	4.3E-01	1003	1.43	3.5E-05	1933
8577	TMEFF1	0.87	2.2E-01	987	1.08	3.6E-01	1933
4175	MCM6	1.63	2.3E-06	1063	1.65	7.7E-05	841
643008	LOC643008	NA	NA	NA	0.88	3.1E-01	841
83879	CDCA7	1.05	7.5E-01	451	1.18	1.9E-01	841
5984	RFC4	1.39	1.8E-03	1063	1.71	6.9E-10	1933
23594	ORC6L	1.66	1.6E-06	1031	1.53	1.1E-06	1933
6515	SLC2A3	0.94	5.7E-01	1035	0.86	9.0E-02	1933
57211	DKFZP56D 0462	0.84	1.6E-01	901	1.07	4.5E-01	1933
79791	FBXO31	1.00	9.7E-01	1031	1.09	5.0E-01	841
1633	DCK	1.21	7.6E-02	1035	1.48	5.6E-06	1933
51514	L2DTL	1.89	2.6E-09	1003	1.82	3.2E-06	841
1284	COL4A2	1.19	1.3E-01	901	1.15	1.0E-01	1933
9833	KIAA0175	1.83	6.7E-09	1063	2.09	1.1E-16	1933
92140	MTDH	1.08	4.6E-01	1035	1.13	3.3E-01	841
51377	UCH37	0.98	8.6E-01	1031	1.17	2.0E-01	841
51560	RAB6B	1.10	4.4E-01	1003	0.79	5.6E-02	841
160897	GPR180	0.90	4.9E-01	402	0.89	3.4E-01	841
79888	FLJ12443	1.43	1.6E-03	901	1.17	6.5E-02	1933
8293	SERF1A	1.10	7.8E-01	53	0.78	3.6E-03	1933
8476	PK428	1.15	1.9E-01	1063	1.22	2.2E-02	1933
10403	HEC	1.12	3.1E-01	1063	2.15	<1E-16	1933
8833	GMPS	1.29	2.0E-02	1038	1.76	2.4E-10	1933
1894	ECT2	1.35	5.0E-03	1031	1.31	3.1E-02	841
4318	MMP9	1.13	2.5E-01	1063	1.25	9.5E-03	1933
5019	OXCT	1.01	9.4E-01	1063	0.9	2.2E-01	1933
2781	GNAZ	1.19	1.2E-01	1035	1.03	7.5E-01	1933

2321	FLT1	0.92	5.1E-01	721	0.79	6.8E-02	841
2131	EXT1	0.98	8.7E-01	1063	1.16	9.6E-02	1933
56942	DC13	1.69	7.3E-07	1031	1.89	6.0E-13	1933
81624	DIAPH3	1.35	1.8E-02	899	1.89	5.5E-07	841
169714	QSOX2	1.03	8.6E-01	402	0.77	3.4E-02	841
286052	LOC286052	NA	NA	NA	0.91	4.5E-01	841
51203	LOC51203	1.77	4.7E-08	1031	2.27	<1E-16	1933
85453	TSPYL5	1.06	5.9E-01	899	1.04	6.8E-01	1933

Table 8.10: Overexpression of *HLXB9* is a better indicator of relapse in breast cancer patients in patients diagnosed with luminal A tumours than a number of genes used in the MammaPrint gene panel. Results from analysing overall survival in KMplotter and *BreastMark*.

Figure A5: A map showing the approximate localisation of RP5-1121A15 probe on 7q. The probe contains both the *NOM1* and *HLXB9* gene and is localised between 156 and 157 (mb) on chromosome 7.

Chapter 8: References

Abate-Shen C (2002) Deregulated homeobox gene expression in cancer: cause or consequence? NatRevCancer 2 (10): 777–785.

Agarwal S, Tafel A, Kanaar R (2006) DNA double-strand break repair and chromosome translocations. DNA Repair (Amst) 5 (9-10): 1075–81.

Allford S, Grimwade D, Langabeer S, et al (1999) Identification of the t (15 ; 17) in AML FAB types other than M3 : evaluation of the role of molecular screening for the PML / RARa rearrangement in newly diagnosed AML. Br J Haematol 105 (1): 198–207.

Almstrup K, Hoei-Hansen CE, Nielsen JE, et al (2005) Genome-wide gene expression profiling of testicular carcinoma in situ progression into overt tumours. Br J Cancer 92 (10) 1934–1941

Almstrup K, Leffers H, Lothe R, et al (2007) Improved gene expression signature of testicular carcinoma in situ. Int J Androl 30 (4): 292–303.

Alseraye FM, Zuo Z, Bueso-Ramos C, et al (2011) Trisomy 11 as an isolated abnormality in acute myeloid leukemia is associated with unfavorable prognosis but not with an NPM1 or KIT mutation. Int J Clin Exp Pathol 4 (4): 371–7.

Althuis MD. et al, 2005. Global trends in breast cancer incidence and mortality 1973-1997. International journal of epidemiology, 34(2), pp.405–12.

Apostolou P, Fostira F (2013) Hereditary breast cancer: the era of new susceptibility genes. Biomed Res Int 2013:747318.

Arber S, Han B, Mendelsohn M, et al (1999) Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity. Neuron 23(4): 659–74.

Arnone P, Zurrida S, Viale G, et al (2010) The TNM classification of breast cancer: Need for change. Updates Surg 62(2): 75–81.

Bailey P, Chang DK, Nones K, et al (2016) Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 531 (7592): 47–52.

Bain BJ (2001) Overview. Cytogenetic analysis in haematology. Best Pract Res Clin Haematol 14(3):463–77.

Bakhoum SF, Compton DA (2012) Science in medicine Chromosomal instability and cancer : a complex relationship with therapeutic potential. J Clin Invest 122(4): 1138–1143.

Balgobind BV, Heuvel-Eibrink MM, Menezes RX, et al (2011) Evaluation of gene expression signatures predictive of cytogenetic and molecular subtypes of pediatric acute myeloid leukemia. Haematologica 96(2): 221-30.

Ballabio E, Cantarella CD, Federico C, et al (2009) Ectopic expression of the *HLXB9* gene is associated with an altered nuclear position in t(7;12) leukaemias. Leukemia 23(6):1179–1182.

Bangs CD, Donlon TA (2005) Metaphase chromosome preparation from cultured peripheral blood cells. Curr Protoc Hum Genet Chapter 4:Unit 4.1.

Belloni E, Martucciello G, Verderio D, et al (2000) Involvement of the *HLXB9* Homeobox Gene in Currarino Syndrome. Am J Hum Genet 66 (1):312–319.

Bennett JM, Catovsky D, Daniel M-T, et al (1976) Proposals for the Classification of the Acute Leukaemias French-American-British (FAB) Co-operative Group. Br J Haematol 33(4): 451–458.

Bertucci F, Finetti P, Birnbaum D (2012) Basal breast cancer: a complex and deadly molecular subtype. Curr Mol Med 12(1): 96–110.

Betz BL, Hess JL (2010) Acute myeloid leukemia diagnosis in the 21st century. Arch Pathol Lab Med 134(10):1427–33.

Bhatlekar S, Fields JZ, Boman BM (2014). Hox genes and their role in the development of human cancers. J Mol Med 92(8):811-23. doi: 10.1007/s00109-014-1181-y. Epub 2014 Jul 5.

Björkholm M, Hultcrantz M, Derolf ÅR (2014) Leukemic transformation in myeloproliferative neoplasms: Therapy-related or unrelated? Best Pract Res Clin Haematol 27 (2):141–153.

Bomken SN, Josef Vormoor H (2009) Childhood leukaemia. Paediatr Child Health (Oxford) 19(8):345–350.

Boultwood J (2013) CUX1 in leukemia: dosage matters. Blood 121:869–71.

Boyle S, Gilchrist S, Bridger J, et al (2001) The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Hum Mol Genet 10(3):211–219.

Branco MR, Branco T, Ramirez F, Pombo A (2008) Changes in chromosome organization during PHA-activation of resting human lymphocytes measured by cryo-FISH. Chromosom Res 16(3):413–426.

Braoudaki M, Tzortzatou-Stathopoulou F (2012) Clinical cytogenetics in pediatric acute leukemia: an update. Clin Lymphoma Myeloma Leuk 12(4):230–7.

Březinová J, Zemanová Z, Ransdorfová Š, et al (2007) Structural aberrations of chromosome 7 revealed by a combination of molecular cytogenetic techniques in myeloid malignancies. Cancer Genet Cytogenet 173(1):10–16.

Bruin AM De, Voermans C, Nolte M a (2014) Impact of interferon- g on hematopoiesis. Blood 124(16):2479–2487.

Buchholz TA, Wazer DE (2002) Molecular biology and genetics of breast cancer development: a clinical perspective. Semin Radiat Oncol 12(4):285–95.

Burke HB (2016) Predicting Clinical Outcomes Using Molecular Biomarkers. Biomark cancer 8: 89–99.

Cairns P (2011) Renal Cell Carcinoma. Cancer Biomark 9 (1-6):461–473.

Caldarella A, Buzzoni C, Crocetti E, et al (2013) Invasive breast cancer: A significant correlation between histological types and molecular subgroups. J Cancer Res Clin Oncol 139(4):617–623.

Cancer Research UK Statistics, http://www.cancerresearchuk.org/health-professional/cancer-statistics

Cazzola M, Malcovati L, Invernizzi R (2011) Myelodysplastic/myeloproliferative neoplasms. Hematology 2011:264– 272.

Chambeyron S, Bickmore WA (2004) Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev 18(10):1119–1130.

Chang KL, Weiss LM (2007) Blood and Bone Marrow : Myeloid Leukaemias and Related Neoplasms. The Cancer Handbook.

Charames GS, Bapat B (2003) Genomic instability and cancer. Curr Mol Med 3(7):589–596.

Chen F, Zhang Y, Şenbabaoğlu Y, et al (2016) Multilevel Genomics-Based Taxonomy of Renal Cell Carcinoma. Cell Reports 14(10):2476–2489.

Chen H, Sukumar S (2003) Role of Homeobox Genes in Normal Mammary Gland Development and Breast Tumorigenesis. J Mammary Gland Biol Neoplasia 8(2):159–175.

Choi H-J, Kim H-R, Shin M-G, et al (2011) Spectra of chromosomal aberrations in 325 leukemia patients and implications for the development of new molecular detection systems. J Korean Med Sci 26(7):886–92.

Corces MR, Corces VG (2016) The three-dimensional cancer genome. Curr Opin Genet Dev 36:1-7.

Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2:292–301.

Cremer T, Cremer M (2010) Chromosome territories. Cold Spring Harb Perspect Biol 2(3): a003889.

Crétolle C, Pelet A, Sanlaville D, et al (2008) Spectrum of *HLXB9* gene mutations in Currarino syndrome and genotype-phenotype correlation. Hum Mutat 29(7): 903–910.

Curigliano G (2012) New drugs for breast cancer subtypes: Targeting driver pathways to overcome resistance. Cancer Treat Rev 38(4): 303–310.

Curtis C, Shah SP, Chin SF, et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature, 486(7403), 346–352. http://doi.org/10.1038/nature10983

Cuturilo G, Hodge JC, Runke CK, et al (2016) Phenotype analysis impacts testing strategy in patients with Currarino syndrome. Clin Genet 89(1): 109–114.

Davies EL (2012) Breast cancer. Medicine Journal 40(1): 5–9.

Davoli A, Hocevar B a, Brown TL (2010) Progression and treatment of HER2-positive breast cancer. Cancer Chemother Pharmacol 65(4): 611–23.

de Jonge HJ, Huls G, de Bont ES (2011) Gene expression profiling in acute myeloid leukaemia. Neth J Med 69(4): 167–76.

de Rooij JD, Zwaan CM, van den Heuvel-Eibrink M (2015) Pediatric AML: From Biology to Clinical Management. J Clin Med 4(1): 127–149.

De Weer A, Poppe B, Vergult S, et al (2010) Identification of two critically deleted regions within chromosome segment 7q35-q36 in EVI1 deregulated myeloid leukemia cell lines. PLoS One 5(1): e8676.

Deguchi Y, Kehrl JH (1991) Selective expression of two homeobox genes in CD34-positive cells from human bone marrow. Blood 78(2): 323–8.

Delaunay J, Vey N, Leblanc T, et al (2003) Prognosis of inv(16)/t(16;16) acute myeloid leukemia (AML): a survey of 110 cases from the French AML Intergroup. Blood 102(2): 462–9.

Desai SS, Kharade SS, Parekh VI, et al (2015) Pro-oncogenic Roles of *HLXB9* Protein in Insulinoma Cells through Interaction with Nono Protein and Down-regulation of the c-Met Inhibitor Cblb (Casitas B-lineage Lymphoma b). J Biol Chem 290(42): 25595–25608.

Desai SS, Modali SD, Parekh VI, et al (2014) GSK-3β Protein Phosphorylates and Stabilizes *HLXB9* Protein in Insulinoma Cells to Form a Targetable Mechanism of Controlling Insulinoma Cell Proliferation. J Biol Chem 289(9): 5386–5398.

Desmedt C, Voet T, Sotiriou C, et al. (2012). "Next Generation Sequencing in Breast Cancer: First Take Home Messages." Current opinion in oncology 24.6: 597–604.

Dieppois G, Stutz F (2010). Connecting the transcription site to the nuclear pore: a multi-tether process that regulates gene expression. J Cell Sci 123(Pt 12):1989-99. doi: 10.1242/jcs.053694

Dirix M, Van Becelaere T, Berkenbosch L, et al (2015) Malignant transformation in sacrococcygeal teratoma and in presacral teratoma associated with Currarino syndrome: A comparative study. J Pediatr Surg 50(3):462–464.

Döhner H, Estey EH, Amadori S, et al (2010) Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood 115(3):453–74.

Dormoy V, Jacqmin D, Lang H, Massfelder T (2012) From development to cancer: lessons from the kidney to uncover new therapeutic targets. Anticancer Res 32(9):3609–17.

Dreijerink KMA, Goudet P, Burgess JR, Valk GD (2014) Breast-Cancer Predisposition in Multiple Endocrine Neoplasia Type 1. N Engl J Med 371:583–584.

Dreijerink KMA, Mulder KW, Winkler GS, et al (2006) Menin links estrogen receptor activation to histone H3K4 trimethylation. Cancer Res 66(9):4929–4935.

Ebert BL (2010) Genetic deletions in AML and MDS. Best Pract Res Clin Haematol 23(4):457–461.

Eden T (2010) Aetiology of childhood leukaemia. Cancer Treat Rev 36(4):286–97.

Engreitz JM, Agarwala V, Mirny LA (2012) Three-Dimensional Genome Architecture Influences Partner Selection for Chromosomal Translocations in Human Disease. PLoS One 7(9): e044196

Ernst T, Chase AJ, Score J, et al (2010) Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet 42(8): 722–726.

Eroles P, Bosch A, Alejandro Perez-Fidalgo J, Lluch A (2012) Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways. Cancer Treat Rev 38(6): 698–707.

Estey E (2014) Acute myeloid leukemia: 2014 update on risk-stratification and management. Am J Hematol 89(11): 1063-1081.

Estey E, Döhner H (2006) Acute myeloid leukaemia. Lancet 368(9550): 1894–907.

Falini B, Tiacci E, Martelli M, et al (2010) New classification of acute myeloid leukemia and precursor-related neoplasms: changes and unsolved issues. Discov Med 10(53):281–292.

Federico C, Cantarella CD, Mare P, et al (2008) The radial arrangement of the human chromosome 7 in the lymphocyte cell nucleus is associated with chromosomal band gene density. Chromosoma. 117(4): 399-410.

Ferguson S, Gautrey H, Strathdee G (2011) The dual role of *HLXB9* in leukemia. Pediatr Blood Cancer 56(3):349–352. doi: 10.1002/pbc.22679.

Ferrara F, Schiffer CA (2013) Acute myeloid leukaemia in adults. Lancet 381(9865): 484–95.

Finlan LE, Sproul D, Thomson I, et al (2008) Recruitment to the Nuclear Periphery Can Alter Expression of Genes in Human Cells. PLoS Genet 4(3): e1000039.

Flanagan SE, De Franco E, Lango Allen H, et al (2014) Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man. Cell Metab 19(1):146–154.

Foucar K (2009) Myelodysplastic/Myeloproliferative Neoplasms. Am J Clin Pathol 132(2):281–289.

Gabay C, Ben-Bassat H, Schlesinger M, Laskov R (1999) Somatic mutations and intraclonal variations in the rearranged Vkappa genes of B-non-Hodgkin's lymphoma cell lines. Eur J Haematol 63(3):180–91.

Galea MH, Blamey RW, Elston CE, Ellis IO (1992) The Nottingham prognostic index in primary breast cancer. Breast Cancer Res Treat 22(3): 207–219.

Ganem NJ, Pellman D (2012) Linking abnormal mitosis to the acquisition of DNA damage. J Cell Biol 199(6):871–881.

Garbe, J.C. et al., 2009. Molecular distinctions between the stasis and telomere attrition senescence barriers demonstrated by long-term culture of normal human mammary epithelial cells. Cancer Research, 69(19): 7557–7568.

Garbe, J.C. et al., 2014. Immortalization of normal human mammary epithelial cells in two steps by direct targeting of senescence barriers does not require gross genomic alterations. Cell Cycle, 13(21), pp.3423–3435.

Garcia-Barceló MM, Lui VC, So M, et al (2009) MNX1 (*HLXB9*) mutations in Currarino patients. J Pediatr Surg 44(10):1892–8.

García-Closas M, Brinton LA, Lissowska J, et al (2006) Established breast cancer risk factors by clinically important tumour characteristics. Br J Cancer 95(1):123–9.

Garcia-Fernàndez J (2005) The genesis and evolution of homeobox gene clusters. Nat Rev Genet 6(12):881–92.

Geyer PK, Vitalini MW, & Wallrath LL (2011). Nuclear organization: taking a position on gene expression. Current Opinion in Cell Biology, 23(3), 354–359. http://doi.org/10.1016/j.ceb.2011.03.002

Gordon DJ, Resio B, Pellman D (2012) Causes and consequences of aneuploidy in cancer. Nat Rev Genet 13(3):189–203.

Greaves M (1999) Molecular genetics, natural history and the demise of childhood leukaemia. Eur J Cancer 35(2):173–85.

Greaves M, Maley CC (2012) Clonal evolution in cancer. Nature 481(7381):306–13.

Gregory TK, Wald D, Chen Y, et al (2009) Molecular prognostic markers for adult acute myeloid leukemia with normal cytogenetics. J Haematol Oncol 2: 23.

Gribble SM, Roberts I, Grace C, et al (2000) Cytogenetics of the chronic myeloid leukemia-derived cell line K562: karyotype clarification by multicolor fluorescence in situ hybridization, comparative genomic hybridization, and locus-specific fluorescence in situ hybridization. Cancer Genet Cytogenet 118(1): 1–8.

Grimwade D, Walker H, Oliver F, et al (2013) The Importance of Diagnostic Cytogenetics on Outcome in AML: Analysis of 1,612 Patients Entered Into the MRC AML 10 Trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. Blood 92(7): 2322–2333.

Gulley ML, Shea TC, Fedoriw Y (2010) Genetic tests to evaluate prognosis and predict therapeutic response in acute myeloid leukemia. J Mol Diagn 12(1): 3–16.

Gyorffy, B, Surowiak P, Budczies J, Lánczky A (2013). Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PLoS ONE, 8(12): e82241.

Hagan DM, Ross AJ, Strachan T, et al (2000) Mutation analysis and embryonic expression of the *HLXB9* Currarino syndrome gene. Am J Hum Genet 66(5): 1504–15.

Harewood L, Schutz F, Boyle S, et al (2010) The effect of translocation-induced nuclear reorganization on gene expression. Genome Res 20(5): 554–564.

Harrison KA, Druey KM, Deguchi Y, et al (1994) A novel human homeobox gene distantly related to proboscipedia is expressed in lymphoid and pancreatic tissues. J Biol Chem 269(31): 19968–75.

Harrison KA, Thaler J, Pfaff SL, et al (1999) Pancreas dorsal lobe agenesis and abnormal islets of Langerhans in *HLXB9*-deficient mice. Nat Genet 23(1):71–5.

Harvey AJ. Pennington CJ, Porter S, et al (2009) Brk Protects Breast Cancer Cells from Autophagic Cell Death Induced by Loss of Anchorage. The American Journal of Pathology, 175(3), 1226–1234.

Hasle H, Alonzo TA, Auvrignon A, et al (2007) Monosomy 7 and deletion 7q in children and adolescents with acute myeloid leukemia: an international retrospective study. Blood 109 (11):4641–7.

Henry NL, Hayes DF (2012) Cancer biomarkers. Mol Oncol 6(2):140–6.

Holland PWH (2013) Evolution of homeobox genes. Wiley Interdiscip Rev Dev Biol 2 (1): 31–45.

Holland PWH, Booth HA, Bruford EA (2007) Classification and nomenclature of all human homeobox genes. BMC Biol 5:47

Holland PWH, Takahashi T (2005) The evolution of homeobox genes: Implications for the study of brain development. Brain Res Bull 66(4-6):484–90.

Hollington P, Neufing P, kalionis B, et al (2004) Expression and Localization of Homeodomain Proteins DLX4, HB9 and HB24 in Malignant and Benign Human Colorectal Tissues. Anticancer Res 24(2B): 955–962

Honda H, Nagamachi A, Inaba T (2015). 7/7Q Syndrome in Myeloid-Lineage Hematopoietic Malignancies: Attempts To Understand This Complex Disease Entity. Oncogene 34(19): 2413–25.

Hosono N, Makishima H, Jerez A, et al (2014) Recurrent genetic defects on chromosome 7q in myeloid neoplasms. Leukemia 28(6): 1348–51.

https://www.nice.org.uk/guidance/dg10

Hu C, Wei W, Chen X, et al (2012) A Global View of the Oncogenic Landscape in Nasopharyngeal Carcinoma: An Integrated Analysis at the Genetic and Expression Levels. PLoS One 7(7): e41055.

Hussain T, Mulherkar R (2012) Lymphoblastoid Cell lines: a Continuous in Vitro Source of Cells to Study Carcinogen Sensitivity and DNA Repair. Int J Mol Cell Med 1(2): 75–87.

Imachi H, Murao K, Dobashi H, et al (2010) Menin, a product of the MENI gene, binds to estrogen receptor to enhance its activity in breast cancer cells: possibility of a novel predictive factor for tamoxifen resistance. Breast Cancer Res Treat 122(2): 395–407.

Inic Z, Zegarac M, Inic M, et al (2014) Difference between Luminal A and Luminal B subtypes according to Ki-67, tumor size, and progesterone receptor negativity providing prognostic information. Clin Med Insights Oncol 8: 107–111.

Ioannou D, Kandukuri L, Simpson JL, Tempest HG (2015) Chromosome territory repositioning induced by PHA-activation of lymphocytes: A 2D and 3D appraisal. Mol Cytogenet 8: 47.

Jagannathan-Bogdan M, Zon LI (2013) Hematopoiesis. Development 140(12): 2463–2467.

Johnson E, Cotter FE (1997) Monosomy 7 and 7q--associated with myeloid malignancy. Blood Rev 11(1): 46–55.

Kim BR, Kim JE, Woo KS, et al (2012) A case of near-triploidy in myelodysplastic syndrome with del(5q) combined with del(1p) and del(13q). Ann Lab Med 32(4): 294–7.

Klaus M, Haferlach T, Schnittger S, et al (2004) Cytogenetic profile in de novo acute myeloid leukemia with FAB subtypes M0, M1, and M2: a study based on 652 cases analyzed with morphology, cytogenetics, and fluorescence in situ hybridization. Cancer Genet Cytogenet 155(1):47–56.

Kondo M (2010) Lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors. Immunol Rev 238(1): 37–46.

Korbling M and Anderlini P (2001) Peripheral blood stem cell versus bone marrow allotransplantation: does the source of hematopoietic stem cell matter? Blood 98 (10): 2900-2910

Kosan C, Godmann M (2016) Genetic and Epigenetic Mechanisms That Maintain Hematopoietic Stem Cell Function. Stem Cell Int. 2016: 5178965.

Kralik JM, Kranewitter W, Boesmueller H, et al (2011) Characterization of a newly identified ETV6-NTRK3 fusion transcript in acute myeloid leukemia. Diagn Pathol 6:19.

Kuipers JE, Coenen EA, Balgobind B V, et al (2011) High IGSF4 expression in pediatric M5 acute myeloid leukemia with t(9;11)(p22;q23). Blood 117(3): 928–35.

Kwong YL (1998) Translocation (3;5)(q21;q34) in Erythroleukemia: A Molecular and In Situ Hybridization Study. Cancer Genet Cytogenet 103(1): 15–19.

Langmuir PB, Aplenc R, Lange BJ (2001) Acute myeloid leukaemia in children. Best Pract Res Clin Haematol 14(1):77–93. doi: 10.1053/beha.2000.0117

Laulier C, Barascu A, Guirouilh-Barbat J, et al (2011) Bcl-2 inhibits nuclear homologous recombination by localizing BRCA1 to the endomembranes. Cancer Res 71(10):3590–602.

Leotta CG, Federico C, Brundo MV, et al (2014) *HLXB9* gene expression, and nuclear location during in vitro neuronal differentiation in the SK-N-BE neuroblastoma cell line. PLoS One 9(8): e105481.

Leshner, M, Devine M, Roloff GW, et al., (2016). Locus-specific gene repositioning in prostate cancer. Mol biol cell, 27(2), pp.236–46.

Li H, Arber S, Jessell TM, Edlund H (1999) Selective agenesis of the dorsal pancreas in mice lacking homeobox gene *HLXB9*. Nat Genet 23(1): 67–70.

Li H, Edlund H (2001) Persistent expression of *HLXB9* in the pancreatic epithelium impairs pancreatic development. Dev Biol 240(1): 247–53.

Li L, Li J, Li G, et al (2012) A tetraploid minimally differentiated acute myeloblastic leukemia with extensive erythrophagocytosis: a case report and literature review. Int J Hematol 96(6): 801–5.

Lian ZQ, Wang Q, Wen-Ping L, et al (2012) Screening of significantly hypermethylated genes in breast cancer using microarray-based methylated-CpG island recovery assay and identification of their expression levels. Int J Oncol 41(2): 629-38.

Liebens FP, Carly B, Pastijn A, Rozenberg S (2007) Management of BRCA1/2 associated breast cancer: a systematic qualitative review of the state of knowledge in 2006. Eur J Cancer 43(2):238–57.

Lieber MR, Ma Y, Pannicke U, Schwarz K (2003) Mechanism and regulation of human non-homologous DNA end-joining. Nat Rev Mol Cell Biol 4(9):712–20.

Madden SF, Clarke C, Gaule P, et al., 2013. BreastMark: an integrated approach to mining publicly available transcriptomic datasets relating to breast cancer outcome. Breast Cancer Res, 15(4): R52.

Manola KN (2009) Cytogenetics of pediatric acute myeloid leukemia. Eur J Haematol 83(5):391–405.

Manola KN, Panitsas F, Polychronopoulou S, et al (2013) Cytogenetic abnormalities and monosomal karyotypes in children and adolescents with acute myeloid leukemia: correlations with clinical characteristics and outcome. Cancer Genet 206(6): 63–72.

Marrone M, Stewart A, & Dotson WD (2015). Clinical utility of gene-expression profiling in women with early breast cancer: an overview of systematic reviews. Genetics in Medicine : Official Journal of the American College of Medical Genetics, 17(7), 519–532. http://doi.org/10.1038/gim.2014.140

Martens JHA, Stunnenberg HG (2010) The molecular signature of oncofusion proteins in acute myeloid leukemia. FEBS Lett 584(12):2662–9.

Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M (2010) Genetic susceptibility to breast cancer. Mol Oncol 4(3):174–191.

Mazo IB, Massberg S, von Andrian UH (2011) Hematopoietic stem and progenitor cell trafficking. Trends Immunol 32(10):493–503.

Mcclellan JS, Majeti R (2013) The cancer stem cell model: B cell acute lymphoblastic leukaemia breaks the mould. EMBO Mol Med 5(1):7–9.

McNerney ME, Brown CD, Wang X, et al (2013) CUX1 is a haploinsufficient tumor suppressor gene on chromosome 7 frequently inactivated in acute myeloid leukemia. Blood 121(6):975–983.

Meaburn KJ, Agunloye O, Devine M, et al (2016) Tissue-of-origin-specific gene repositioning in breast and prostate cancer. Histochem Cell Biol 145(4): 433–446.

Meaburn KJ, Gudla PR, Khan S, et al (2009) Disease-specific gene repositioning in breast cancer. J Cell Biol 187(6): 801–812.

Meaburn KJ, Misteli T (2007) Cell biology: Chromosome territories. Nature 445(7126): 379–381.

Meaburn KJ, Misteli T (2008) Locus-specific and activity-independent gene repositioning during early tumorigenesis. J Cell Biol 180(1):39–50.

Meaburn KJ, Misteli T, Soutoglou E (2007) Spatial genome organization in the formation of chromosomal translocations. Semin Cancer Biol 17(1):80–90.

Mercher T, Coniat MB, Khac FN, et al (2002) Recurrence of OTT – MAL Fusion in t (1 ; 22) of Infant AML-M7. Genes Chromosomes Cancer 33(1): 22–8.

Miller KR, Patel JN, Ganapathi MK, et al (2016) Biological role and clinical implications of homeobox genes in serous epithelial ovarian cancer. Gynecol Oncol. 141(3): 608-15

Mrózek K, Heerema NA, Bloomfield CD (2004) Cytogenetics in acute leukemia. Blood Rev 18(2):115–36.

Murphy LC, Watson P (2002) Steroid receptors in human breast tumorigenesis and breast cancer progression. Biomed Pharmacother 56(2): 65–77.

Murray AJ, Davies DM (2013) The genetics of breast cancer. Surg 31(1): 1–3.

Nagel S, Ehrentraut S, Meyer C, et al (2014) Oncogenic deregulation of NKL homeobox gene MSX1 in mantle cell lymphoma. Leuk Lymphoma 55(8): 1893–903.

Nagel S, Ehrentraut S, Meyer C, et al (2015) Aberrantly Expressed OTX Homeobox Genes Deregulate B-Cell Differentiation in Hodgkin Lymphoma. PLoS One 10(9): e0138416.

Nagel S, Kaufmann M, Scherr M, et al (2005a) Activation of *HLXB9* by juxtaposition with MYB via formation of t(6;7)(q23;q36) in an AML-M4 cell line (GDM-1). Genes Chromosom Cancer 42(2): 170-8.

Nagel S, Scherr M, Quentmeier H, et al (2005b) *HLXB9* activates IL6 in Hodgkin lymphoma cell lines and is regulated by PI3K signalling involving E2F3. Leukemia 19 (5): 841–6.

Naiel A, Vetter M, Plekhanova O, et al (2013) A novel three-colour fluorescence in situ hybridization approach for the detection of t(7;12)(q36;p13) in acute myeloid leukaemia reveals new cryptic three way translocation t(7;12;16). Cancers (Basel) 5(1): 281–295.

Nambiar M, Raghavan SC (2011) How does DNA break during chromosomal translocations? Nucleic Acids Res 39(14): 5813–5825.

Narod SA (2012) Tumour size predicts long-term survival among women with lymph node-positive breast cancer. Curr Oncol 19(5): 249–253.

Neufing P, Kalionis B, Horsfall D, et al (2003) Expression and localization of homeodomain proteins DLX4/HB9 in normal and malignant human breast tissues. Anticancer 23(2B): 1479–88.

Nguyen LV, Vanner R, Dirks P, Eaves CJ (2012) Cancer stem cells: an evolving concept. Nat Rev Cancer 12:133–143.

Novotny GW, Nielsen JE, Sonne SB, et al (2007) Analysis of gene expression in normal and neoplastic human testis: new roles of RNA. Int J Androl 30(4):316–27.

O'Donovan MR, Johns S, Wilcox P (1995) The effect of PHA stimulation on lymphocyte sub-populations in whole-blood cultures. Mutagenesis 10(4):371–4.

Odden JP, Holbrook S. & Doe CQ (2002). Drosophila HB9 is expressed in a subset of motoneurons and interneurons, where it regulates gene expression and axon pathfinding. The Journal of neuroscience : the official journal of the Society for Neuroscience, 22(21), pp.9143–9.

Ohgami RS, Arber DA (2015) The diagnostic and clinical impact of genetics and epigenetics in acute myeloid leukemia. Int J Lab Hematol 37:122–132.

Okabayashi T, Shima Y, Sumiyoshi T, et al (2013) Diagnosis and management of insulinoma. World J Gastroenterol 19(6):829–837.

Olsen J, Eiholm S, Kirkeby LT, et al (2016) CDX2 downregulation is associated with poor differentiation and MMR deficiency in colon cancer. Exp Mol Pathol 100(1):59–66.

Ozsaran Z, Alanyah SD (2013) Staging of Breast Cancer. In: Haydaroglu A, Ozyigit G (eds) Principles and Practice of Modern Radiotherapy Techniques in Breast Cancer. Springer.

Pan FC, Brissova M, Powers AC, et al (2015) Inactivating the permanent neonatal diabetes gene Mnx1 switches insulin-producing β-cells to a δ-like fate and reveals a facultative proliferative capacity in aged β-cells. Development 142(21):3637–48.

Paradis V (2013) Histopathology of Hepatocellular Carcinoma. In: Vauthey J, Brouquet A (eds) Multidisciplinary Treatment of Hepatocellular Carcinoma. Springer, pp 21–33

Park J, Kim M, Lim J, et al (2009) Three-way complex translocations in infant acute myeloid leukemia with t(7;12)(q36;p13): the incidence and correlation of a *HLXB9* overexpression. Cancer Genet Cytogenet 191(2): 102–5.

Parker JS, Mullins M, Cheang MC, et al (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 27(8): 1160–7.

Pelechano, V. & Steinmetz, LM (2013). Gene regulation by antisense transcription. Nat Rev Genet, 14(12), pp.880–893.

Pereira B, Chin SF, Rueda OM, et al. (2016). The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nature Communications,10;7:11479. doi: 10.1038/ncomms11479.

Perou CM, Sørlie T, Eisen MB, et al (2000) Molecular portraits of human breast tumours. Nature 406 (6797):747–52.

Pombo A, Dilon N (2015). Three-dimensional genome architecture: players and mechanisms. Nature Reviews Molecular Cell Boiology.16, 245–257(2015) doi:10.1038/nrm3965

Potapova TA, Zhu J, Li R (2013) Aneuploidy and chromosomal instability: a vicious cycle driving cellular evolution and cancer genome chaos. Cancer Metastasis Rev 32(3-4): 377-89.

Purizaca J, Meza I, Pelayo R (2012) Early Lymphoid Development and Microenvironmental Cues in B-cell Acute Lymphoblastic Leukemia. Arch Med Res 43(2):89–101.

Rabbitts TH (1994) Chromosomal translocations in human cancer. Nature 372 (6502):143–9.

Ragoczy T, Bender MA, Telling A et al. (2006). The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. Genes Dev 20(11):1447-57

Rakha EA, Reis-Filho JS, Baehner F, et al (2010) Breast cancer prognostic classification in the molecular era: the role of histological grade. Breast Cancer Res 12(4): 207.

Rhodes DR, Ateeq B, Cao Q, et al (2009) AGTR1 overexpression defines a subset of breast cancer and confers sensitivity to losartan, an AGTR1 antagonist. Proc Natl Acad Sci U S A 106 (25): 10284–9.

Rhodes DR, Kalyana-Sundaram S, Mahavisno V, et al (2007) Oncomine 3.0: Genes, Pathways, and Networks in a Collection of 18,000 Cancer Gene Expression Profiles. Neoplasia 9(2):166–180.

Rhodes DR, Yu J, Shanker K, et al (2004) ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. Neoplasia 6(1): 1–6.

Rondón-Lagos M, Verdun Di Cantogno L, Marchiò C, et al (2014). Differences and homologies of chromosomal alterations within and between breast cancer cell lines: a clustering analysis. Molecular cytogenetics, 7(1): 8

Ross AJ, Ruiz-Perez V, Wang Y, et al (1998) A homeobox gene, *HLXB9*, is the major locus for dominantly inherited sacral agenesis. Nat Genet 20 (4): 358–361.

Rubin P, Hansen JT (2012) Hematopoietic System. In: TNM Staging Atlas with Oncoanatomy, 2nd ed. Lippincott Williams & Wilkins.

Segeren CM, Veer MB Van (1996) The FAB classification for acute myeloid leukaemia-is it outdated ? Neth J Med 49(3): 126–131.

Shaffer LG, Rosenfeld JA, Dabell MP, et al (2012a) Detection rates of clinically significant genomic alterations by microarray analysis for specific anomalies detected by ultrasound. Prenat Diagn 32 (10): 986–95.

Shaffer LG, Schultz RA, Ballif BC (2012b) The use of new technologies in the detection of balanced translocations in hematologic disorders. Curr Opin Genet Dev 22(3): 264–71.

Sherwood RI, Chen TY, Melton DA (2009) Transcriptional dynamics of endodermal organ formation. Dev Dyn 238(1): 29–42.

Shi K, Parekh VI, Roy S, et al (2013) The embryonic transcription factor *HLXB9* is a menin interacting partner that controls pancreatic β-cell proliferation and the expression of insulin regulators. Endocrine-Related Cancer 20(1): 111–22.

Shin G, Kang T-W, Yang S, et al (2011) GENT: Gene Expression Database of Normal and Tumor Tissues. Cancer Inform 10:149–157.

Siletz A, Kniazeva E, Jeruss JS, Shea LD (2013a) Transcription Factor Networks in Invasion-Promoting Breast Carcinoma-Associated Fibroblasts. Cancer Microenviron 6(1):91–107.

Siletz A, Schnabel M, Kniazeva E, et al (2013b) Dynamic Transcription Factor Networks in Epithelial-Mesenchymal Transition in Breast Cancer Models. PLoS One 8(4):e57180.

Simmons HM, Oseth L, Nguyen P, et al (2002) Cytogenetic and molecular heterogeneity of 7q36/12p13 rearrangements in childhood AML. Leukemia 16(12): 2408-16.

Skok JA, Brown KE, Azuara V, et al (2001) Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes. Nat Immunol 2(9):848–54.

Sommer S, Fuqua SA(2001) Estrogen receptor and breast cancer. Semin Cancer Biol 11(5): 339–52.

Sorlie T, Tibshirani R, Parker J, et al (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100(14):8418–23.

Sotiriou C, Wirapati P, Loi S et al., 2006. Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst. 98(4):262–72.

Stephens PJ, Greenman CD, Fu B, et al (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 144(1):27–40.

Strachan T, Read AP (2011) Human Molecular Genetics 4, 4th Edition. Garland Science/Taylor & Francis Group Pg 52.

Strimbu K, Tavel JA (2011) What are Biomarkers? Curr Opin HIV AIDS 5(6):463–466.

Svigen T, Tonissen KF (2003) Altered HOX gene expression in human skin and breast cancer cells. Cancer Biol Ther 2(5): 518–523.

Szczerbal I, Foster HA, Bridger JM (2009) The spatial repositioning of adipogenesis genes is correlated with their expression status in a porcine mesenchymal stem cell adipogenesis model system. Chromosoma 118(5): 647–663.

Taddei A, Van Houwe G, Hediger F, et al. (2006). Nuclear pore association confers optimal expression levels for an inducible yeast gene. Nature 8;441(7094):774-8.

Takash, W., Cañizares, J., Bonneaud, N., Poulat, F., Mattéi, Marie-Geneviève, Jay, P., & Berta, P. (2001). SOX7 transcription factor: sequence, chromosomal localisation, expression, transactivation and interference with Wnt signalling. Nucleic Acids Research, 29(21), 4274–4283.

Taketani T, Taki T, Sako M, et al (2008) MNX1-ETV6 fusion gene in an acute megakaryoblastic leukemia and expression of the MNX1 gene in leukemia and normal B cell lines. Cancer Genet Cytogenet 186(2):115–9.

Takizawa T, Gudla PR, Guo L, et al (2008) Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. Genes Dev 22(4):489–498.

Talkowski ME, Ernst C, Heilbut A, et al (2011) Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. Am J Hum Genet 88(4):469–481.

Tallman MS, Kim HT, Paietta E, et al (2004) Acute monocytic leukemia (French-American-British classification M5) does not have a worse prognosis than other subtypes of acute myeloid leukemia: a report from the Eastern Cooperative Oncology Group. J Clin Oncol 22(7):1276–86.

Tang Y, Wang Y, Kiani MF, Wang B (2016) Classification, treatment strategy and associated drug resistance in breast cancer. Clin Breast Cancer pii: S1526-8209(16)30108-2.

Terwijn M, Zeijlemaker W, Kelder A, et al (2014) Leukemic stem cell frequency: A strong biomarker for clinical outcome in acute myeloid leukemia. PLoS One 9:7–9. doi: 10.1371/journal.pone.0107587

Thaler J, Harrison K, Sharma K, et al (1999) Active suppression of interneuron programs within developing motor neurons revealed by analysis of homeodomain factor HB9. Neuron 23(4): 675–87.

Thompson SL, Bakhoum SF, Compton DA (2010) Mechanisms of chromosomal instability. Curr Biol 20: R285–95.

Tommasi S, Karm DL, Wu X, et al (2009) Methylation of homeobox genes is a frequent and early epigenetic event in breast cancer. Breast Cancer Res 11(1):R14

Tosi S, Giudici G, Rambaldi A, et al (1999) Characterization of the human myeloid leukemia-derived cell line GF-D8 by multiplex fluorescence in situ hybridization, subtelomeric probes, and comparative genomic hybridization. Genes Chromosomes Cancer 24(3): 213–21.

Tosi S, Harbott J, Hass O, et al (1996) Classification of deletions and identification of cryptic translocations involving 7q by fluoresence in situ hybridisation. Leukemia 10(4): 644–649.

Tosi S, Harbott J, Teigler-Schlegel A, et al (2000) t(7;12)(q36;p13), a new recurrent translocation involving ETV6 in infant leukemia. Genes Chromosomes Cancer 29(4): 325–32.

Tosi S, Hughes J, Scherer SW, et al (2003) Heterogeneity of the 7q36 breakpoints in the t(7;12) involving ETV6 in infant leukemia. Genes Chromosomes Cancer 38(2):191–200.

Tosi S, Kamel YM, Owoka T, et al (2015) Paediatric acute myeloid leukaemia with t(7;12)(q36;p13) rearrangement: a review of the biological and clinical management aspects. Biomark Res 3(1) 21.

Tosi S, Scherer SW, Giudici G, et al (1999) Delineation of multiple deleted regions in 7q in myeloid disorders. Genes Chromosomes Cancer 25(4): 384–92.

van der Straaten H, van Biezen A, Brand R, et al (2005) Allogeneic stem cell transplantation for patients with acute myeloid leukemia or myelodysplastic syndrome who have chromosome 5 and/or 7 abnormalities. Haematologica 90(10):1339–45.

Vardiman JW, Thiele J, Arber DA, et al (2009) The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood 114(5):937–51.

von Bergh ARM, van Drunen E, van Wering ER, et al (2006) High Incidence of t (7 ; 12)(q36 ; p13) in Infant AML but Not in Infant ALL , with a Dismal Outcome and Ectopic Expression of *HLXB9*. Genes Chromosomes Cancer 45(8):731–9.

Wan Q, Dingerdissen H, Fan Y, et al (2015) BioXpress: an integrated RNA-seq-derived gene expression database for pan-cancer analysis. Database(Oxford) pii:bav019.

Watson MB, Greenman J, Drew PJ, et al (2004). Variation between independently cultured strains of the MDA-MB-231 breast cancer cell line identified by multicolour fluorescence in situ hybridisation. Cancer Therapy 2: 167–172.

Weigelt B, Geyer FC, Reis-filho JS (2010) Histological types of breast cancer : How special are they ? Mol Oncol 4(3): 192–208.

Welborn J, Jenks H, Taplett J, Walling P (2004) Inversion of chromosome 12 and lineage promiscuity in hematologic malignancies. Cancer Genet Cytogenet 148(2): 91–103.

Wendik B., Maier E. & Meyer D (2004). Zebrafish mnx genes in endocrine and exocrine pancreas formation. Developmental biology, 268(2), pp.372–83.

Wiech T, Stein S, Lachenmaier V, et al (2009) Spatial allelic imbalance of BCL2 genes and chromosome 18 territories in nonneoplastic and neoplastic cervical squamous epithelium. Eur Biophys J 38(6): 793–806.

Wiech T, Timme S, Riede F, et al (2005) Human archival tissues provide a valuable source for the analysis of spatial genome organization. Histochem Cell Biol 123(3): 229–238.

Wiemels J (2012) Perspectives on the causes of childhood leukemia. Chem Biol Interact 196(3): 59–67.

Wildenhain S, Ingenhag D, Ruckert C, et al (2012) Homeobox protein HB9 binds to the prostaglandin E receptor 2 promoter and inhibits intracellular cAMP mobilization in leukemic cells. J Biol Chem 287(48): 40703-12.

Wildenhain S, Ruckert C, Röttgers S, et al (2010) Expression of cell-cell interacting genes distinguishes *HLXB9*/TEL from MLL-positive childhood acute myeloid leukemia. Leukemia 24(9): 1657–60.

Wilkens L, Jaggi R, Hammer C, et al (2011) The homeobox gene *HLXB9* is upregulated in a morphological subset of poorly differentiated hepatocellular carcinoma. Virchows Arch 458 (6): 697–708.

Xu J, Chen Y, & Olopade OI. (2010) MYC and Breast Cancer. Genes & Cancer, 1(6), 629–640.

Yamamoto M, Cid E, Bru S, Yamamoto F (2011). Rare and frequent promoter methylation, respectively, of TSHZ2 and 3 genes that are both downregulated in expression in breast and prostate cancers. PloS one, 6(3): e17149.

Yohe S (2015) Molecular Genetic Markers in Acute Myeloid Leukemia. J Clin Med 4(3): 460–478.

Yoshida A, Maoate K, Blakelock R, et al (2010) Long-term functional outcomes in children with Currarino syndrome. Pediatr Surg Int 26(7): 677–81.

Zhang Y, Rowley JD (2006) Chromatin structural elements and chromosomal translocations in leukemia. DNA Repair (Amst) 5(9-10): 1282–97.

Zink D, Amaral MD, Englmann A, et al (2004) Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. J Cell Biol 166(6): 815–825.

Zu S, Winberg J, Arnberg F, et al (2011) Mutation analysis of the motor neuron and pancreas homeobox 1 (MNX1, former *HLXB9*) gene in Swedish patients with Currarino syndrome. J Pediatr Surg 46(7): 1390–5.

Zwaan CM, Kaspers GJL (2004) Possibilities for tailored and targeted therapy in paediatric acute myeloid leukaemia. Br J Haematol 127(3): 264–79.