A potential role for SOX7 in leukaemogenesis

A thesis submitted to the University of Manchester for

the degree of Doctor of Philosophy

in the Faculty of Medical and Human Sciences

2014

Sara Cuvertino

School of Medicine

Institute of Cancer Sciences

List of Contents

Final word count: 60219

List of Tables

List of Figures

List of Abbreviations

7-AAD 7-aminoactinomycin D
ABL Abelson
Acetyl-LDL acetylated low-densitylipoproteins
AGM aorta-gonad-mesonephros
AID Activation-induced cytidine deaminase
ALL cute Lymphoblastic Leukaemia
AKT protein kinase B
AML Acute Myeloid Leukaemia
APC allophycocyanin
ATP adenosine triphosphate
BCR B cell receptor
BER base excision repair
BLK B lymphoid tyrosine kinase
BL blast
BL-CFC BL colony-forming cell
BLNK B cell linker protein
bp base pairs
BP1 beta protein 1
Bry brachyury
BSA bovine serum albumin
cKIT kit oncogene
C constant
Ca^{2+} calcium
CAS casanova
CD cluster of differentiation
CDH congenital distrophia
cDNA complementary DNA
CFC colony-forming cell
ChIP chromatin IP
CLP common lymphoid precursor
CLL Chronic Lymphoblastic Leukaemia
CML Chronic Myeloid Leukaemia
CMR1 cold and menthol receptor1
CMP common myeloid precursor
CMV citomegalovirus
CNS central nervous system
CRC colorectal cancer
CTC circulating tumour cells
D diversity region
DAG diacylglycerol
DL1 delta like 1
DMEM Dulbeco's modified Eagle medium
DNA deoxyribonucleic acid
Dox doxycycline
DSB double-strand DNA breaks
E embryonic day
E12/E47 E proteins
EBs embryoid bodies
EBF1 early B cell factor 1
EdU 5-ethynyl-2-deoxyuridine
EMP erythroid-myeloid progenitors
EryP erythroid
ES embryonic stem cells
ETS E26 transformation-specific
FACS fluorescence-activated cell sorting
FCM flow cytometry
FCS foetal calf serum
FGF fibroblast growth factor
FLK1 foetal liver kinase 1
FLT fms-related tyrosine kinase
FOXO forkhead box O
GATA5 GATA binding factor 5
gDNA genomic DNA
GFP green fluorescent protein
GM-CSF granulocyte/macrophage colony-stimulating factor
GMP granulocyte-macrophage progenitors
GSEA gene enrichment analysis
Gy gray
HeLa Henrietta Lacks
HEK 293 cells human embryonic kidney 293 cells
hESC human embryonic stem cells
HLH helix-loop-helix

HMG high mobility group
HOX homeotic gene
Hprt hypoxanthine guanine phosphoribosyl transferase
HRP horseradish peroxidase
Hrs hours
HSA heat stable antigen
HSCs haematopoietic stem cells
HSPC human stem progenitor cells
ID inhibitor of DNA
Ig immunoglobulin
IgH immunoglobulin heavy chain
IgL immunoglobulin light chain
IL interleukin
IMDM Iscove's modified Dulbeco's medium
IP_3 inositol-1,4,5-triohosphate
IRF interferon regulatory factor
ITAM immuno-receptor tyrosine activation motif
IVIS *in vivo* imaging system
J joining region
JAK Janus Kinase
Kb kilo base
kDa kiloDalton
KL cKit ligand
LEF1 lymphoid enhancer binding factor 1
LMPP lympho-primed multipotent progenitor
LSK Lineage Sca1 Kit
LTR long-term reconstitution
M-CSF macrophage colony-stimulating factor
MACS magnetic activated cell sorting
MAPK mitogen-activated protein kinase
MCL mantle cell lymphoma
MDS myelodisplastic syndrome
MEF2C
MEM alpha minimum essential medium alpha
MEP megakaryocyte-erythroid progenitors
MLL mixed-lineage leukaemia
MK megakaryocyte
MMR mismatch repair
MPP multi-potent progenitors
MS1 Mile Sven
MTG monothioglycerol
MW molecular weight
MZB marginal zone B cells
MZP marginal zone progenitor cells
NES nuclear export signal
NK natural killer
NLS nuclear localization signals
NSG NOD-SCID IL2R$\gamma^{-/-}$
OBF OCT binding factor
OCT octamer-binding transcription factor
P-Sp paraaortic splanchnopleura
PAX5 paired box gene 5
PBS phosphate buffered saline
PBX pre B cell leukaemia homeobox
PCNA proliferating cell nuclear antigen
PCR polymerase chain reaction
PDGF platelet-derived growth factor
PE phycoerythrin
PEST proline-glutamic acid-serine-threonine domain
PFA paraformaldeyde
PI3K phosphatidyl-inositol 3-kinase
PIP_3 phosphatidyl-inositol-3,4,5-Trisphosphate
Plat-E Platinum-Ecotropic retroviral packaging-HEK
plpC poly-inosinic-poly-cytidylic acid
qRT-PCR quantitative RT-PCR
RAG Recombination activating gene
RISC RNA-induced silencing complex
RNA ribonucleic acid
RPMI Roswell Park Memorial Institute
RSS recombination signal sequences
RT-PCR reverse transcription PCR
rtTA reverse tetracycline-controlled transcriptional activator
RUNX runt related transcription factor
SCC squamous cell carcinoma

SCF stem cell factor
SCID severe combined immunodeficiency
SCL stem cell leukaemia
SHM somatic hypermutation
siRNA small interfering RNA
SLC surrogate light chain
shRNA short hairpin RNA
SOX SRY-related HMG-box
SRC
SSB single-strand DNA breaks
STAT signal transducer and activator of transcription
SYK spleen tyrosine kinase
T1 transitional type 1
T2 transitional type 2
TGFβ transforming growth factor β
TIE2 TEK tyrosine kinase
TRE tetracycline responsive element
TPO thrombopoietin
UNG uracil DNA glycosylase
V variable region
VE-Cadherin vascular endothelial cadherin
VEGF vascular endothelial growth factor
WHO World Health Organization
WT wild type
XLA X-linked agammaglobulinemia
XID X-linked immunodeficiency

ABSTRACT

The University of Manchester
Sara Cuvertino
Degree of Doctor of Philosophy

A potential role for SOX7 in leukaemogenesis

September 2014

The transcription factor SOX7 plays a central role in the development of the cardio-vascular system. Its expression is tightly regulated at the onset of blood and endothelium specification and its sustained expression in immature blood cells blocks differentiation and promotes proliferation, leading to the accumulation of immature blood cells. This striking effect on early blood specification led to the hypothesis that the mis-expression of *Sox7* may confer a proliferative or survival advantage to adult cells and that *Sox7* might be implicated in the emergence or maintenance of leukaemia initiating cells.

Analysis of *SOX7* expression in multiple cases of human leukaemia revealed that this transcription factor was significantly and specifically expressed in B-Cell Acute Lymphoblastic Leukaemia (B-ALL). Based on this observation, I first investigated in a mouse model the consequences of *Sox7*-enforced expression on the homeostasis of adult haematopoiesis and during the B cell differentiation. *In vitro*, *Sox7*-enforced expression impaired the differentiation of B cells and induced the proliferation of immature progenitor cells. *In vivo*, *Sox7*-enforced expression also blocked B cell differentiation, caused splenic enlargement and induced the accumulation of fibrotic fibres in the bone marrow, all signs of a pre-leukaemic stage. To investigate the role for SOX7 in the formation and maintenance of human leukaemia, I knocked-down the expression of *SOX7* in B-ALL cell lines. This resulted in a significant decrease in the proliferation of these B-ALL cells *in vitro*. Upon engraftment *in vivo*, I observed that the down-regulated expression of *SOX7* induced a significant increase in survival rate.

Altogether, the findings presented in this thesis demonstrate for the first time the correlation between *SOX7* expression and the B-ALL and the maintenance of immature progenitor cells by *Sox7*-enforced expression.

DECLARATION

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

COPYRIGHT STATEMENT

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy, in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations and in The University's policy on presentation of Theses

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor Dr Valerie Kouskoff for giving me the opportunity to work in her laboratory on this project, for her precious time spent on my preparation and training during these 4 years and for her great advises and discussions.

A big thank you goes to Dr Georges Lacaud for his input in my project and for the support during my PhD. Also, I would like to thank my advisor, Dr Tim Somerville for his valuable advises.

A special thank you goes to present and past members of Stem cell biology and Stem cell haematopoiesis groups for the great time spent together in the office and in the lab, for their advise and support during my PhD. Thanks for making these year so great!

This thesis couldn't be done without the help from the institute facilities. I want to thank all the members of the CRUK-MI facilities, in particular FACS, histology, MBCF and BRU facilities for their help and patient for dealing with my requests.

A huge and special thanks goes to Mauri that always support me and never let me down. Thanks for taking care and understand me during this 4 intense years of study. Thanks for keeping me happy with your music!

Then, I want to thank my mum and my brother for being always with me, but also my grandparents and my family for cheering me up.

A special thanks goes also to my Italian friend Maria for all the support during these years and nice time spent together in UK. Also thanks to Giu, Giulietta, Meli and Giulia mora, even if spread around the world we are always together!

Finally, last but not least, a great thank you goes to my friends, Elli, Hadir, Eva, Marija, Becky and Sharmin for the great friendship and wonderful time spent together during these 4 years in Manchester.

Grazie a tutti!

ABOUT THE AUTHOR

Sara Cuvertino was born in 1986 in Turin, Italy. In 2005, Sara studied Biotechnology at the University of Turin in her hometown. In 2008, she continued her study getting a Master's degree in Medical Biotechnologies at the University of Turin. During her undergraduate studies, Sara attended the Laboratory of Cancer Genetics at the IRCC Institute (Candiolo, Turin) working with Dr Maria Flavia Di Renzo and Dr Nadia Dani on the role of the *MET* oncogene in osteoblast transformation.

During her Master course, she took part in a summer studentship program in the laboratory of Dr Dario Alessi at the Phosphorylation Unit in Dundee (Scotland, UK). During the 2 months period, she worked with Dr. Beatrice Maria Filippi and investigated how the AMPK-related kinases are activated by the tumour suppressor kinase LKB1.

Since the beginning of her university study, Sara found really attractive the study of tumour biology for its rapid evolution and for the great contribute that cancer research can give to improve therapies. Grateful to Dr Alessi and her Italian supervisor Dr Di Renzo for their support, Sara decided to apply for the PhD fellowship at the Cancer Research UK-Manchester Institute (CRUK-MI). After being accepted, Sara moved to Manchester, UK, in September 2010 to commence the project presented in this thesis in the Stem Cell Haematopoiesis group of Dr Valerie Kouskoff at the CRUK-MI.

To my family

Chapter 1

1. Introduction

1.1. Haematopoietic development

Haematopoiesis is the process by which all the different cell lineages that form the blood and immune system are generated from a common multipotent stem cell (Cumano and Godin, 2007). During embryogenesis the haematopoietic, vascular and cardiac systems are the first to develop because they become soon essential for the transportation of oxygen to developing tissues of the rapidly growing embryo (Copp, 1995; McGrath and Palis, 2005).

The gastrulation resulting in the formation of the mesoderm germ layer that, in the mouse embryo, begins at embryonic day (E) 6.5 is the first process leading to the generation of blood cells in the embryo (Gardner, 1979; Tam and Behringer, 1997). The mesoderm cells migrate away from the posterior region of the primitive streak and localize between the primitive ectoderm and visceral endoderm germ layers forming blood islands (Lawson et al., 1991; McGrath and Palis, 2005). These cells will form various extra-embryonic structures, such as the chorion, the amnion, the allantois and the yolk sac (Lawson et al., 1991), that are necessary for normal embryonic development (McGrath and Palis, 2005).

Mammalian haematopoiesis is subdivided into two main systems, the embryonic primitive system and the definitive system, of which only the latter persists in the adult life (Orkin and Zon, 2008). During primitive haematopoiesis, the first blood cells are

generated around E7-7.5 in the blood islands of the extra-embryonic yolk sac (McGrath and Palis, 2005; Palis et al., 1995). These first circulating cells are primitive erythroid cells, which are large nucleated erythroblasts carrying out oxygen and nutrients to the developing tissues. These primitive erythroid cells express both embryonic (ζ, βH1 and εy) and adult (α1, α2, β1 and β2) haemoglobin genes (Brotherton et al., 1979; McGrath and Palis, 2005). In addition to primitive erythroid cells, the generation of some megakaryocytes and macrophages also takes place during primitive haematopoiesis (Lacaud et al., 2001; Palis et al., 1999). This first transient wave lasts about 48 hours (Lacaud et al., 2001), and then, as the embryo increases in size, this early circulatory system is replaced by the definitive system between E8.25 and E10.5 (Keller et al., 1993; Palis et al., 1999). The definitive system is characterized by a first transient wave occurring in the blood island that give rises to erythroid-myeloid progenitors (EMP) (Bertrand et al., 2007; Medvinsky et al., 2011). This population migrates into the liver where it is involved in the production of foetal definitive erythrocytes. Between E8.5 and E11.5 in the mouse embryo, haematopoiesis shifts to intra-embryonic para-aortic splachnopleura (P-Sp) site, this site later develops into the aorta-gonads-mesonephros (AGM) region (Dzierzak and Speck, 2008). This definitive haematopoietic system gives rise to the myeloid, lymphoid and definitive erythroid lineages, including the haematopoietic stem cells (HSCs) (Cumano and Godin, 2007; Dzierzak, 2005). The first evidence of intra-embryonic haematopoiesis was revealed from experiments conducted in avian embryos. Transplanting chick yolk sac cells into quail embryo before circulation resulted in production of quail derived haematopoietic cells suggesting that the extra-embryonic haematopoiesis moved to the embryo proper (Dieterlen-Lievre, 1975). Only cells deriving from the AGM region retain the long-term reconstitution (LTR) ability upon engraftment in adult recipient mice (Medvinsky and Dzierzak,

1996). In particular, only clusters associated with the ventral part of the dorsal aorta contains HSC with LTR capacity (Taoudi and Medvinsky, 2007). In contrast, at E10.5 yolk sac-derived cells do not retain LTR ability but can produce B and T cells suggesting the presence of a lympho-myeloid progenitor population generated before the emergence of HSCs in AGM at E10.5 (Boiers et al., 2013; Yoshimoto et al., 2011). From E11.5, HSCs are found in the yolk sac, whether they are *de novo* generated or transported there through the circulation is still a matter of debate.

Thereafter, from E11.5 until birth, the liver becomes progressively the main site of haematopoiesis where progenitors generated previously at other sites migrate, expand and differentiate. After birth and during all adult life, haematopoiesis occurs in the bone marrow (Figure 1.1) (Costa et al., 2012a; Jagannathan-Bogdan and Zon, 2013; Lacaud et al., 2004a; Ottersbach and Dzierzak, 2009).

Interestingly, some studies revealed a role for the placenta in definitive haematopoiesis and in the generation of HSCs. HSCs were detected in this site around E10.5 and expanded until E12.5-E13.5, retaining the ability to reconstitute recipient mice (Gekas et al., 2005; Ottersbach and Dzierzak, 2005). Additionally, other sites, such as the chorion, allantois, vitelline and umbilical arteries have been described as source haematopoietic activity (Downs et al., 2004; Gordon-Keylock et al., 2013; Inman and Downs, 2007). New sites for haematopoietic activities have recently been reported in the mouse head vasculature between E10.5-11.5 (Li et al., 2012b) and in haemogenic endocardium (Nakano et al., 2013). However, the relative contribution of each of the above sites to the final pool of adult HSCs is not yet clear and remains to be determined.

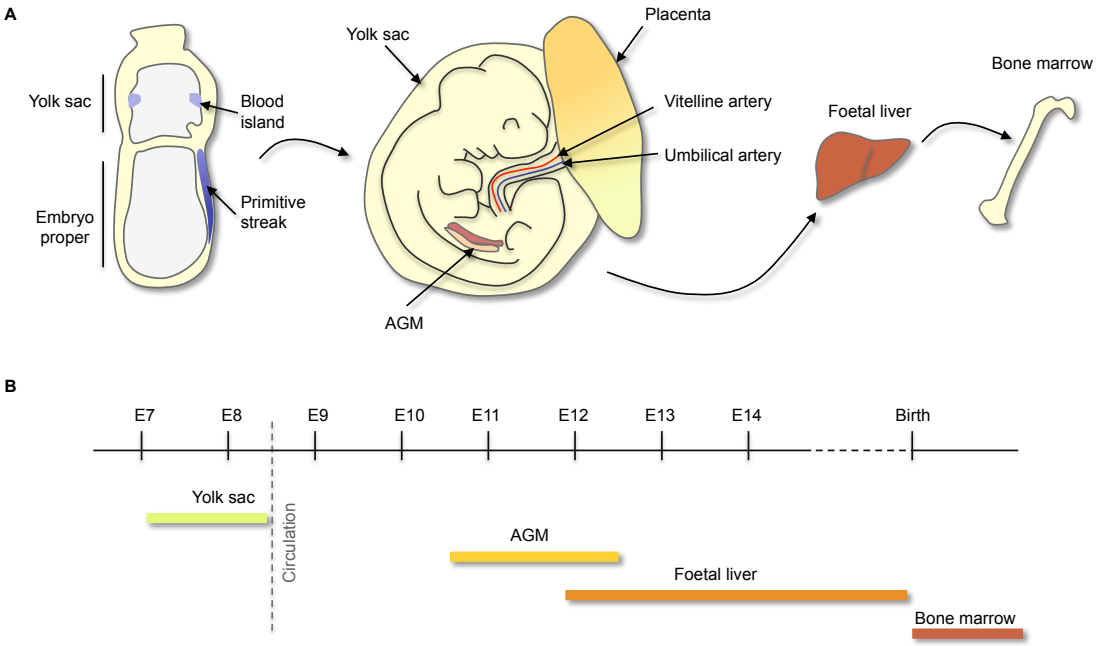


Figure 1.1. Anatomical sites of haematopoiesis.
A) Mesodermal cells generate haemangioblasts through the posterior primitive streak, which migrate in the blood island in the yolk sac. At E7.5 primitive haematopoiesis is initiated in the extra-embryonic part of the embryo. Blood cells emerge at E10.5 in the AGM region, placenta, vitelline and umbilical arteries. From E12 until birth, the foetal liver becomes the main site of haematopoiesis. Finally, after birth and during the adult life the bone marrow replace the foetal liver for blood production. B) Timeline of haematopoietic activity during mouse development. AGM: aorta-gonads-mesonephros; E: embryonic day.

1.2. Blood cell formation

The close spatial and temporal development of the haematopoietic and endothelial cells, that occurs in parallel in the yolk sac, observed in chick embryo by Sabin and colleague led to the hypothesis that a unique precursor called angioblast gives rise to endothelial cells, blood plasma and erythrocytes (Sabin, 1920). Later on, Murray substituted the term angioblast with haemangioblast indicating a common mesodermal progenitor cell that retains the potential to give rise to both lineages: haematopoietic and endothelial (Downs, 2003; Nishikawa, 2012).

In support of this hypothesis haematopoietic cells and blood vessels express many common surface receptors, such as the foetal liver kinase 1 (FLK1), the FMS (colony stimulating factor 1 receptor) like tyrosine kinase 1 for vascular endothelial growth factor (FLT1), the tyrosine kinase with immunoglobulin-like and EGF (epidermal growth factor) -like domains 1 (TIE1) and TEK (endothelial-specific receptor tyrosine kinase) receptor tyrosine kinase for angiopoietin (TIE2) (Downs, 2003; Lacaud et al., 2004b). In addition, the deletion of *Flk1* compromised the formation of both endothelium and haematopoiesis inducing early death of homozygous embryos between E8.5 and E9.5. At E7.5, both yolk sac blood island and blood vessels were not observed in embryo deficient for *Flk1* (Shalaby et al., 1995). Similarly, deficiency in the ETS transcription factor *Etv2* leads to a complete absence of both vasculature and blood cells (Lee et al., 2008; Wareing et al., 2012).

The *in vitro* differentiation of embryonic stem (ES) cell generates three-dimensional structures called embryoid bodies (EBs). Cells within EBs give rise to haematopoietic, endothelial, neuronal and muscle cells (Keller, 1995). Using the ES/EB system, Choi and colleagues identified the *in vitro* equivalent of the haemangioblast, termed the blast colony-forming cell (BL-CFC). Based on this model, the BL-CFC

appears as a transient population before the development of any other haematopoietic lineage precursor around day 2.5-3.5 of EB differentiation (Choi et al., 1998). Indeed, the BL-CFCs express FLK1 and the mesodermal marker *Brachyury* and give rise to blast colonies with endothelial, haematopoietic and vascular smooth muscle potential (Fehling et al., 2003). These data were also supported by *in vivo* studies using *GFP-Bry*$^{+/-}$ transgenic embryos where the haemangioblast population (BRY^{+}FLK1^{+}) was identified in the posterior region of the primitive streak of the embryo (Huber et al., 2004).

At later stage of development, the presence of clusters of haematopoietic cells and HSCs closely associated with the endothelial cells of the dorsal aorta in the AGM region, the vitelline and umbilical arteries and allantois/placenta region led to the hypothesis of the existence of endothelial cells with haematopoietic potential, termed haemogenic endothelium (Figure 1.2) (Ciau-Uitz et al., 2014; Jaffredo et al., 1998). Labelling chick endothelial cells with acetylated low-density lipoproteins (Acetyl-LDL), it was possible to demonstrate the derivation of CD45^{+} haematopoietic cells from Acetyl-LDL positive endothelial cells in the dorsal aorta. Moreover, infecting aortic endothelial cells with *LacZ* retroviral vector, Jaffredo and colleagues also demonstrated that a haemogenic population is expressed transiently in the aorta (Jaffredo et al., 2000). In addition, Nishikawa and colleagues demonstrated the generation of haematopoietic colonies *in vitro* from VE-cadherin^{+} endothelial cells extracted from yolk sack of E9.5 mouse embryos (Nishikawa et al., 1998). Using transgenic mice carrying the *Gfp* reporter gene under the control of *Ly-6A* (*Sca1*) gene permitted to mark all functional HSCs. In the AGM region GFP^{+} cells were localized to the endothelial layer of the wall of the dorsal aorta confirming the generation of haematopoietic cells from endothelial cells (de Bruijn et al., 2002).

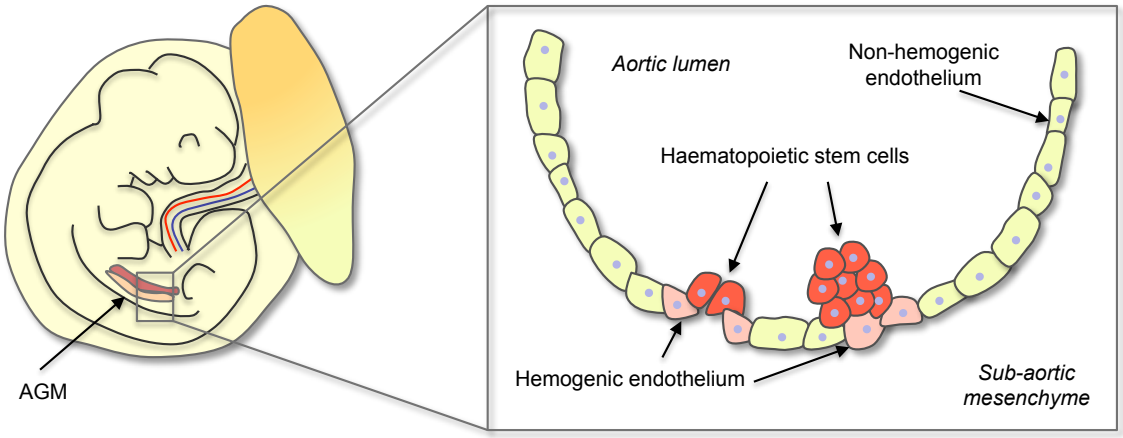


Figure 1.2. Generation of haematopoietic cells in the dorsal aorta.
Haemogenic endothelial cells line wall of the dorsal aorta in the AGM region. Subsequently, haemogenic endothelial cells differentiate and generate the haematopoietic cluster containing haematopoietic stem cells. Finally, haematopoietic progenitor cells are released in the circulation. AGM: aorta-gonads-mesonephros.

Experiments performed using inducible *VE-cadherin-Cre* line permitted to trace the endothelial cells in AGM region. As already reported, VE-cadherin^{+} cells generated HSCs and the ability to produce haematopoietic cells decreased from the early to later mesenchymal populations (Zovein et al., 2008).

Additionally, Oberlin and colleagues observed the origin of haematopoietic cells from the aorta in human embryo. Human CD34^{+}CD31^{+}CD45^{-} endothelial cells were sorted from the dorsal aorta and after culturing on stromal cells generated both lymphoid and myeloid cells (Oberlin et al., 2002).

Recent studies have now demonstrated that the haemangioblast also give rise to haematopoietic precursors through the formation of an intermediate haemogenic endothelial stage (Eilken et al., 2009; Lancrin et al., 2009). This intermediate cell population is characterized by the expression of cKIT and TIE2, is SCL-dependent and gives rise to the primitive and definitive haematopoietic progenitors (Figure 1.3). The generation of definitive haematopoietic cells, characterized by the expression of cKIT,

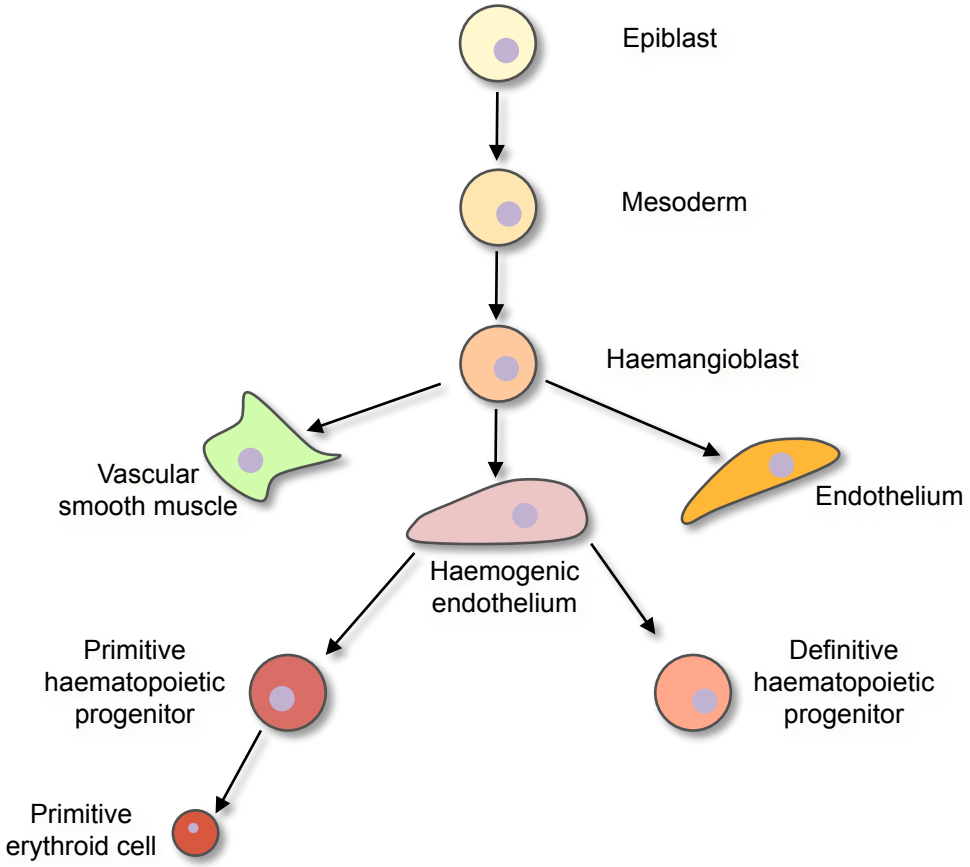


Figure 1.3. Scheme of haemangioblast differentiation.
The haemangioblast originates from the mesoderm cell and represents the common precursor for endothelial, haematopoietic and vascular smooth muscle lineages. The haemogenic endothelium gives rise to the two distinct waves of haematopoiesis, the primitive and definitive populations.

CD41 and CD45 and the down-regulation of TIE2, requires the transcription factor RUNX1. In contrast, primitive erythroid cells, that are CD45 negative, are RUNX1-independent (Lancrin et al., 2010).

Live imaging technology together with confocal microscopy was used to observe the emergence of haematopoietic cells in real-time in the aorta of *Ly6A-GFP* mouse embryo (E9-E11). GFP^+ haematopoietic cells observed to bud off from ventral $CD31^+GFP^+$ endothelial cells (Boisset et al., 2010). Few months later, Lam and colleagues showed that RUNX1 expression tracked the emergence of haematopoietic cells from endothelial cells in the dorsal aorta of zebrafish *Runx1-GFP* embryos. Using *Runx1-GFP/Kdr1-mCherry* double mutant, they specifically observed the expression of RUNX1 in $mCherry^+$ endothelial cells that subsequently develop haematopoietic cells (Lam et al., 2010).

In addition, other studies established a hierarchical organization during HSC development in the dorsal aorta. VE-$cadherin^+CD45^-CD41^+$ cells were identified as pre-HSC type I and were distributed throughout the dorsal aorta. These cells then matured into definitive HSCs via the acquisition of CD45 expression that characterized a second population called pre-HSC type II (Rybtsov et al., 2011).

Past these stages of early haematopoietic development, the HSCs, characterized by the properties of self-renewal and multi-potentiality, reside in the foetal liver, then, in the bone marrow and generate, upon specific signals, multipotent progenitors (MPP). These progenitors are part of the so-called LSK compartment characterized by the expression of SCA1 and cKIT and the absence of the lineage differentiation markers (Spangrude, 1994). Multipotent progenitors retain the ability to generate both lymphoid and myeloid cells. The exact progression for the development of the haematopoietic lineages generates strong debates, however to date a classical view of the

haematopoietic hierarchy remains commonly accepted (Ema et al., 2014). In this classical hierarchy, MPP cells generate common myeloid precursor (CMP) and common lymphoid precursor (CLP), which later differentiate into more specific lineages (Orkin and Zon, 2008; Seita and Weissman, 2010). CMP gives rise to two populations: megakaryocyte-erythroid progenitors (MEP) and granulocyte-macrophage progenitors (GMP). The former generates megakaryocytes and erythrocytes while the latter gives rise to granulocytes and macrophages. In contrast, CLP generates dendritic cells, natural killer (NK) cells, T and B lymphocytes (Figure 1.4) (Cumano and Godin, 2007; Orkin, 2000). In addition, both CMP and GMP can give rise to dendritic cells compared to MEP that cannot (Traver et al., 2000; Manz et al., 2001).

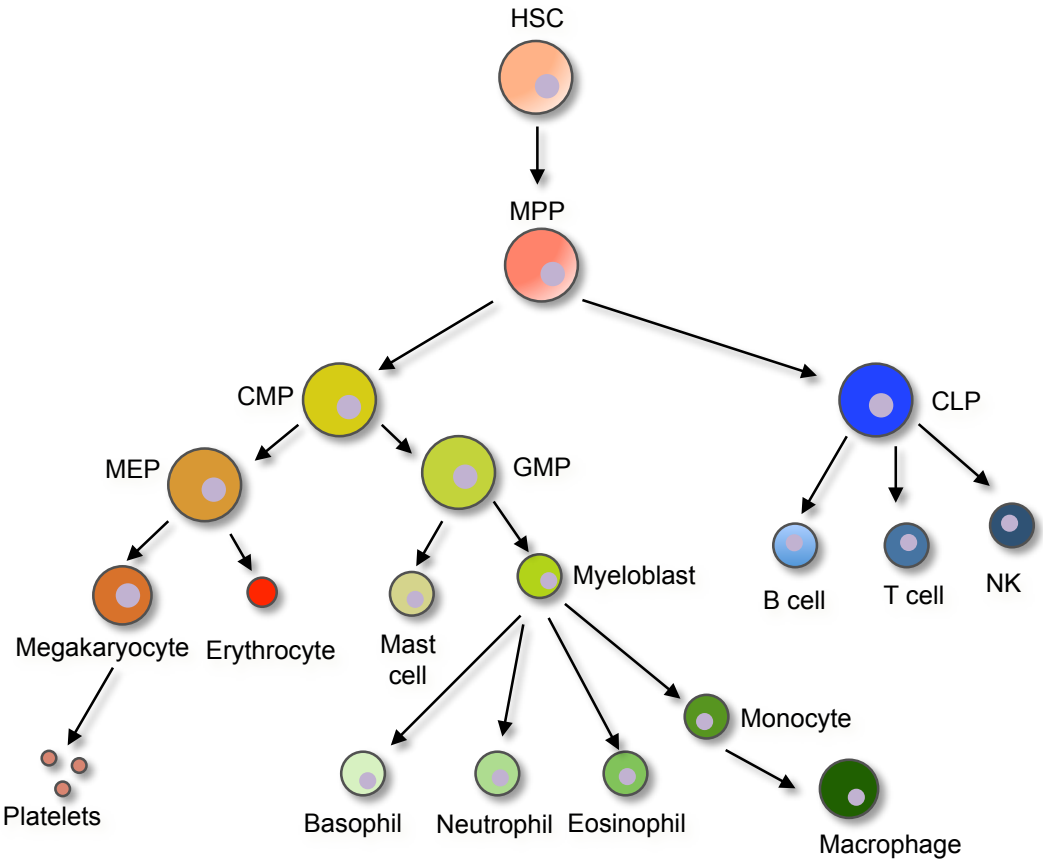


Figure 1.4. Scheme of haematopoietic hierarchy.
Haematopoietic stem cells (HSC) generate multi-potent progenitors (MPP) that give rise to common myeloid (CMP) and lymphoid (CLP) progenitors. Subsequently, CMP gives rise to two different populations: megakaryocyte-erythroid progenitors (MEP) and granulocyte-macrophage progenitors (GMP) that generate mature myeloid cells while CLP generates lymphoid cells.

All these populations were identified through analysing the expression of specific cell surface markers combined with functional assays. In the mouse haematopoietic hierarchy, the multipotent progenitors differ from HSCs for the expression of the CD34 marker. While cells progress to later stages, the oligopotent progenitors lose their expression of SCA1 acquiring the expression of IL7Rα and FcγR for CLP and CMP, respectively. In particular, different levels of FcγR expression characterize the myeloid compartment. Absence, low or high expression levels of this marker distinguish MEP, CMP and GMP, respectively (Orkin, 2000). Similarly to the mouse haematopoietic development, similar populations characterize the human haematopoietic hierarchy with discrimination between subsets through the expression of different cell surface markers (Seita and Weissman, 2010). Human HSCs are characterized by the $LIN^-CD34^+CD38^-CD90^+$ immuno-phenotype. Multipotent progenitors lose the expression of CD90, while oligopotent progenitors acquire the expression of CD38. The CLP population is characterized by the expression of the CD10 marker while the myeloid compartment expresses IL3Rα. Similar to FcγR expression in mouse, low and high expression levels of IL3Rα distinguish CMP and GMP. In contrast, MEP is negative for the expression of IL3Rα (Figure 1.5). Finally, the generation of mature blood cells progress via the maturation of lineage restricted progenitors.

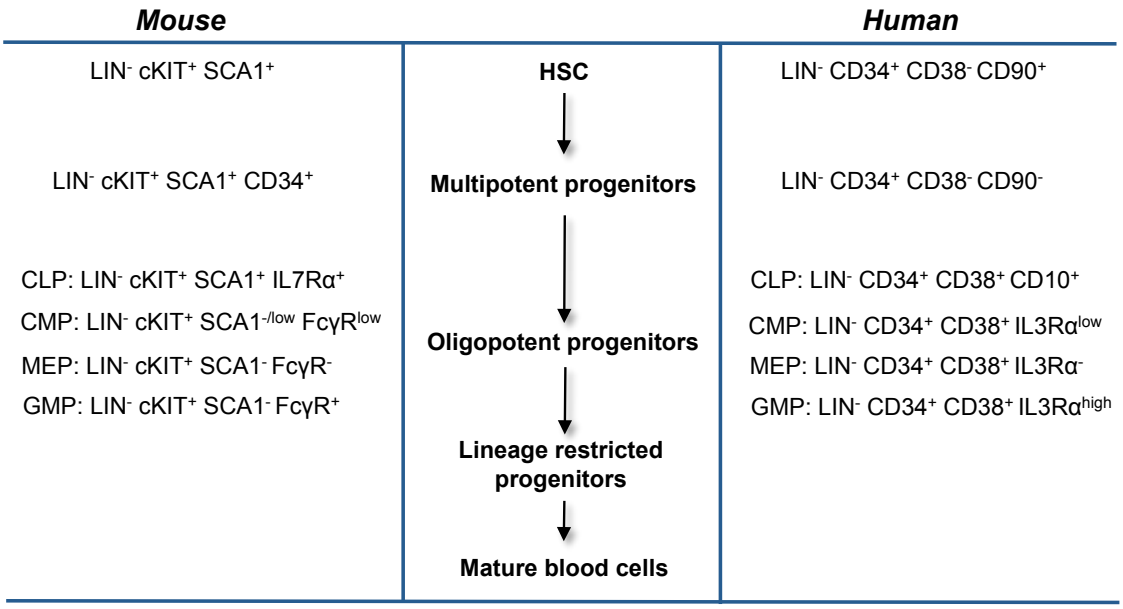


Figure 1.5. Scheme of cell surface markers for haematopoietic progenitors in mouse and human haematopoietic development.
Haematopoietic stem cells (HSC) are located at the top of the haematopoietic hierarchy. HSC are characterized by the properties of self-renewal and multipotency. During the maturation process, HSCs progressively lose their self-renewal capacity and become more committed to mature blood lineages. Multipotent progenitors (MPP) differentiate into oligopotent progenitors that subsequently generate mature blood cells through lineage restricted progenitors.

1.3. Introduction to B lymphocytes: roles and functions

Lymphocytes are small round cells of about 9 microns in size and represent 20% of the white blood cells in the peripheral blood. They are classified as mononuclear cells as they contain a round non-lobulated nucleus and show a high nucleus:cytoplasm ratio. The life span of lymphocytes varies between few days and months or years for memory cells (Sprent and Tough, 1994).

The term B lymphocyte refers to a group of cells that express on their cell surface different immunoglobulin (Ig) receptors recognizing specific antigenic epitopes. Antigenic epitopes are molecules from the body (self) or from pathogens (non-self) that are toxic for the organism and that can be recognized and bound by a specific antibody. Indeed, the term antigen originates from antibody generator (Khan et al., 2013).

The immunoglobulins consist of two heavy chains and two light chains. The former is composed of one variable region and three constant regions while the latter consists of one constant and one variable region (Woof and Kerr, 2004). During B cell differentiation the rearrangement of the heavy chain precedes the light chain rearrangement. The rearrangement between the D and J region occurs in the pro B cells while the formation of the VDJ region occurs in the late pro B cells (Gonzalez et al., 2007). Once the heavy chain rearrangement is completed, cells express the pre-BCR complex on their cell surface. This complex consists of the heavy chain; Igα-Igβ dimer and the two surrogate light chain molecules, VpreB and λ5 (Kitamura et al., 1991). Cells that expressed functional pre-BCR proceed with maturation. Once activated, the main role of B lymphocytes is the production of immunoglobulins or antibodies produced to kill pathogens (Obino and Lennon-Dumenil, 2014).

As part of the humoral immune response, B cells regulate or mediate other functions important for the immune system. Acting as antigen-presenting cells, they can

initiate T-cell immune response. B cells are also involved in cytokines production, the regulation of lymphoid tissue organization and wound healing. Roles in transplanted tissue rejection and tumour immunity have also been described (Figure 1.6) (Bao et al., 2014; LeBien and Tedder, 2008).

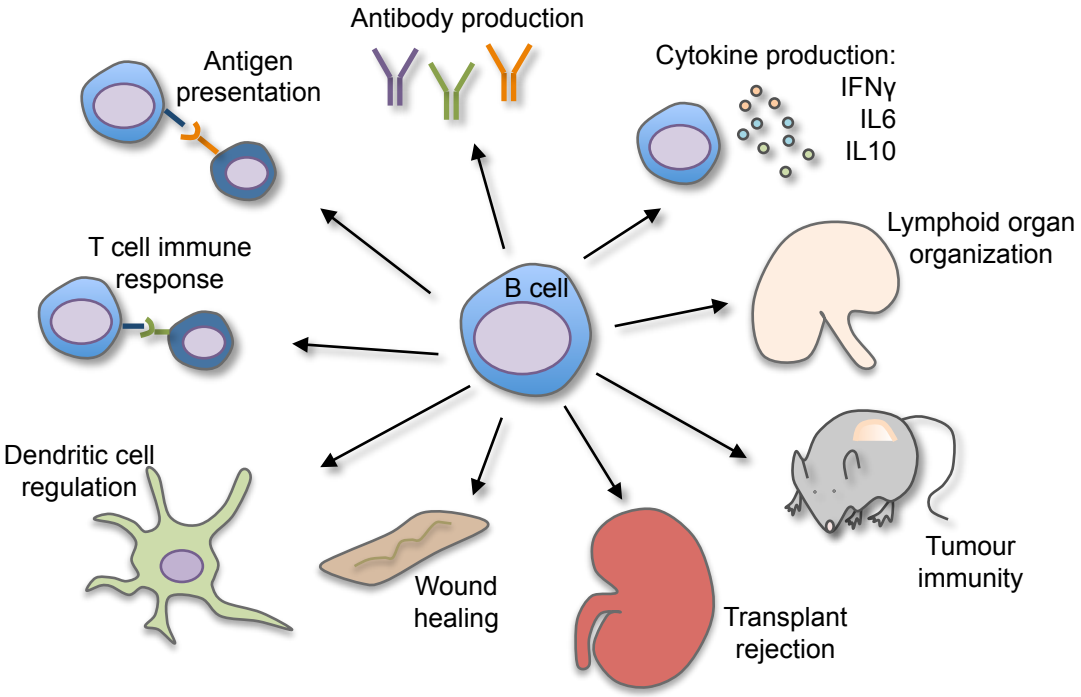


Figure 1.6. Scheme of B cell functions.
The two main functions of B cells are the production of antibodies and the presentation of the antigen to T cells. Once activated, B cells produce cytokines that initiate the T cell immune response or regulate dendritic cells. B cells are involved in the lymphoid organ organization and wound healing but also in the rejection of transplanted organs or tumour immunity.

1.4. Stages of B cell development in mouse

B cell development initiates in the foetal liver during mid and late gestation. Then, it moves to the bone marrow after birth where it continues throughout life. Only 10% of immature B cells leave the bone marrow and enter into the periphery (Sprent and Tough, 1994). These immature B cells enter into the spleen through the blood circulation, where they develop into mature B cells (Loder et al., 1999).

Each stage of the mouse B cell development can be identified by the expression of specific surface markers and genes (Figure 1.7). Mouse B cells can be sub-divided into two sub-populations based on Hardy's classification: less mature cells expressing the protein tyrosine phosphatase B220 (CD45R) and CD43 (leukosialin) while the more mature cells express only the surface marker B220 (Hardy and Hayakawa, 1991).

Then, the $B220^{+}CD43^{+}$ population can be further sub-divided into three different fractions according to their expression of BP-1 and CD24 (heat stable antigen, HSA). BP-1 is a type II trans-membrane protein that retains glutamyl-aminopeptidase activity while CD24 is a cell adhesion molecule, anchored to the cell surface through a glycosyl-phosphatidylinositol (Wu et al., 1989). The fraction A represents the pre-pro B cells and lacks the expression of both BP-1 and CD24 markers. In fraction B, cells called pro B cells initiate the immunoglobulin heavy chain (IgH) D-J rearrangement and express only the CD24 marker. The last fraction within the immature $B220^{+}CD43^{+}$ population is the fraction C representing the late-pro B cells. In this sub-population cells initiate the IgH V-DJ rearrangement and express both BP-1 and CD24 markers. Then, B cell development proceeds into more mature steps. Three different sub-populations can also be identified in the $B220^{+}CD43^{-}$ fraction according to the expression of IgM and IgD. Fraction D contains small pre B cells characterized by the absence of both immunoglobulins. Then, cells proceed to fraction E in which the immature B cells

express only IgM, having successfully rearranged a functional light chain to form a cell surface expressed immunoglobulin. Finally, the last step of B cell differentiation in the bone marrow is represented by the fraction F. This fraction contains mature B cells that express both IgM and IgD (Hardy and Hayakawa, 1991; Hardy et al., 2007; Miosge and Goodnow, 2005).

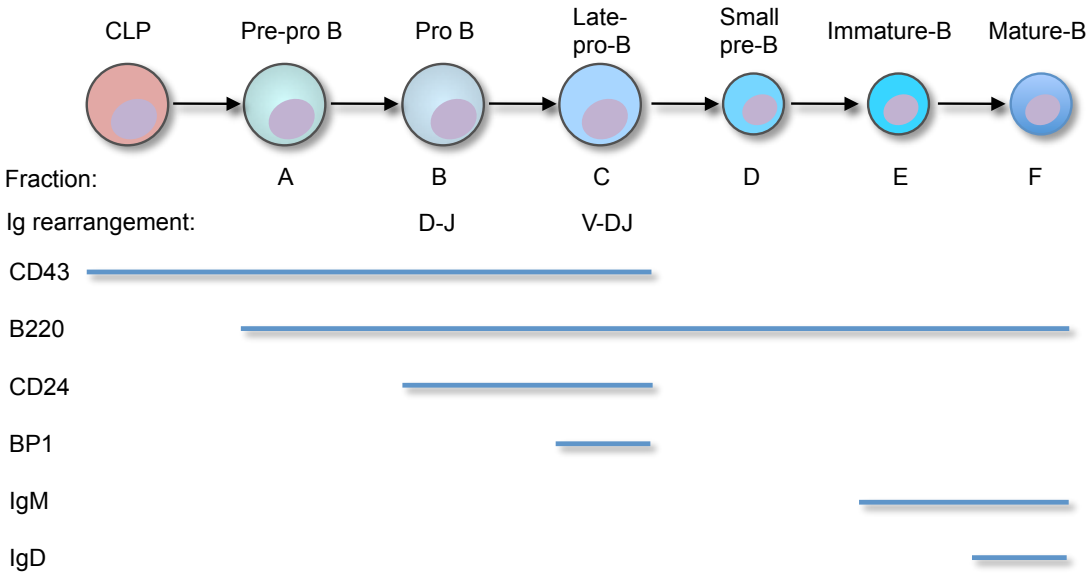
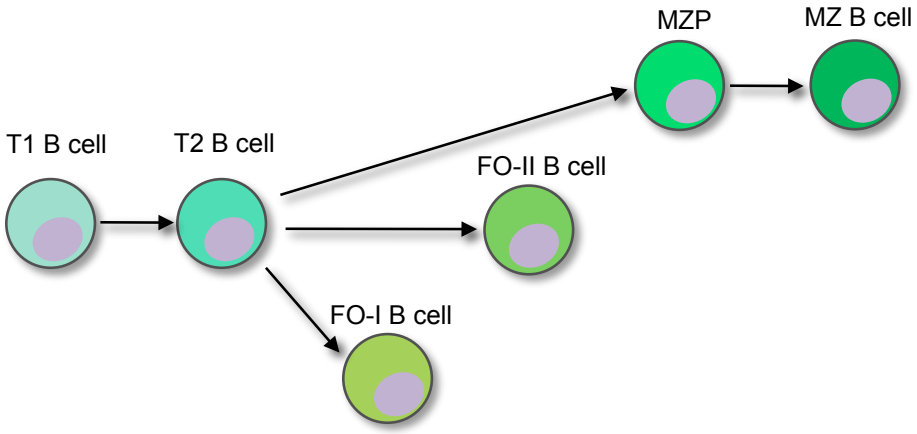


Figure 1.7. Scheme of mouse B cell development.
B cells are generated by common lymphoid progenitors (CLP). The immature populations ($B220^{+}CD43^{+}$) are subdivided based on the expression of CD24 and BP1 markers. The mature ($B220^{+}CD43^{-}$) populations instead are subdivided based on the expression of IgM and IgD.

The mouse peripheral B cell maturation occurs in the spleen (Figure 1.8). Immature B cells $IgM^{high}IgD^{-}$ move from the bone marrow and enter the red pulp of the spleen. Several studies have been performed to identify subsets of immature and mature spleen B cells. Loder and colleagues classified two immature transitional populations by staining with IgM, IgD, CD23 and CD21 antibodies. Once in the spleen, the IgM^{high} immature B cells are called transitional type 1 (T1) where they acquire low level of IgD, CD23 and CD21 markers. After about one day, T1 cells move into the white pulp of the spleen. These cells, called transitional type 2 (T2), acquire high level of IgD and CD21 and intermediate level of CD23 (Loder et al., 1999). Allman and colleagues have proposed another classification by introducing the expression of AA4.1 marker to discriminate three immature transitional B cells. The transitional type 1 (T1) population is IgM^{high}, $CD23^{-}$ and $AA4.1^{+}$. The transitional type 2 (T2) cells are IgM^{high}, $CD23^{+}$ and $AA4.1^{+}$ while the transitional type 3 (T3) cells are IgM^{low}, $CD23^{+}$ and $AA4.1^{+}$ (Allman et al., 2001).

Then, follicular B cells proceed through B cell maturation expressing intermediate level of IgM, high level of CD21 and CD23 but no AA4.1. Finally, another cell population resides in the white pulp and in particular in the marginal zone where it takes its name from, marginal zone B cells (MZB). The MZB cells, generated by the marginal zone progenitor cells (MZP), are characterized by high expression of both IgM and CD21 markers and lack of CD23 expression (Loder et al., 1999; Pillai and Cariappa, 2009; Srivastava et al., 2005).



IgM	+	high	high	high	high	high
IgD	-	high	low	high	high	low
CD23	-	+	+	+	+	-
CD21	low	mid	mid	mid	high	high
AA4.1	+	+	-	-	-	-

Figure 1.8. Scheme of mouse peripheral B cell development.
Once naïve cells enter the spleen, they differentiate into transitional cells. There are two populations of transitional cells: T1 and T2 B cells. T2 B cells then can differentiate into follicular type 1, type 2 or marginal zone progenitor cells. The latter generates marginal zone B cells that reside in the spleen red pulp area.

1.5. Stages of B cell development in human

Similar to mouse B lymphopoiesis, human B cell development occurs in the bone marrow where haematopoietic stem cells differentiate into naïve mature B lymphocytes thought a series of maturing B cell progenitors. The second stage of B cell differentiation takes place in secondary lymphoid tissues such as lymph nodes and spleen (Nagasawa, 2006).

The classification of human B cells has been based on the murine ones but differences in biology between the two species resulted in the identification of some specific markers only for human B cells (LeBien and Tedder, 2008). It is possible to

distinguish five different sub-populations during human B cell maturation in the bone marrow (Figure 1.9). The first sub-population is the pro B cell characterized by the expression of CD22 and CD34 and high level of CD38. Pro B cells migrate to a niche rich in IL7-expressing cells promoting the production of several transcription factors such as *PAX5*, *PU.1*, *EBF1*, *E2A*, *RAG1* and *RAG2* (Nagasawa, 2006). Upon recombination between the D and J segments of the IgH locus initiated by *RAG1* and *RAG2*, pro B cells mature into pre B-I cells. These cells express CD19 together with high level of CD10 and CD38. At this stage, the second part of the IgH locus rearrangement occurs between the V segment and the D-J joint. Then, cells with a functional V-DJ exon differentiate into pre B-II cells that express CD20 and high level of CD45. With the acquisition of IgM, $CD10^{+}CD20^{+}$ cells differentiate into immature B cells. Finally, after passing the negative selection process where cells with an auto-reactive BCR are deleted, naïve mature B cells leave the bone marrow and circulate thought the blood circulation. This fully functional population is characterized by the expression of CD22, IgM and IgD and does not depend on the second step maturation in the spleen as it happens in mouse B cell development (Hystad et al., 2007; LeBien, 2000; Perez-Andres et al., 2010).

Naïve mature B cells then move in the spleen. Descatoire and colleagues recently identified four different populations within $IgD^{+}IgM^{+}CD27^{-}$ B cells using first the MEM55 antibody against CD45RB and ABCB1 marker and then CD24 and CD38 markers. The expression of the transporter ABCB1 was detected by the extrusion of the mito-tracker green (MTG) fluorescent dye. The naïve B cells are negative for both markers, $MTG^{-}MEM55^{-}$. Transitional cells are characterized only by the expression of MTG ($ABCB1^{-}$) and CD24 markers. Transitional type 1 cells are separated from transitional type 2 cells based on their high expression of CD38 marker. The last two

sub-populations identified are: $MTG^{-}MEM55^{+}$ and $MTG^{+}MEM55^{+}$. When cultured on OP9hDLL1, these cells acquire the expression of CD27 becoming marginal zone B-like cells. Furthermore, $MTG^{-}MEM55^{+}$ cells, based on their gene expression profile and non-mutated Ig status, represent a marginal zone progenitor population while $MTG^{+}MEM55^{+}$ represent the marginal zone B cells (Figure 1.10) (Bemark et al., 2012; Descatoire et al., 2014; Koethe et al., 2011).

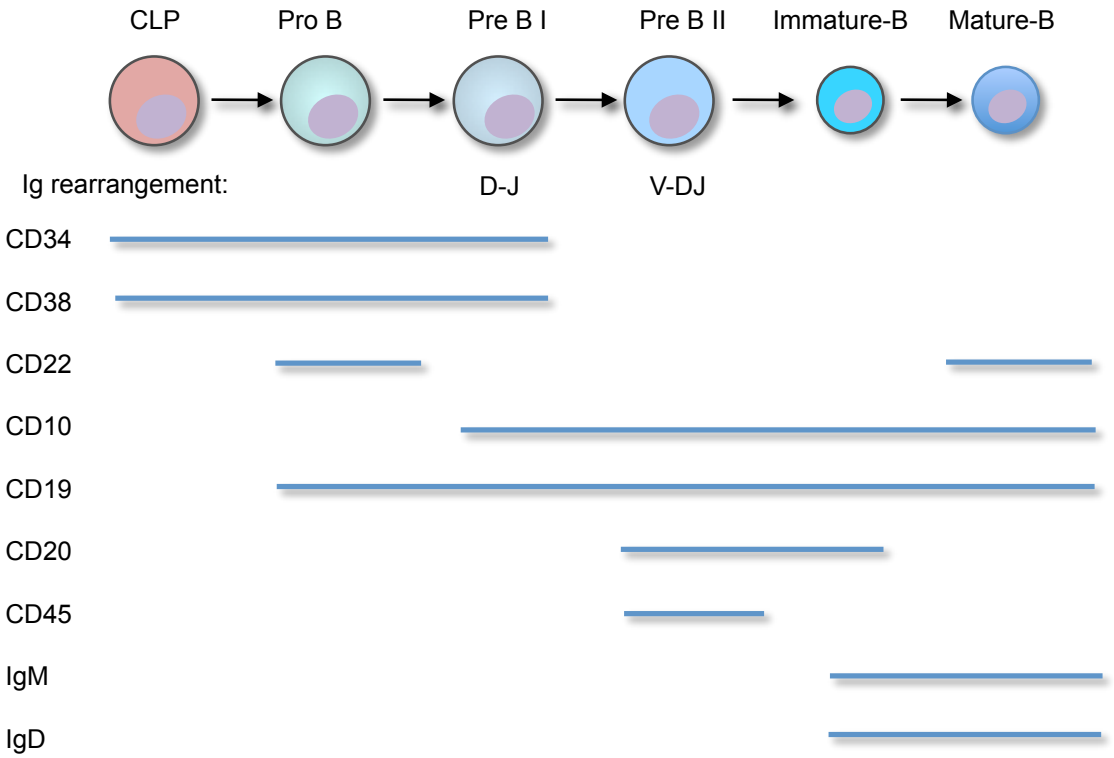
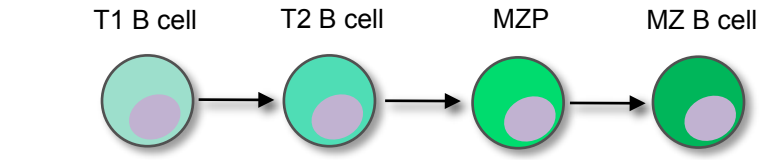


Figure 1.9. Scheme of human B cell development.
B cells are generated by common lymphoid progenitors (CLP). Each step of human B cell development is characterized by different sub-populations that express specific markers.



	T1 B cell	T2 B cell	MZP	MZ B cell
ABCB1	-	-	+	-
CD24	+	+	-	-
CD38	++	+	-	-
MEM55	-	-	+	+
CD27	-	-	-	+
IgD	-	-	-	+
IgM	-	-	-	+

Figure 1.10. Scheme of human peripheral B cell development.
Naïve mature B cells move in the spleen and differentiate into transitional cells. Transitional B cells express CD24 marker. $ABCB1^{+}$ cells represent the marginal zone progenitors while $ABCB1^{-}MEM55^{+}CD27^{+}$ $IgM^{+}IgD^{+}$ cells represent the marginal zone B cells.

1.6. Transcription factors important for B cell development

All the different stages of B cell differentiation described in the previous section are not only characterized by the expression of cell surface markers but are also coordinated by specific sets of transcription factors that drive and control the progress of the B cell maturation (Figure 1.11).

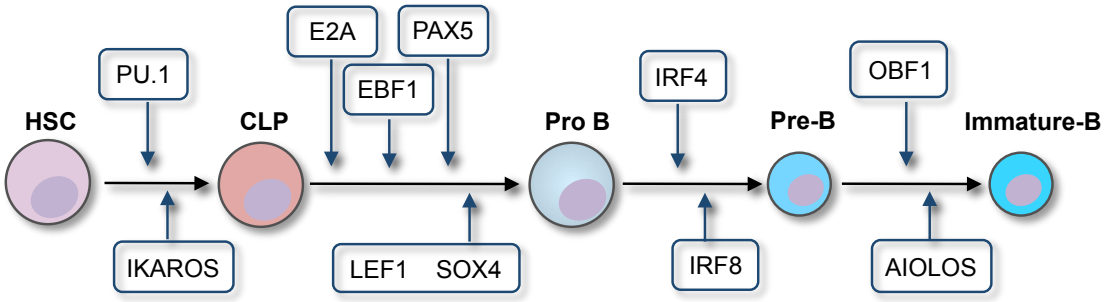


Figure 1.11. Scheme of transcription factors that regulate B cell differentiation. Each stage of B cell differentiation is driven by different genes. The commitment into common lymphoid progenitors (CLP) is induced by PU.1 and IKAROS transcription factors. Later on, the expression of three key transcription factors, E2A, early B cell factor 1 (EBF1), paired box protein 5 (PAX5) occurs in a sequential order to drive the differentiation into Pro B cells. Interferon-regulatory factor 4 and 8 (IRF4/8) control the maturation into Pre-B cells. Finally, OCT (octamer-binding transcription factor)-binding factor 1 (OBF1) and AIOLOS are important for the final step of B cell maturation into immature B cells.

Two important transcription factors, PU.1 and IKAROS, are fundamental for the commitment of the haematopoietic stem cells to common lymphoid progenitors. PU.1 is a transcription factor belonging to the ETS family which expression is important for the early stages of haematopoietic development. PU.1 contains the ETS DNA binding domain at the C-terminal that recognizes the sequence 5'-GGAA/T-3' and two transactivation domains at the N-terminal, one rich in glutamine and the other rich in acidic residues. PU.1 is also characterized by the presence of a PEST domain, important

for protein-protein interaction. A lack of B lymphocytes, T lymphocytes, macrophages and neutrophils has been reported in mice deficient for *Pu.1*. Moreover, these mice died at embryonic day 18.5 (E18.5) (McKercher et al., 1996; Scott et al., 1994). The deficiency in *Pu.1* also induced a reduction in progenitors corresponding to a decreased expression of FLT3 and IL7Rα. This diminution in the expression of IL7Rα is due to the absence of PU.1, which normally binds to the IL7Rα promoter (DeKoter et al., 2002). PU.1 has also been shown to be important for myeloid differentiation. Interestingly, the levels of PU.1 expression are extremely critical for cell fate decision toward B lymphoid or myeloid lineages. Indeed, DeKoter and colleagues, transducing foetal liver haematopoietic *Pu.1*$^{-/-}$ progenitors with PU.1 virus, have described that only low levels of PU.1 promote B cell differentiation while high levels of PU.1 induce myeloid development (DeKoter and Singh, 2000).

The second key transcription factor that drives B cell specification at the very early stages of the haematopoietic hierarchy is IKAROS. *Ikaros*, a key gene for the lymphoid development, encodes for five different isoforms, *Ik-1-5*, all important for lymphocyte development. The IKAROS proteins contain six zinc finger domains, four of which located at the N-terminal and two at the C-terminal. The N-terminal DNA binding domain recognizes the 5'-GGGAA/T-3' motif while the C-terminal domain induces the formation of homodimers or heterodimers (Georgopoulos et al., 1997). Mice homozygous for the deletion of the DNA binding domain in *Ikaros* gene lack T and B lymphocytes and also progenitors. In contrast, erythroid and myeloid lineages were intact in these mutant mice mutant (Georgopoulos et al., 1994). In the mouse strain *Plastic* (*Plstc*), a single point mutation in the *Ikaros*-DNA-binding domain resulted in a severe phenotype with lack of B and T lymphocytes and impaired erythroid and myeloid differentiation. This severe phenotype might be explained by the fact that

IKAROS cannot bind the DNA because of the point mutation and as a consequence all IKAROS complexes are inactivated (Papathanasiou et al., 2003).

Three transcription factors have been found to be essential during the differentiation of CLP towards pro B cells. The first one is the gene that encodes for E2A, basic helix-loop-helix (HLH) transcription factor giving rise to E12 and E47 proteins through alternative splicing. These proteins, together with HEB and E2-2, are called E proteins because of their specific binding to the E box motif, 5'-CANNTG-3'. E proteins can form homodimers and, through the recruitment of co-activators, can act as transcriptional activator. Otherwise, E proteins, when forming heterodimers with Class II bHLH, can act as transcriptional activator or repressor depending on the recruited proteins. Finally, when E proteins heterodimerise with inhibitor of DNA binding (ID) proteins, they cannot bind to the DNA and therefore targeted genes are not transcribed (Kee, 2009). To demonstrate the critical role for E2A during B cell differentiation, the insertion of a *neomycin resistance* cassette into the *E12* exon was used to generate $E2a^{-/-}$ mutant mice. These mice lacked pre-B and mature B cells together with a failure in Ig rearrangement. Moreover, a dramatic decrease in the levels of *Pax5*, *Rag1*, *mb-1* and *CD19* was reported in the foetal liver of those mutant mice (Bain et al., 1994). These results were confirmed by the studies performed by Zhuang and colleagues in which *E2a* inactivation was done by the deletion of the bHLH domain, essential for both DNA binding and dimers formation. $E2a^{bHLH}$ mice showed a high percentage in neonatal death. The surviving mice were smaller in size compared to wild-type controls. Moreover, they lacked B cells while all other blood lineages were intact (Zhuang et al., 1994). Both E2A isoforms, E12 and E47, act in concert to promote B cell differentiation, although E47 is necessary at an earlier stage than E12. Bone marrow cells from mice deficient for *E12* or *E47* have been analysed. A blockage at the

pre pro B stage was observed in *E47*$^{-/-}$ mice while only a slight decrease in pre B cells was reported in *E12*$^{-/-}$ mice (Beck et al., 2009).

The second transcription factor involved in the early stages of B cell differentiation is EBF1. This protein is characterized by a DNA binding domain at the N-terminal that binds the 5'-CCCNNGGG-3' motif, a HLH dimerization domain and a transactivation domain at the C-terminal. The role of EBF1 in B cell development has been shown using *Ebf1*$^{-/-}$ mice. The deletion of *Ebf1* was induced by the replacement of the DNA binding domain with the *neomycin resistance* cassette. These mice have a complete block in B cell maturation at the pro B stage, lacking of IgM^{+}B220^{+} mature B cells. This phenotype resembles *E2a* deficiency with normal T lymphocytes and myeloid differentiation (Lin and Grosschedl, 1995). The B cell impairment in *Ebf1*$^{-/-}$ B220^{+}IL7^{+} B cell progenitors can be rescued by the expression of EBF1. Cells reactivated the expression of B cell genes such as *λ5*, *VpreB*, *Rag1* and *Rag2* and underwent V-DJ rearrangement of the *IgH* locus (Medina et al., 2004). Moreover, Maier and colleagues reported the synergistic activation of *Mb-1* transcription by EBF1 and RUNX1 during early B cell differentiation. Using a luciferase assay, they showed that EBF1, RUNX1 and CBFβ activated the *Mb1* promoter more strongly than single factors (Maier et al., 2004). Methylation of the *Mb1* promoter decreases towards B cell commitment. In mice deficient for *E2a* or *Ebf1,* where B cell development is impaired, the *Mb1* promoter remains hypermethylated suggesting a role in chromatin remodelling of *Mb1* promoter for both genes (Hagman and Lukin, 2006).

The last main transcription factor involved in the early stages of the B cell differentiation is PAX5. This protein contains a DNA-binding domain at the N-terminal and a transactivation domain at the C-terminal. The DNA-binding domain recognizes the 5'-GCATC/GAT/AGCGTA/GC/A-3' sequence. *Pax5* deletion causes a block in B

cell developmental at an early pro B cell stage with a severe impairment in the V-DJ rearrangement. *Pax5*$^{-/-}$ pro B cells, cultured in the absence of IL-7, did not undergo further differentiation in contrast to control pro B cells maturing into IgM^{+} B cells. However, the re-expression of *Pax5* in *Pax5*$^{-/-}$ pro B cells rescued the B cell phenotype *in vitro* (Nutt et al., 1999). Using conditional *Pax5* inactivation in pro B cells, Mikkola and colleagues showed a requirement for *Pax5* to initiate B cell commitment and the ability to reconstitute T cell in *Rag2*$^{-/-}$ mice *in vivo* (Mikkola et al., 2002). The *Pax5*$^{-/-}$ pro B cells can be expanded *in vitro* and induced to differentiate into osteoclasts, T cells, NK-cells, macrophages, granulocytes and dendritic cells (Nutt et al., 1999). However, *Pax5*$^{-/-}$ pro B cells engrafted *in vivo* failed to reconstitute the host with the entire haematopoietic system (Nutt et al., 1999; Rolink et al., 1999) suggesting that the loss of *Pax5* induces the maintenance of progenitor cells with multi-lineage potential but not long-term *in vivo* repopulation capability.

Apart from these three key transcription factors, two other transcription regulators, both members of the high-mobility group (HMG) box family, are important for the progression to pro B cells. One is SOX4 that belongs to the SOX family of transcription factors. SOX4 protein is part of the SOX C group, together with SOX11, SOX22 and SOX24. This protein is composed by a HMG DNA-binding domain at the N-terminal and a trans-activation domain rich in serine at the C-terminal (Bowles et al., 2000). Mice deficient for *Sox4* died at E14 due to general oedema and valvular insufficiency. Upon engraftment of *Sox4*$^{-/-}$ foetal liver cells in recipient mice, reconstitution of the myeloid and T cell compartment was normal. In contrast, B cell development was impaired with a reduction in CD43^{+}B220^{+} pro B cells and the absence of more mature CD43^{-}B220^{+} pre B cells in the bone marrow. When culture in the presence of IL7, *Sox4*$^{-/-}$ foetal liver cells were less responsive and therefore proliferated

more slowly than control cells (Schilham et al., 1996).

The other member of the HMG box family involved in B cell development towards pro B cells is LEF1. Together with TCF1, TCF3 and TCF4, LEF1 is part of the LEF1/TCF family of transcription factors and contains a HMG DNA-binding domain at the C-terminal and a β-catenin binding domain at the N-terminal. The context-dependent activation domain occupies the central part of the protein and promotes the formation of multi-protein enhancer complexes. Using a *LacZ* model, Reya and colleagues analysed the expression of *Lef1 in vivo*. They observed the expression of LEF1 during B cell development. LEF1 is expressed in foetal liver and bone marrow mainly in $B220^{+}BP1^{-}$ B cells. Moreover, in $Lef1^{-/-}$ mice B cell maturation was blocked at the pro B stage in the foetal liver. The reduced growth and survival of B cells was related to defects in normal WNT signalling (Reya et al., 2000).

Two key transcription factors belonging to the interferon regulatory factor (IRF) family are critical for the differentiation into pre B cells. These proteins are composed of a DNA-binding domain at the N-terminal that recognizes the 5'-AAAGGAAGTGAAACCA-3' motif and a trans-activation and regulatory domain at the C-terminal. At 10 to 15 weeks of age, an expansion of T and B lymphocytes was observed in $Irf4^{-/-}$ mice. A closer analysis of B cell development showed no differences in early B cell maturation. Defects were observed in the peripheral B cell maturation with a strong reduction in $CD23^{+}B220^{+}$ B cells and the absence of $CD23^{high}B220^{+}$ B cells. In addition, these mice showed a decrease in immunoglobulin production that resulted in an impaired immune response (Mittrucker et al., 1997). In contrast, the deletion of IRF8, induced by the insertion of the *neomycin resistance* cassette into the second exon, impaired the production of IFNγ and induced defects in macrophage production that evolved in chronic myelogenous leukaemia (Holtschke et al., 1996).

Interestingly, the combination of the two single knock-outs arrested B cell development at the pre B stage. *Irf4*$^{-/-}$*Irf8*$^{-/-}$ lacked immature IgM^{+} B cells in the bone marrow and IgM^{+} peripheral B lymphocytes. In these mice, the pre B cell population showed a high proliferative rate due to an increase in the expression levels of the pre-BCR complex. This increase in the expression of the pre-BCR resulted from a specific failure in the down-regulation of *λ5* and *VpreB* genes (Lu et al., 2003).

Finally, the last step of B cell development is regulated by OBF1 and AIOLOS. The octamer-binding transcription factor (OCT)-binding factor 1 is composed of a POU-domain interaction site at the N-terminal and a trans-activation domain at the C-terminal. *Obf1* null mice showed normal early B cell development. In contrast, they showed impairment in peripheral B cell with an important reduction in mature IgM^{+}IgD^{+} cells. In addition, antibody production was affected by the loss of *Obf1* and this phenotype correlated with a lack of germinal centers (Nielsen et al., 1996). Similarly to IKAROS, the AIOLOS protein contains six zinc finger domains, four of which located at the N-terminal and two at the C-terminal. The N-terminal DNA-binding domain recognizes the 5'-GGGAA/T-3' motif while the C-terminal domain induces the formation of homodimers or heterodimers. *Aiolos*$^{-/-}$ mice showed an increase in pre B cells and a diminution in marginal zone B cells, B1 B cells and recirculating IgM^{+}IgD^{+} cells. These mice did not have problem in antibody production but instead showed spontaneous germinal centers formation. Moreover, *Aiolos* deficient mice developed systemic lupus erythematous, an autoimmune disease, characterized by the production of auto-antibodies (Wang et al., 1998). Sun and colleagues performed analysis on mice double knock-out for both *Obf1* and *Aiolos*. Surprisingly, the production of auto-antibodies was completely abolished suggesting an essential role for OBF1 in the autoimmune response. The early stages of B cell development were not

affected by the loss of these two transcription factors. Instead, the immature $B220^{+}IgM^{+}$ population was severely decreased. These data supported the requirement of those two factors in the transition from pre B cells to immature cells (Sun et al., 2003).

1.7. Immunoglobulin rearrangement

The somatic gene rearrangement of the immunoglobulin and T cell receptor is a specific process occurring only in B and T cells, respectively.

Four different regions compose the immunoglobulin heavy chain (*IgH)* locus, which are: variable (V), diversity (D), joining (J) and constant (C). Each of them is, then, subdivided in several exons. The first step of the *IgH* locus rearrangement results in the rearrangement of one segment of D region to one of the J region. The second part of the rearrangement brings one segment of V region next to the DJ region forming the VDJ region (Figure 1.12). Consequently, the enhancer, located between the J and C regions is now closer and can activate the promoter associated with the V region. The pre-mRNA is transcribed and can be processed in two different ways. One generates VDJ with Cμ while the other produces VDJ together with Cδ (Gonzalez et al., 2007).

The rearrangement of the light chain can occur on the lambda (λ) or kappa (κ) genes. The former one is composed by several segments of the V region and four segments of the C region, each of them separated by J region. The latter contains several segments of the V region and J region and only one C region (Figure 1.13). The λ light chain recombination only occurs when the rearrangement of the κ chain is not successful (Gonzalez et al., 2007).

Finally, once both heavy and light chains are translated, they are assembled together in the endoplasmic reticulum and then the immunoglobulin is expressed at the cell surface membrane.

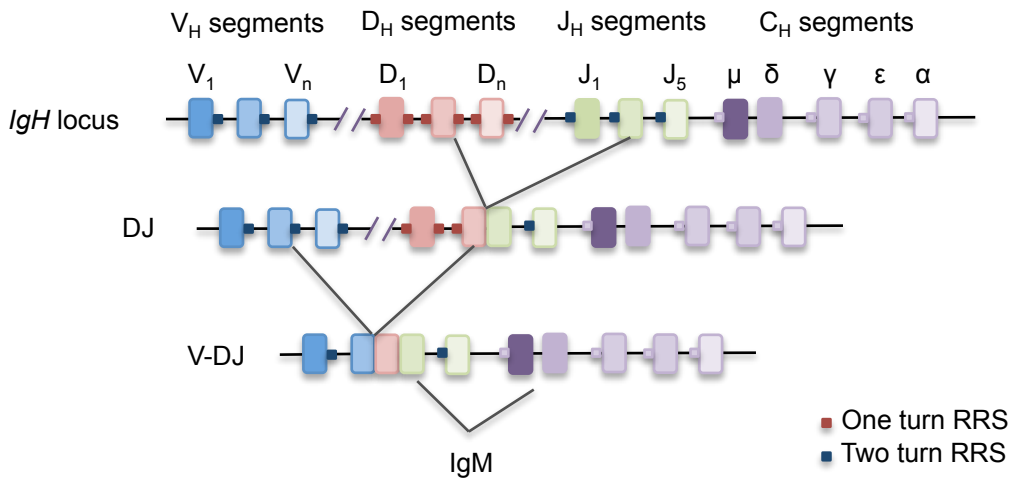


Figure 1.12. Schematic diagram of IgH rearrangement.
The *IgH* locus is composed by four different regions: V (V_n=1000), D (D_n=15), J and C. The first step of the *IgH* locus rearrangement occurs between one segment of D region to one of the J region. The second step, instead, completes the rearrangement bringing one segment of V region next to the DJ region. The DNA rearrangement is catalyzed by RAG1 and RAG2 that recognize the recombination signal sequences (RSS). In the heavy chain the one turn RSS is located close to the D region while the two turn RSS is close to the V and J regions.

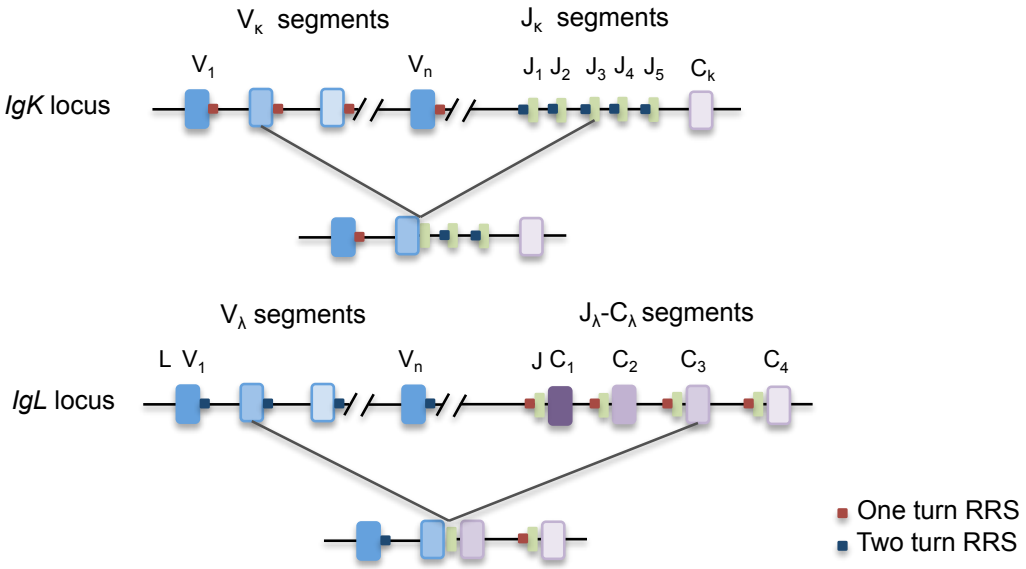


Figure 1.13. Schematic diagram of light chain rearrangement
The *IgK* locus is composed by the V (V_n=300), J and one C regions. The rearrangement occurs between one segment of the V region and one segment of the J region. The rearrangement of the *IgL* locus occurs only when the rearrangement of the κ chain is not successful. The *IgL* locus is composed by several segments of the V region and four segments of the C region, each of them separated by J region. The rearrangement brings one segment of the V region next to one segment of the J region. RAG1 and RAG2 recognizing the recombination signal sequences (RSS) catalyzed the DNA recombination. In the κ light chain the one turn RSS is close to the V region while the two turn RSS is near the J region. These positions are opposite in the λ light chain, one turn RSS on J region and two RSS on V region.

The DNA rearrangement is catalysed by two proteins: RAG1 and RAG2. These proteins recognize the recombination signal sequences (RSS). The RSS are a conserved nonamer and heptamer separated by 12 or 23 base pair depending on one or two turns of the DNA helix. The DNA recombination occurs always between one and two turns RSS. In particular, in the heavy chain the one turn RSS is located close to the D region while the two turns RSS is close to the V and J regions. On the contrary, in the κ light chain the one turn RSS is close to the V region while the two turns RSS is near the J region. These positions are opposite in the λ light chain, one turn RSS on J region and two RSS on V region (Male et al., 2006).

It has been shown that RAG1 and RAG2 are sufficient to induce Ig recombination. Their expression is restricted to lymphoid organs but RAG1 is also expressed in the central nervous system in mice (Chun et al., 1991). Deletion of *Rag1* abrogates the maturation of B and T lymphocytes. T cells were negative for CD4 and CD8 in the thymus while B cell were IgM^-IgD^- both in the spleen and in the bone marrow. Moreover, no detectable IgM was reported in the serum (Mombaerts et al., 1992). Similarly to *Rag1*$^{-/-}$ mice, the deletion of *Rag2* induced a block in B cell development with no IgM^+ cells in bone marrow, spleen and serum. All the $B220^+$ cells in the bone marrow were also $CD43^+$ suggesting a blockage in the immature stage of B cell maturation before Ig rearrangement. As in *Rag1* mutant, T cells were not detected in *Rag2* null mice and the TCR rearrangement was undetectable (Shinkai et al., 1992).

Once the immunoglobulin binds to an antigen, B cells can undergo a process called switch recombination. During this process, cells change their heavy chain constant region resulting in a production of different classes of immunoglobulin. The switch recombination does not alter the specificity of the antigen binding site but instead alters the antibody's effector function. Indeed the constant region is bound by

different protein receptors such as Fc receptors that change the antibody function. There are five different classes of immunoglobulin: IgM, IgD, IgE, IgG and IgA. The IgG class is subdivided in four sub-classes both in mice and humans while two classes only in human form the IgA class. The switch recombination is activated by T-cell dependent responses together with signals from cytokines such as IL4, TGFβ and IFNγ (Stavnezer, 1996). The recombination occurs upstream of the C region between specific sequences containing a tandem repeat sequence rich in G, called switch (S) regions that vary in length from 1 to 12 Kb (Dunnick et al., 1993). Activation-induced cytidine deaminase (AID) deaminases cytosines in uracils in S regions. The repair of the uracils by base excision repair (BER) induces single-strand DNA breaks (SSB). Then, the mismatch repair (MMR) converts SSB into double-strand DNA breaks (DSB) initiating the end-joining intra-chromosomal DNA recombination (Stavnezer et al., 2008).

The other process occurring in mature B cells is the somatic hypermutation (SHM) of the variable region. Similarly to switch recombination, the SHM is initiated by AID that replaces amino acid in the V regions of H and L chains. Most of the mutations are single base changes; more frequently transitions C to T or G to A than transversions C to A or G and G to C or T (Peled et al., 2008). AID, deaminating a cytidine, creates a uridine:guanosine mismatch. This brings three different pathways. In the first one, the replication machinery read the U as T creating a transition C to T mutation. In the second pathway uracil DNA glycosylase enzymes (UNG) removes the U leaving a site without a base that is, then, filled with any nucleotides by BER. The third way of generating mutation is through MMR that recognizes and excises the strand containing the U base. Then, proliferating cell nuclear antigen (PCNA) recruits error-prone polymerases to fill the gap. The SHM process increases the affinity and the specificity of binding between immunoglobulin and antigen and is mostly detected in

activated and memory B cells (Neuberger and Milstein, 1995; Peled et al., 2008).

1.8. IL7 signalling

The IL7 receptor (IL7R) is composed by two chains: α and common γ (γc). The first one is mainly expressed in lymphoid cells and has a high specificity for binding to IL7 while the second one is expressed by different haematopoietic cells and is important for initiating the cascade downstream of IL7R (Jiang et al., 2005).

Knock-out for *Il7* induced a decrease in spleen cellularity and impaired B cell development. Interestingly, this phenotype was more severe in mice deficient for *Il7r* (Carvalho et al., 2001). Inserting the *neomycin resistance* cassette in the third exon of the *Il7r* gene generated *Il7r*$^{-/-}$ mice. The cellularity of thymus, spleen and lymph node of these mice was reduced. The reduction in thymus cellularity reflected a T cell impairment characterized by a blockage at the CD4^{-}CD8^{-}CD25^{+} stage. Moreover, B cell development was blocked as well. In particular, *Il7r*$^{-/-}$ mice were able to generate B cells only until the pre pro B cell stage (Peschon et al., 1994).

Differences between mice and human in IL7-dependent B cell development have been reported. Interestingly, while IL7R is essential for B cell generation in mice, patients with severe combined immunodeficiency (SCID) caused by *IL7Rα* deficiency have been shown to maintain normal CD19^{+} B lymphocytes but to have diminished natural killer cells and T-lymphocytes. In addition, SCID patients showed normal *IL7* mRNA levels (Puel et al., 1998). Giliani and colleagues have identified more cases of SCID due to *IL7R* mutations. Their analyses confirmed the role of IL7R in human T cell development. Moreover, mutations found in α chain might not compromise the function of the IL7 receptor (Giliani et al., 2005).

Both IL7R chains do not have tyrosine kinase activity; therefore they require

other kinases to initiate the downstream signalling. The α chain is associated with Janus Kinase 1 (JAK1) while the γ_c chain is associated with JAK3.

The JAK family is formed by four kinases: JAK1-3 and TYK2 that play important roles in signalling (Quintas-Cardama and Verstovsek, 2013). JAK3 is mainly expressed in haematopoietic cells and its deletion in mice induced a reduction in both B and T cells (Nosaka, 1995).

When IL7 binds to its receptor, the two receptor-chains dimerize bringing JAK1 close to JAK3. This close association between the two kinases induces a trans-phosphorylation. Activated JAK kinases phosphorylate the tyrosine residues on the α chain of the IL7R (Clark et al., 2014; Corfe and Paige, 2012); the phosphorylation on the IL7Rα recruits the signal transducer and activator of transcription 5A (STAT5A) and 5B (STAT5B). After being phosphorylated, STAT5 proteins dimerize and enter in the nucleus where they induce the transcription of target genes (Clark et al., 2014; Corfe and Paige, 2012) (Figure 1.13).

The STAT family comprises seven members (STAT1, 2, 3, 4, 5A, 5B and 6) that play critical roles in signal transduction pathways (Paukku and Silvennoinen, 2004). Different studies have been done to demonstrate the critical role for STAT5 proteins for lymphoid development. Mice carrying a deletion for *Stat5a/b* died prenatally. *Stat5*$^{-/-}$ foetuses were anaemic with a reduction in spleen and thymus cellularity. Analysis of B cell development suggested a decrease in $CD19^+B220^+$ cells in the spleen and foetal liver together with an accumulation of pre pro B cells (Yao et al., 2006).

The phosphorylation on the IL7Rα can also recruit phosphatidyl-inositol 3-kinase (PI3K), an enzyme that phosphorylates the third position of the phosphatidyl-inositol generating phosphatidyl-inositol-3,4,5,-trisphosphate (PIP_3). Activation of the PI3K induces phosphorylation of AKT. Finally, AKT regulates by phosphorylating

FOXO that enters in the nucleus and induces the transcription of target genes (Clark et al., 2014; Corfe and Paige, 2012) (Figure 1.14).

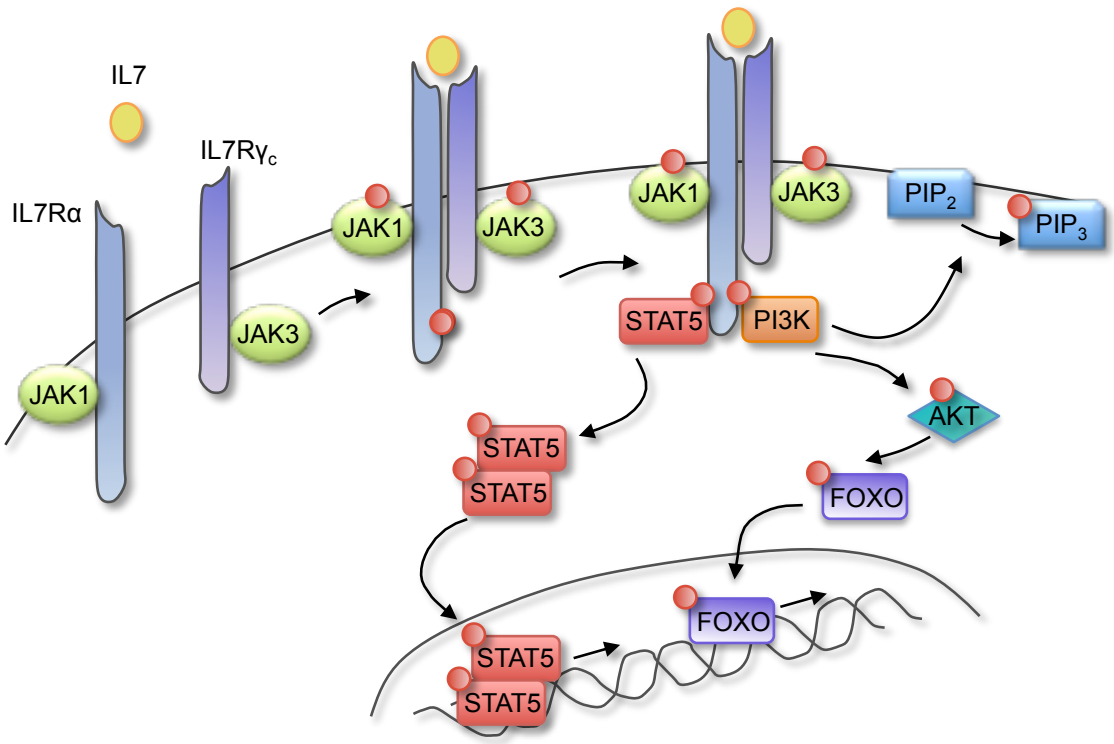


Figure 1.14. Schematic representation of IL7 signaling pathway.
IL7 binds to the IL7 receptor that is formed by two chain: α and γ_c. After IL7R dimerization, the two JAK kinases activate and phosphorylate the tyrosine residues on IL7Rα. This phosphorylation recruits signal transducer and activator of transcription 5 (STAT5). After being phosphorylated, STAT5 proteins dimerize, enter in the nucleus and induce the transcription of target genes. IL7Rα can also recruit phosphatidyl-inositol 3-kinase (PI3K). PI3K generates phosphatidyl-inositol-3,4,5,-trisphosphate (PIP_3) and phosphorylates AKT. Finally, AKT phosphorylates FOXO that enters in the nucleus and induces the transcription of target genes.

The PI3K kinase is composed of one regulatory subunit, p50α, p55α or p85α, and one catalytic subunit, p110. Mice deficient for *p85* and its isoforms died after birth. To bypass the problem of this early lethality in *p85α*$^{-/-}$ mice, Fruman and colleagues investigated the role of PI3K in lymphocyte development injecting *p85α*$^{-/-}$ ES cells in *Rag2*$^{-/-}$ blastocysts. *p85α*$^{-/-}$*Rag2*$^{-/-}$ mice had a decrease in B220^{+} cells in the spleen and in serum immunoglobulin but an increase in pro B cells in the bone marrow (Fruman et al., 1999). In contrast, mice lacking the catalytic subunit, *p110δ*, showed a reduction in B1 cells and CD21^{high}CD23^{low} marginal zone B cells in the spleen. Similarly to *p85α*$^{-/-}$ *Rag2*$^{-/-}$, defect in the production of immunoglobulin was reported in mice deficient for *p110δ* (Clayton et al., 2002). In addition, the deletion of both *p110δ* and *p110α* induced a block at the pre B cell stage and impairment in IgM^{+} immature cells suggesting an essential role for *p110α* in B cell development (Ramadani et al., 2010).

1.9. BCR signalling

During B cell maturation, immunoglobulin rearrangement is controlled at two main checkpoints. The first one occurs in pre B cells to check the functional rearrangement of the heavy chain (*IgH*) locus while the second one controls the functional rearrangement of the light chain loci (*Igκ* or *Igλ*) in immature B cells.

In the pre B cell stage, only cells that have successfully completed the *IgH* rearrangement and expressed a pre-BCR at the cell surface, are positively selected and can progress to the proliferative stage. This phase is called proliferative phase resulting in five cell divisions and is driven by both IL7 and pre B cell receptor (BCR) signalling. Successively, cells undergo a maturation phase that involves the rearrangement of the light chain κ locus (Miosge, 2005). The Pre-BCR complex is composed of the heavy chain, Igμ, the two surrogate light chain (SLC) molecules, VpreB and λ5, and Igα-Igβ

dimer. *Igμ* mutant mice lacked $B220^+$ mature cells in blood and spleen and IgM in the serum. Moreover, B cells, unable to assemble the pre-BCR receptor, were stopped at the pre B cell stage in the bone marrow suggesting an important role for the heavy chain μ during B cell maturation (Kitamura et al., 1991). In Mundt's study, *VpreB1* and *VpreB2* sequences were disrupted by the insertion of *neomycin* and *hygromycin resistance* cassettes, respectively. *VpreB1*$^{-/-}$*VpreB2*$^{-/-}$ mice showed defects in B cell maturation. Specifically, cells were blocked at the pre B cell stage. Indeed, a marked decrease in the number of IgM^+ immature and IgM^+IgD^+ mature B cells was observed in the bone marrow of these double mutant mice (Mundt et al., 2001).

Similarly to *VpreB1* and *VpreB2*, Kitamura and colleagues have reported the importance of λ5 in B cell development by disrupting the expression of this gene with a *neomycin resistance* cassette inserted in the λ locus. The deletion of *λ5* induced a blockage at the pre B stage of the B cell maturation in the bone marrow while the peripheral B cells was partially affected. B cells in blood and spleen were activated but were dramatically reduced in number (Kitamura et al., 1992).

The pre-BCR complex interacts with the Igα-Igβ heterodimer to perform and initiate signalling. Igα (CD79a, mb1) and Igβ (CD79b, B29) are two signal transducers carrying an immuno-receptor tyrosine activation motif (ITAM) in their cytoplasmic tail (Miosge and Goodnow, 2005).

Transgenic experiments were used to demonstrate the essential function of Igα and Igβ in the BCR complex. The truncated Igβ and Igα proteins were generated introducing a stop codon in their locus. The two transgenic mice showed a B cell impairment leading to an increase in pro B cells and a decrease in pre B cells. These defects were more pronounced in *Igα*$^{-/-}$ than in *Igβ*$^{-/-}$. When crossed together, in double mutant mice the blockage was at pre B stage as reported in *Igβ*$^{-/-}$ (Reichlin et al., 2001).

To investigate the function of the ITAM phosphorylation of Igα in B cell, tyrosine residues were replaced by phenylalanine residues within the ITAM motif ($Ig\alpha^{FF/FF}$). This mutation did not affect B cell differentiation in the bone marrow. However, peripheral B cells, marginal zone B cells and B1 cells were reduced in number and the T cell dependent response was less efficient. This phenotype changed dramatically when $Ig\alpha^{FF/FF}$ was combined with truncated *Igβ*. Cells were blocked at the pro B cell stage demonstrating the importance of the phosphorylation of ITAM for B cell maturation (Kraus et al., 2001).

Cells, that positively pass the pre-BCR checkpoint, proceed to the maturation phase where they initiate the rearrangement of the immunoglobulin light chain (*IgL)* locus. During the rearrangement of the light chain, allelic exclusion occurs resulting in the generation of a single light chain. This process increases the diversity and the specificity of antibody. The light chain rearrangement stops when one light chain and one heavy chain associated together. Consequently, cells differentiate and are selected for their ability not to react to the self in the bone marrow prior to enter into the blood circulation (Nossal, 1994).

Antigen binding to the BCR complex induces the phosphorylation of tyrosine residues within the ITAM motif in the cytoplasmic tail of Igα-Igβ heterodimer by SRC kinases such as LYN, FYN and BLK. This phosphorylation recruits and phosphorylates the spleen tyrosine kinase (SYK). $Syk^{-/-}$ fetuses showed haemorrhages and died after birth. When $Syk^{-/-}$ foetal liver cells were transplanted in *Rag1* deficient mice, B cell impairment was observed. No $IgM^{+}IgD^{+}$ mature B cells were detected in the bone marrow and reduced number of pre B cells was also observed suggesting impairment in the pre-BCR signal transduction that partially blocked the progression and maturation of pro B cell (Turner et al., 1995). This partial blockage in B cell maturation was

probably due to the activation of a SYK-independent pre-BCR signalling through ZAP-70, the only other kinase in the SYK family. To address this possibility, Schweighoffer and colleagues generated mice deficient for both *Syk* and *Zap-70*. The absence of both kinases induced an arrest at the pro B stage with no detectable pre B cells and a failure of heavy chain allelic exclusion affecting therefore a correct BCR signalling (Schweighoffer et al., 2003).

Activated SYK initiates the downstream cascade for BCR signalling. It phosphorylates B cell linker protein (BLNK) that successively recruits the SRC-related tyrosine kinase BTK. SYK, together with BTK, phosphorylates and activates PLCγ2. PLCγ2 generates diacylglycerol (DAG) and inositol-1,4,5-triohosphate (IP_3) from phosphatidyl-inositol-4,5-bisphosphate (PIP_2). DAG activates the protein kinase C (PKC) while IP_3 induces the release of Ca^{2+} from the endoplasmic reticulum. The increase in Ca^{2+} induces the translocation of NFAT into the nucleus where it regulates the expression of target genes. In addition, the mitogen-activated protein kinase (MAPK) signalling is also activated by SOS, BNLK, VAV1 (Figure 1.15) (Miosge and Goodnow, 2005).

To confirm the essential role for BLNK in B cell development, *Blnk* mutant mice were generated. An accumulation of pro B cells was reported suggesting a blockage in the maturation to pre B cells. Moreover, less IgM^+ mature peripheral B cells were detected in the spleen (Pappu et al., 1999).

Another element crucial for BCR signalling is the SCR-related tyrosine kinase BTK. Mutation in this kinase resulted in X-linked agammaglobulinemia (XLA) in human and X-linked immunodeficiency (xid) in mouse. This immunodeficiency disease is characterized by the absence of circulating B cells. Inactivating the *Btk* gene with the insertion of the *LacZ* gene, Hendriks and colleagues reported defect in the maturation

step to $IgM^{low}IgD^{high}$ population in the spleen and in particular during the transition from pre B cell to immature B cells in the bone marrow (Hendriks et al., 1996). In *Btk* deficient mice, the phosphorylation of PLCγ2 was markedly reduced resulting in an arrest of phosphatidyl-inositol hydrolysis and calcium release suggesting the importance of the BTK-PLCγ2 crosstalk during the BCR signalling (Takata and Kurosaki, 1996).

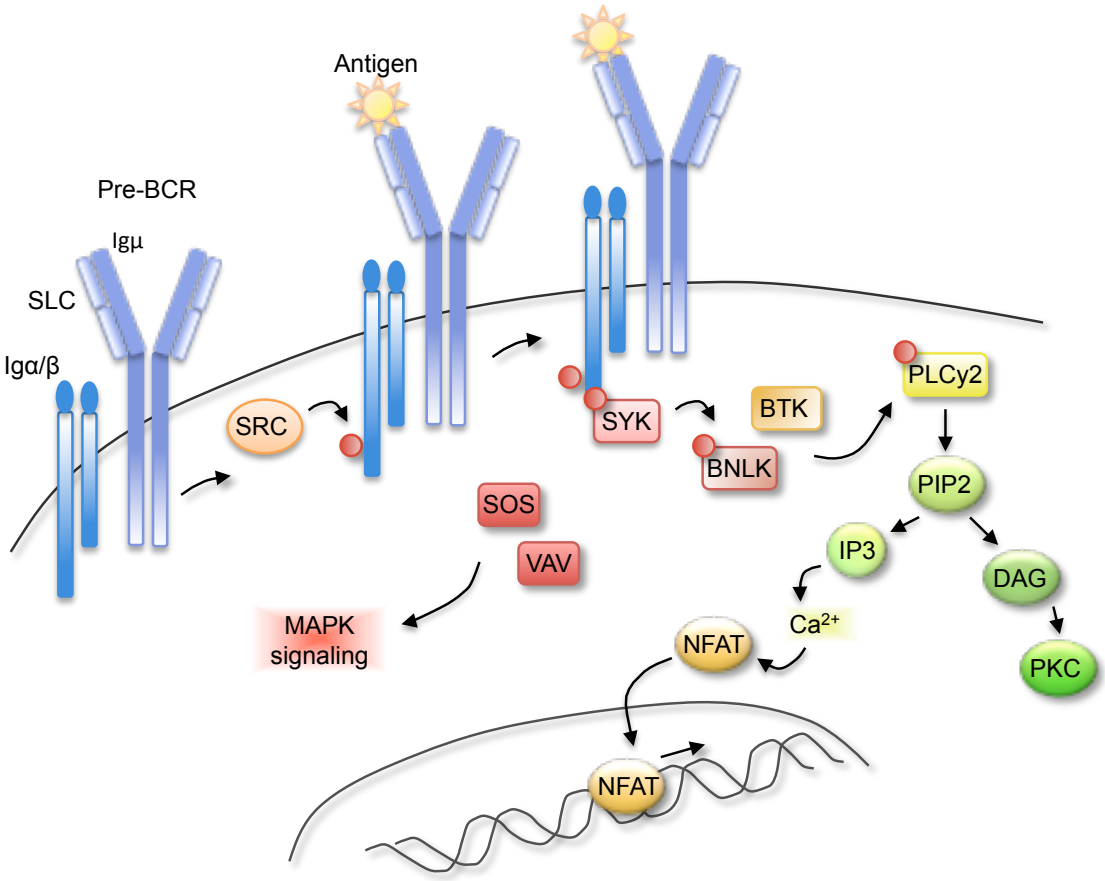


Figure 1.15. Schematic representation of BCR signaling pathway.
Pre-BCR is composed by the heavy chain (Igμ), two surrogate light chain (SLC) molecules (VpreB and λ5) and Igα-Igβ dimer. The antigen binds to the BCR complex and induces SRC kinases to phosphorylate the ITAM motif in the Igα-Igβ cytoplasmic tail. This phosphorylation activates spleen tyrosine kinase (SYK) that initiates the BCR signalling. SYK phosphorylates B cell linker protein (BLNK) and together with BTK, phosphorylates and activates PLCγ2. PLCγ2 generates diacylglycerol (DAG) and inositol-1,4,5-triohosphate (IP_3). DAG activates the protein kinase C (PKC) while IP_3 induces the release of Ca^{2+} from the endoplasmic reticulum inducing the translocation of NFAT into the nucleus where regulates target gene expression. Another signalling pathway activated during the BCR cascade is the mitogen-activated protein kinase (MAPK) signaling.

1.10. The SOX family of transcription factors

The Sex-determining region Y-box (SOX) family of transcription factors was first defined in mammals based on sequence conservation within the High Mobility Group (HMG) box initially identified in the testis-determining factor SRY (sex determining region Y) (Gubbay et al., 1990; Sinclair et al., 1990). The HMG domain is a DNA-binding motif of about 79 amino acids and consists of three alpha helices forming an L-shape with an angle of about 80° between the two arms (Kondoh and Kamachi, 2010). The HMG domain superfamily is composed of two subfamilies: TCF/SOX/MATA and HMG/UBF. The former group typically binds to a single specific sequence while members of the latter group have a HMG domain that is less sequence-specific in binding (Bowles et al., 2000; Soullier et al., 1999). The SOX family usually binds the minor groove of the DNA, specifically to the sequence motif 5'-(A/T)(A/T)CAA(A/T)G-3', causing unwinding and bending of the DNA helix (Bowles et al., 2000; Laudet et al., 1993; Stros et al., 2007). This binding characteristic distinguishes the SOX family proteins from other transcription factors, which usually bind to the major groove of the DNA. The DNA-binding affinity of HMG domain can be modulated by phosphorylation. Four phosphorylation sites are localized within the conserved HMG domain and control the SOX binding activity (Siino et al., 1995). Apart from being responsible for DNA binding, the HMG domain is also necessary for intracellular transport regulation and interactions with partner proteins. Two nuclear localization signals (NLS) have been identified within the HMG domain, located at the N and C termini of the conserved domain, respectively. These NLS motifs are required for a complete nuclear translocation (Sudbeck and Scherer, 1997). The N-terminal NLS binds to calmodulin, a calcium binding protein involved in the calcium signalling, or to exportin-4, a nuclear export receptor. On the contrary, the C-terminal interacts with

importin-β that mediates nuclear translocation through the nuclear membrane. Acetylation of the lysine residues in the C-terminal NLS motif enhances SOX nuclear import while deacetylation of these residues relocates the SOX proteins to the cytoplasm. Another conserved motif within the HMG domain is the short sequence named nuclear export signal (NES). The export receptor CMR1 binds to the NES sequence and facilitates the cytoplasmic translocation of SOX proteins (Malki et al., 2010).

The SOX family comprises 20 members in mice and humans (Schepers et al., 2002). Based on sequence and structural motif, the SOX family is divided into eight groups, from A to H with the group B subdivided into B1 and B2 (Table 1.1) (Bowles et al., 2000). Within each group, SOX proteins have at least 70% of amino acid sequence identity whereas SOX proteins from different groups have low amino acid sequence similarities. All SOX proteins, however, share more than 50% amino acid similarity within the HMG box. This family of transcription factors has been implicated in cell fate decisions in numerous developmental processes and can act as activators or repressors depending on the cellular context and their associated interacting proteins (Wegner, 1999). A number of SOX proteins contain, within the C-terminal region, trans-activation domains whose activity is often promoter and partner-dependent (Kamachi et al., 2000).

Sox Group	Genes	Functions	Species and Chromosomal location
A	*Sry*	Testis determination	Mouse Y(3cM); Human Yp11.3
B1	*Sox1*	Lens development	Mouse 8(4cM); Human 13q34
	Sox2	Neural induction	Mouse 3(15cM); Human 3q26.3
	Sox3	Neural determination	Mouse X(24.3cM); Human Xq27
B2	*Sox14*	Interneuron specification, limb development	Mouse 9(53cM); Human 3q22
	Sox21	CNS patterning	Mouse 14(50cM); Human 13q32
C	*Sox4*	Heart, lymphocyte, tymocyte development	Mouse 13(20cM); Human 6q22.3
	Sox11	Neuronal, glia maturation	Mouse 12(11cM); Human 2p25
	Sox12	Development of tissues	Mouse 2(86cM); Human 20p13
D	*Sox5*	Chondrogenesis	Mouse 6(69.5cM); Human 12p11.1
	Sox6	Chondrogenesis	Mouse 7(55cM); Human 11p15.3
	Sox13	Arterial wall, pancreatic islets development	Mouse 1(70cM); Human 1q31
E	*Sox8*	Development of tissues	Mouse 17(8cM); Human 16p13.3
	Sox9	Chondrogenesis, sex determination	Mouse 11(69.5cM); Human 17q25
	Sox10	Neural crest specification	Mouse 15(46.5cM); Human 22q13
F	*Sox7*	Vascular and tissues development	Mouse 14(28cM); Human 8p22
	Sox17	Endoderm specification	Mouse 1(7cM); Human 8q11.2
	Sox18	Vascular, hair follicle development	Mouse 2(96cM); Human 20p13.3
G	*Sox15*	Myogenesis	Mouse 11(39cM); Human 17p13
H	*Sox30*	Male germ cell maturation	Mouse 11(20cM); Human 5q35

Table 1.1. *Sox* family members

1.11. SOX F subfamily group

The SOX family of transcription factors are subdivided into eight groups, as mentioned above. The F group subfamily comprises 3 members: SOX7 together with its close homologues SOX17 and SOX18 (Bowles et al., 2000).

1.11.1. *SoxF* genes: gene and protein structure

The murine *Sox7* gene is located on chromosome 14 band D while its human counterpart is found on chromosome 8, at position p22 and contains 2 exons and one intron. Both murine and human SOX7 proteins are around 380 amino acids in length and share 87.4 per cent identity (Takash et al., 2001; Taniguchi et al., 1999). The high sequence similarities observed between mouse and human SOX7 proteins suggests conservation in their functions (Schepers et al., 2002).

The murine *Sox17* and *Sox18* are located on chromosome 1A1 and 2H4 respectively,

while their human counterparts are located on chromosome 8, at position q11.23 and on chromosome 20q13.33, respectively. The conserved intronic positions within the *Sox7*, *Sox17* and *Sox18* loci suggest that those genes have a common ancestor (Bowles et al., 2000; Takash et al., 2001). Like other SOX proteins, SOX F factors selectively bind, through their HMG domain, the heptameric consensus DNA sequence, 5'-(A/T)(A/T)CAA(A/T)G-3'. Subsequently, SOX F proteins are able to activate transcription via a C-terminal trans-activation domain adjacent to the HMG domain (Hosking et al., 1995; Taniguchi et al., 1999). Finally, a short amino acid motif (DXXEFD/EQYL) mediating β-catenin interaction has been identified in all members of the SOX F group, with the exception of SOX17 in the zebrafish (Figure 1.16) (Sinner et al., 2004; Takash et al., 2001).

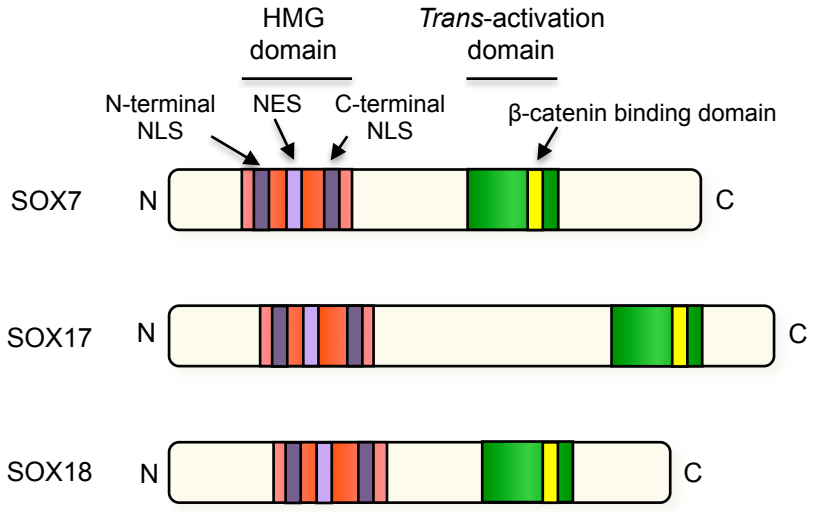


Figure 1.16. Schematic representation of the mouse SOX F proteins.
The SOX F proteins are composed by two domains: the HMG domain and the trans-activation domain. Two nuclear localization signals (NLS) within the HMG domain are located at the N and C termini of the conserved domain, respectively. The N-terminal NLS domain binds to calmodulin while the C terminal NLS interact with importin-β. The trans-activation domain activates the transcription. A short amino acid motif (DXXEFD/EQYL) mediating β-catenin interaction is located within the trans-activation domain.

1.11.2. *SoxF* genes: mode of transcriptional activation

SOX proteins bind to specific DNA-binding sites through their HMG domains. However, the SOX proteins are not sufficient on their own to promote a regulatory function; additional binding sites for other transcription factors are usually found nearby the SOX-binding sites. Transcriptional activation or repression is then mediated through cooperation between SOX proteins and binding partner (Figure 1.17) (Kamachi et al., 2000; Kondoh and Kamachi, 2010).

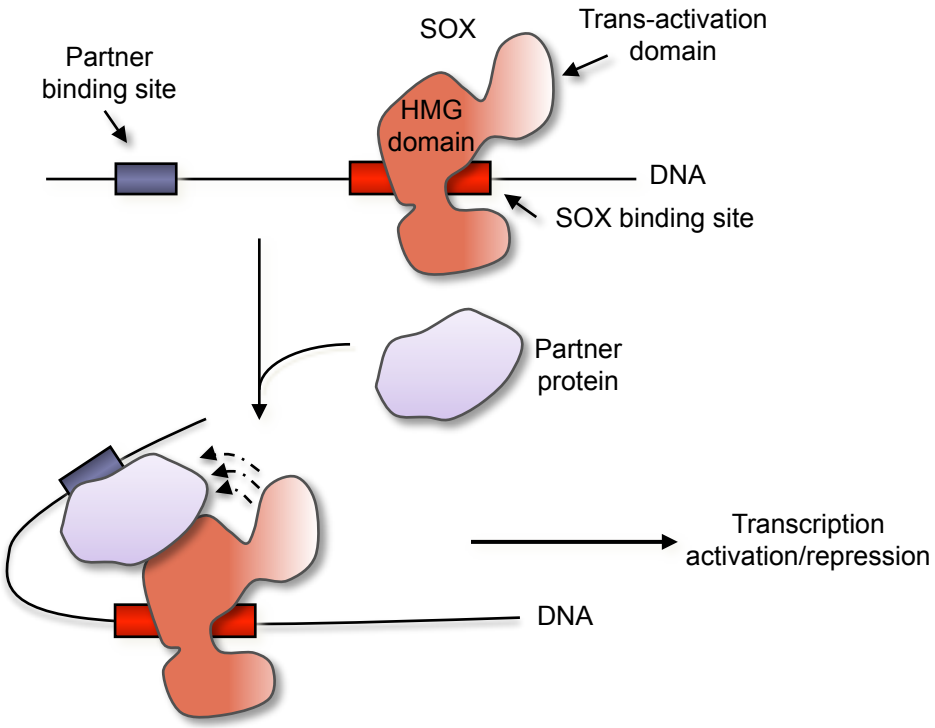


Figure 1.17. Schematic representation of SOX-partner factor interaction.
SOX protein binds to specific DNA-binding sites through their HMG domains. Then, when the SOX-partner factor interacts with SOX and its DNA-binding site, this protein complex is able to mediate its transcriptional activity.

The ability of SOX F factors to orchestrate gene transcription depends on different molecular properties including DNA-binding and bending, specific target selection, and interaction with co-factors, partner proteins (MEF2C, β-catenin, SMAD3) and growth factor signalling pathways (VEGFs). In particular, SOX18 interacts via its HMG domain with myocyte enhancer factor 2C (MEF2C) in endothelial cells (Hosking et al., 2001) and SOX17 interacts with SMAD3 repressing its transcriptional activity in lung epithelial cells (Lange et al., 2009).

In addition, some studies reported that SOX7 is implicated as a potential modulator of many essential signalling pathways, such as fibroblast growth factor (Murakami et al., 2004), WNT/β-catenin (Kormish et al., 2010; Takash et al., 2001) and Nodal (Zhang et al., 2005a). Indeed, it has been reported that SOX7 is involved in the activation of *Fgf3* transcription, competing with GATA4, in mouse embryonal carcinoma cell line while SOX17 showed a much weaker activity (Murakami et al., 2004). Over-expression of SOX7 and SOX17 induced β-catenin degradation in colon cancer cell lines. Guo and colleagues, using *Sox7* mutant construct, showed that β-catenin binding motif is necessary for *Sox7* binding to β-catenin (Guo et al., 2008). Overall, the biochemical nature of SOX F interaction and function still remains unclear.

1.11.3. *SoxF* genes: roles and functions

The role of the *SoxF* genes is best characterized in the development of the cardiovascular system. A detailed analysis of *SoxF* expression in the cardiac region of developing mouse embryo reveals that *Sox7* and *Sox18* are expressed in the pre-cardial region at E8.25 and in the heart tube at E8.5, while *Sox17* is undetectable (Sakamoto et al., 2007). A first observation on the role for *Sox18* in cardiovascular development was suggested by the cardiovascular defects seen in semi-dominant mutations in *Sox18*.

Moreover, the co-injection of *Sox7* and *Sox18* morpholinos strongly inhibited cardiogenesis in Xenopus suggesting an essential role of the two *SoxF* genes in this development process (Zhang et al., 2005b). $FLK1^+$ progenitors derived from embryonic stem cells expressed high level of *Sox7* that subsequently decreased when $FLK1^+$ cells progressed to cardiogenic development (Nelson et al., 2009). Recently, Wat and colleagues reported cardiac defects and congenital diaphragmatic hernia (CDH) in mice deficient for *Gata4* or *Sox7*. Deletion of the second exon of *Sox7* resulted in death before E14.5 in $Sox7^{\Delta ex2/\Delta ex2}$ embryos prior to diaphragm and cardiovascular development (Wat et al., 2012). In addition, duplication of SOX7 was found in child born with cardiac defects (Long et al., 2013).

All three *SoxF* members are also co-expressed in vascular endothelial cells (Sakamoto et al., 2007; Takash et al., 2001). Knocking-down of these genes causes limited effect on blood circulation. Both *Sox18* and *Sox7* are expressed at very early steps of endothelial differentiation. In fact, expression of *Sox18* begins at E7.5 in yolk sac blood islands and endothelial cells of the allantois. By E8.25-8.5 *Sox7* and *Sox18* but not *Sox17* are detected in the nascent pair of dorsal aorta (Pennisi et al., 2000; Sakamoto et al., 2007). At E9-9.5, *SoxF* expression becomes evident in smaller branching vessels and inter-somitic vasculature (Pennisi et al., 2000; Sakamoto et al., 2007; Young et al., 2006). At that stage, which corresponds to primitive lymphangiogenesis, *Sox18*, but not the other two *SoxF* genes, is expressed in lymphatic endothelial cells. *Sox18* deficiency causes lymphatic dysfunction suggesting that *Sox18* is critical for lymphatic vasculature in mice (Francois et al., 2008; Irrthum et al., 2003). Several studies have been shown that simultaneous knock-down of both *Sox7* and *Sox18* in zebrafish resulted in fusion between the major vessels with a failure in the formation of arterio-venous identity (Cermenati et al., 2008; Herpers et al., 2008; Pendeville et al.,

2008). Moreover, *Sox7* and *Sox18* regulated arterial specification via binding of Notch ligand Delta-like ligand 4 (DLL4) (Sacilotto et al., 2013). In addition, recent findings show that in certain strain (mixed 129/CD1 genetic background) of mice *Sox7* and *Sox17* are uniquely up-regulated during lymphatic development in the absence of *Sox18* and are able to functionally substitute for *Sox18* both *in vitro* and *in vivo* (Hosking et al., 2009). Other studies revealed that both *Sox7* and *Sox17* are highly expressed in the extra-embryonic endoderm (Hudson et al., 1997; Murakami et al., 2004; Shirai et al., 2005). During endoderm formation, GATA5 acts to induce Casanova (CAS) expression down-stream of Nodal signalling. Over-expression of CAS in zebrafish *one-eyed-pinhead* (*Oep*) mutant embryos, which lack of Nodal signalling, induced the expression of the endodermal markers SOX17 and FOXA2 (Alexander and Stainier, 1999; Aoki et al., 2002). Seguin's study on human ES cells revealed that stable endoderm progenitors can be established by constitutive expression of *Sox7* or *Sox17*, producing extra-embryonic endoderm and definitive endoderm progenitors, respectively (Seguin et al., 2008). Indeed, a role for *Sox17* in endoderm development was highlighted using *Sox17*$^{-/-}$ embryos. These embryos showed defects in gut definitive endoderm formation (Kanai-Azuma et al., 2002). Other studies reported that SOX7 and SOX17 regulated endoderm formation via binding of the parietal endoderm-specific enhancer of laminin-α1. Silencing of *Sox7* with short interfering RNA in embryonal carcinoma cells suppressed laminin-α1 production and GATA4 and GATA6 expression (Futaki et al., 2004; Niimi et al., 2004).

Finally, SOXF proteins were also reported to be involved in the blood development program. Kim and colleagues have shown that SOX17 is specifically expressed in foetal and neonatal HSCs. *Sox17* deletion leads to severe haematopoietic defects including loss of foetal and neonatal but not adult HSCs (Kim et al., 2007).

Over-expression of *Sox17* in adult haematopoietic bone marrow cells enhanced the multi-lineage reconstitution resembling foetal haematopoiesis with a high myeloid:lymphoid ratio. Moreover, ectopic *Sox17* expression induced an increase in immature progenitor cells, spleen size together with extramedullary haematopoiesis but it is not sufficient to cause leukaemia (He et al., 2011). Two recent studies reported a role for SOX17 in the formation of haemogenic endothelium. Over-expression of *Sox17* in endothelial cells derived from human embryonic stem cells (hESC) maintained the proliferation of $CD34^{+}CD43^{+}CD45^{-/low}$ haemogenic endothelial cells (Nakajima-Takagi et al., 2013). In addition, the use of *Sox17-GFP* mice also showed that *Sox17* is an important transcription factor expressed in haemogenic endothelium and HSC development (Clarke et al., 2013).

Sox7 and *Sox18* are also expressed at the onset of blood development; both being transiently expressed in $FLK1^{+}$ mesodermal precursor cells undergoing specification to the haematopoietic program. *Sox7* is expressed in gastrulating embryos (E7.5) in mesodermal masses in the yolk sac. Using inducible *Sox7* ES cells, the enforced expression of *Sox7* blocked the haematopoietic differentiation while enhanced the proliferation of haematopoietic precursors (Gandillet et al., 2009). Similarly to enforced expression of *Sox7*, over-expression of *Sox18* induced proliferation of early haematopoietic precursors. In contrast, the enforced expression of *Sox17* enhanced cell death (Serrano et al., 2010). Recently, it has been shown that SOX7 is expressed in haemogenic endothelium cells. SOX7 binds the promoter of the endothelial marker VE-cadherin and induced its expression (Costa et al., 2012b).

1.12. *Sox* genes and solid tumours

The expression of many *Sox* genes is up-regulated in several tumour types in human: their role in tumour development, growth and metastasis however remains to be determined. Few studies to date have clearly implicated SOX transcription factors in carcinogenesis. One of the most significant publications reported that the maintenance of glioma-initiating stem cells was dependent on an autocrine TGF-β pathway controlling *Sox2* expression. SOX2, a crucial factor for the maintenance of neural stem cells plays an essential role in the retention of glioma-initiating cells. Inhibition of TGF-β signalling decreased the tumorigenicity of glioma-initiating cells by promoting their differentiation, as the expression of *Sox2* was no longer maintained (Ikushima et al., 2009). The expression of *Sox2* has been also reported in lung and esophageal squamous cell carcinomas and its expression is important for cell proliferation (Bass et al., 2009). In addition, the expression of *SOX2* has been analysed in primary breast tumour samples and cutaneous squamous cell carcinoma (SCC). Its expression correlated with high KI67 proliferation index, high risk of recurrence in breast cancer and with tumour expansion in SCC suggesting a possible role for *SOX2* as prognostic marker in these two types of cancer (Finicelli et al., 2014; Siegle et al., 2014). Another study reported the correlation between *SOX2* expression and poor prognosis in cancer, in particular in colorectal cancer (CRC). Moreover, Lundberg and colleagues showed that *SOX2* expression is regulated by BRAF. Indeed, blocking BRAF downstream signalling in CRC cell line transduced with mutated *BRAF* induced a decrease in *SOX2* expression (Lundberg et al., 2014). Since *SOX2* expression increased in ovarian cancer samples, Wang and colleagues have also investigated its role in this type of cancer. They found, using an *in vitro* migration assay, that the over-expression of *SOX2* enhanced migration and invasion of ovarian cancer cell lines suggesting a role for *SOX2* in metastasis

formation (Wang et al., 2013).

In addition, *Sox1* has been identified as an antigen expressed in para-neoplastic neurological syndrome associated with small cell lung cancer (Tschernatsch et al., 2009). Indeed, the use of SOX1-antibodies has been reported to be highly specific for this disorder. Moreover, the presence of SOXB1 group (SOX1, 2, 3) antibodies has been detected in small cell lung carcinoma and the correlation between clinical parameters of less aggressive disease and seroreactivity suggests that these antibodies are indicators of better prognosis in this kind of tumour (Vural et al., 2005). Several studies describe the presence of *SoxC* (*Sox4*, *11*, *12*) gene expression in different tumours. *Sox4* and *Sox11* are highly expressed in most medulloblastomas. Increased *Sox4* expression is also associated with prostate, colon, non-small cell lung tumours and high expression of *Sox11* is associated with gliomas, non-B cell lymphomas and epithelial ovarian tumours (Medina et al., 2009; Penzo-Mendez, 2010). In addition, increased level of *Sox4* has been reported in endometrial cancers. Interestingly, Saegusa and colleagues showed that *Sox7* induced the up-regulation of *Sox4* expression, showing a complex network between SOX members (Saegusa et al., 2012).

Sox genes of the group E and D are expressed in various types of gliomas. In particular, *Sox5*, *Sox9* and *Sox10* are generally expressed at levels similar to those in adult brain tissues while *Sox6*, *Sox8* and *Sox13* are up-regulated in many oligodendrogliomas (Schlierf et al., 2007).

Regarding *SoxF* genes, *Sox*17 has been shown to negatively regulate canonical WNT/β-catenin signalling, to inhibit human hepatocellular carcinoma cells growth (Jia et al., 2010), to be down-regulated in colon cancer and induced by WNT activation in the early stage of gastrointestinal tumourigenesis but down-regulated by methylation during malignant progression (Du et al., 2009). In addition, methylation of the promoter

was also observed in circulating tumour cells (CTC) from primary breast cancer (Chimonidou et al., 2013).

Analysis of the promoter methylation was also reported for *SOX18* in primary samples of non-small cell lung cancer and surrounding tissues. This study showed an increased methylation in non-small cell lung cancer tissues compared to healthy lung tissues (Azhikina et al., 2011). In addition, Duong and colleagues studied the role of SOX18 in tumour lymphangiogenesis and observed that *SOX18* is expressed in lymphatic vessels during tumour growth in mice engrafted with melanoma cells. In contrast, partial loss of *Sox18* dramatically decreased the lymphatic vessel density and diameter. Moreover, the impairment in tumour neolymphangiogenesis correlated with a decrease in cancer cell metastasis (Duong et al., 2012). Recently, it has been shown that SOX18 is expressed in ductal breast carcinoma and ovarian cancer. Its high expression correlated with poor prognosis in both tumours and therefore it can be used as prognostic marker (Pula et al., 2014; Pula et al., 2013).

Sox7 is considered a tumour related gene as its expression is down-regulated in several types of cancer, in particular, prostate, colorectal, lung and breast cancer (Guo et al., 2008; Stovall et al., 2014). The restoration of *Sox7* expression induces colorectal cancer cell apoptosis, inhibits proliferation and colony formation suggesting that *Sox7* is a tumour suppressor gene in this specific cellular context (Zhang et al., 2009). A role for SOX7 as tumour suppressor is also reported in lung cancer. Indeed, its expression is down-regulated and associated with methylation of CpG islands in the upstream region of the *SOX7* gene (Hayano et al., 2013). Moreover, SOX7 expression can be used as a prognostic marker in patients with lung adenocarcinoma since decreased SOX7 level correlated with poor prognosis (Li et al., 2012a). The hypermethylation of *SOX7* promoter was also observed in breast cancer cell lines. In these cell lines the re-

expression of *SOX7* inhibited proliferation *in vitro* and tumour growth *in vivo* (Stovall et al., 2013). Despite *SOX7* down-regulation in several tumours, one example has been reported for increased expression level for *Sox7* in pancreatic and gastric cancer cell lines (Katoh, 2002).

1.13. Leukaemogenesis

Leukaemogenesis is the pathological process underlying leukaemia development. Leukaemia is the term used to describe malignant cancers of the bone marrow and blood characterized by the uncontrolled increase in blood cells.

Leukaemias are clinically subdivided into a variety of categories. According to the classical division, these cancers are divided into acute and chronic forms. The acute form is characterized by the rapid increase of more undifferentiated blood cells that makes the bone marrow unable to produce healthy blood cells. In this type of leukaemias the treatment is immediately required because the malignant cells can enter into the bloodstream and spread to the body. The chronic forms of leukaemia are characterized by an excessive increase of more differentiated blood cells and, in this case, the treatment is not necessary immediately since the disease takes months or years to progress (Pui et al., 2012).

In addition, leukaemias are also subdivided into two sub-groups, lymphoblastic or lymphocytic and myeloid or myelogenous, according to which types of blood cells are involved in the malignancy. The term lymphoblastic indicates that the mutation takes place in cells that normally form lymphocytes. Instead, the myeloid term indicates that the affected cells derive from myeloid lineage: red cells, white cells and platelets. Today's leukaemia classification is based on the World Health Organization (WHO) classification that requires morphology, immuno-phenotyping, cytogenetic and

molecular studies (Vardiman, 2010). Thus, the four major types of leukaemia are Acute Myeloid Leukaemia (AML), Acute Lymphoblastic Leukaemia (ALL), Chronic Myeloid Leukaemia (CML) and Chronic Lymphoblastic Leukaemia (CLL). ALL is the most common paediatric cancer while the three other types occur mainly in adults. In particular, ALL occurs in 75-80 per cent of childhood leukaemias while AML occurs in 20 per cent. In contrast, 2 per cent of childhood leukaemias represent CML and only few cases of CLL are reported (Miller et al., 1995).

ALL were first subdivided in three different categories based on cell morphological feature following the French-American-British (FAB) criteria: L1, L2 and L3 (Bennett et al., 1976). However, the introduction of immuno-phenotype, cytogenetic and molecular genetic analysis in the diagnosis of leukaemia changed the FAB classification. Today's classification is based on the WHO classification that recognizes two main groups of ALL: precursor B cell and precursor T cell leukaemia/lymphoma. The most recurrent mutations in ALL are *TEL-AML1*, *E2A-PBX*, *BCR-ABL1*, *MLL-AF4* and hyperdiploidy (Collins-Underwood and Mullighan, 2010).

AML are subdivided in eight different categories accordingly to the FAB classification: M0-M7 (Bennett et al., 1976). As for ALL, the WHO classification integrates the immuno-phenotypic, cytogenetic and molecular characteristics to the FAB classification. For AML, the main molecular genetics mutations are translocations between chromosome 8 and 21 (*RUNX1/RUNX1T1* or *AML/ETO*), inversions in chromosome 16 (*CBFB/MYH11*) and translocations between chromosome 15 and 17 (*PML/RARA*) (Valk et al., 2004). Furthermore, mutations within the tyrosine kinase receptor FLT3 and c-KIT have been shown to confer AML in association with mutations in other genes such as *CEBPA*, *WT1*, *NRAS* and *NPM1* (Betz and Hess, 2010).

For CLL, the four most common chromosome abnormalities are deletion of 13q14**,** trisomy 12**,** deletion at 11q23 and 17p13 deletion or mutation that involves the *P53* tumour suppressor gene (Butler and Gribben, 2010). Finally, most CML are characterized by a translocation where parts of the two chromosomes 9 and 22 swap places. The translocation results in the production of a *BCR-ABL* fusion gene created by the Abelson tyrosine kinase *ABL1* gene on chromosome 9, region q34 and a part of the break point cluster *BCR* gene on chromosome 22, region q11 (An et al., 2010).

1.14. B cell-Acute Lymphoblastic Leukaemia

ALL has an annual incidence of 39.9 per million and occurs more than AML in children. Moreover, ALL shows a peak of incidence between the age of 2 and 5 years and is more frequent in male than female (Bhayat et al., 2009). Symptoms and signs of ALL are quite variable but include fever in about 50-60% of patients. Patients are neutropenic and anaemic resulting in fatigue, lethargy and bone pain. Moreover, patients are characterized by pallor, ecchymosis and bone tenderness. Liver, spleen and lymph nodes are often enlarged due to extramedullary haematopoiesis. Some patients also manifest ocular haemorrhages. Furthermore, the increase in white blood cell count ($>10x10^9$/L) and the decrease in platelet counts ($50x10^9$/L) resulting in thrombocytopenia are confirmed by blood analysis (Pui, 2012b).

Diagnosis can be done only with blood analysis but most frequently is in association with bone marrow aspirate. Fibrosis of the bone marrow can cause problems during the aspiration. Therefore, biopsy procedures became necessary. A morphology analysis is associated with the analysis of the immuno-phenotype of the leukaemic cells for a definitive diagnosis (Pui and Evans, 2006).

ALL are primarily induced by the mis-expression of oncogenes, mutated

transcription factors and fusion proteins generated by chromosomal translocations. These alterations transform haematopoietic stem or progenitor cells into leukaemic cells altering the balance between differentiation and self-renewal capacity.

The most frequent mutation occurring in childhood B-ALL leukaemias are hyperdiploidy (25%), TEL-AML1 (22%), MLL rearrangement (8%), E2A-PBX1 (5%), BCR-ABL (3%) and MYC (2%). The frequency of these alterations is slightly changed in adult B-ALL leukaemia. 25% of ALL patients have BCR-ABL, 10% MLL rearrangements, 7% hyperdiploidy, 4% MYC, 3% E2A-PBX and only 2% TEL-AML1 (Figure 1.18) (Pui et al., 2004).

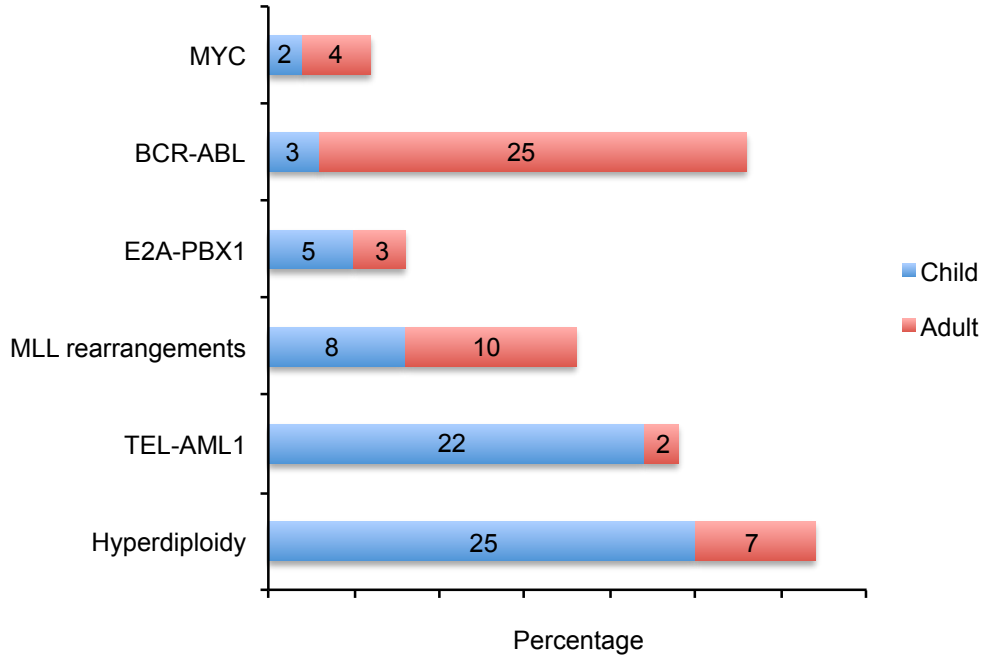


Figure 1.18. Frequency of the most common mutations in child and adult ALL.
The blue bar represents mutations in childhood ALL while the red bar shows mutations in adult ALL. Percentages corresponding each mutation are reported on the bars. The most common mutation in childhood ALL are hyperdiploidy and TEL-AML1 while BCR-ABL is mainly detected in adult ALL.

The translocation between chromosome 12 p13 and 21 q22 forms the fusion protein TEL-AML1, also known as ETV6-RUNX1, and is associated with good prognosis in ALL. The protein-protein interaction domain of the TEL is fused to the DNA-binding and transcriptional regulatory domain of AML1. TEL is part of the ETS family transcription factors and it has been found to rearrange not only with *AML1* but also with other genes such as *ABL, JAK2* (Zelent et al., 2004). RUNX1 is a transcription factor belonging to the RUNX transcription factor family together with RUNX2 and RUNX3. It is characterized by the DNA-binding RUNT domain and forms a complex with CBFβ. This complex is essential in haematopoiesis and has a main role in both lymphoid and myeloid lineages (Speck, 2001).

The fusion protein binds to the DNA, recruits histone deacetylases and therefore inhibits transcriptional activity. Over-expression of *Tel-aml1* in bone marrow cells impaired B cell differentiation leading an increase in pre and pro B cells *in vivo*. Moreover, an accumulation of $SCA1^{+}cKIT^{+}$ cells has been reported together with an increased self-renewal capacity *in vitro*. Despite this increase in progenitors, the TEL-AML1 fusion protein is not sufficient to induce leukaemia *in vivo* (Fischer et al., 2005). In another model, retroviral transduction of *Tel-aml1* in bone marrow cells induced transplanted mice to develop ALL; however, only one in nine mice developed B-ALL (Bernardin et al., 2002). *Tel-aml1* transgenic mice have been generated in which the expression of the fusion gene was under the control of the immunoglobulin heavy chain promoter. These mice did not develop leukaemia suggesting the requirement of other genes or mutation for leukaemogenesis (Andreasson et al., 2001). Indeed, Li and colleagues showed that low-dose irradiation could accelerate the formation of leukaemia. Similarly, loss of *Cdkn2a* in *Tie2-Cre/Tel-aml1* mice decrease the survival rate to 128 days compared to 164 days for *Cdkn2a* null mice (Li et al., 2013). Recently, another mouse model has

been developed for TEL-AML1 in which the expression of the fusion gene is under the control of the *Etv6* promoter. Since *Tel-aml1* mice did not alter the haematopoiesis, they were then crossed with *Sleeping Beauty* transposase. The generated mice developed B-ALL showing an accumulation of pro B cells. Genomic DNA from those mice was also analysed for identifying potential mutations acting together with the fusion gene. They identified mutations in *Ebf1*, *E2a*, *Epor*, *Met*, *Cbpa* and *Rasgrf1*, all possible candidates to induce leukaemia together with *Tel-aml* (van der Weyden et al., 2011).

Translocations involving the mixed-lineage leukaemia (MLL) transcription factor are frequent in ALL. The fusion occurs between the N-terminal portion of MLL and the C-terminal part of one of more than 70 partners. MLL, orthologue of the Drosophila Trithorax, is a histone-lysine N-methyltransferase and regulates the appropriate expression of HOX genes during development. *Mll* knock-out mice die around E10.5 with hypoplasia of the branchial arches and involution of the maxillary process (Yu et al., 1998).

The MLL-AF4 translocation is associated with poor prognosis and a quick evolution of B-ALL. *AF4* gene is a member of AF4/LAF4/FMR2 gene family and encodes for a transcription factor (Nilson et al., 1997). Several studies have been done to analyse the aetiology of the disease. Analysis performed on identical twins suggested an in utero origin of the fusion between MLL and AF4 (Greaves, 2003). The presence of the MLL-AF4 translocation is important for the highly aggressive phenotype of B-ALL. Indeed, the inhibition of MLL-AF4 using short interfering RNAs induced impairment in proliferation and clonogenic potential in human leukaemia cell lines. In addition, HOX genes were decreased together with stem cell genes inducing haematopoietic differentiation. After transplantation, the inhibition of MLL-AF4 in leukaemia cell lines prolonged survival *in vivo* (Thomas et al., 2005b). The absence of MLL-AF4 human

models makes it difficult to understand the formation of B-ALL induced by this common translocation. In contrast, different murine models have been proposed but they do not recapitulate the human disease correctly. *Mll-af4* knock-in mice generated by homologous recombination, by Chen and colleagues, results in an increase proliferation of both myeloid and lymphoid lineages. Moreover, they developed mixed lymphoid and myeloid hyperplasia and B cell lymphoma only after a very long latency period (Chen et al., 2006). The difficulties in producing a good model for MLL-AF4 in mouse brought, for the first time, Montes and colleagues to study the impact of this translocation in human cells. The over-expression of *MLL-AF4* in human cord blood derived $CD34^+$ human stem progenitor cells (HSPC) induced an increase in clonogenic potential and proliferation *in vitro*. Upon transplantation *in vivo*, cord blood derived $CD34^+$ HSPC over-expressing *MLL-AF4* showed multi-lineage haematopoietic engraftment capacity and no differences in blood cell count and organs size and failed to initiate leukaemia (Montes et al., 2011). Similarly, *MLL-AF4* expressing human embryonic stem cells (hESC) activated the expression of HOX genes and differentiated towards hemogenic precursors displaying an endothelial fate to the disadvantage of haematopoietic commitment. In addition, the *MLL-AF4* expression in hESC derived $CD45^+$ haematopoietic cells did not induce proliferation *in vitro* and failed to engraft *in vivo* (Bueno et al., 2012). Overall, these studies suggested that MLL-AF4 translocation is not sufficient to induce leukaemia. Targeting the appropriate cell type, inducing MLL-AF4 expression at physiological level and providing secondary mutation, such as FLT3, might be required for the development of the disease.

The pre B cell leukaemia homeobox (PBX1) protein is one HOX-DNA-binding co-factor that generally fuses to the transcriptional trans-activator domain of the E2A (also known as TCF3) transcription factor resulting in the production of E2A-PBX1

fusion protein. This alteration disrupts the expression of *E2a* target genes and *Hox* genes expression. When this fusion protein was over-expressed in mouse bone marrow cells, no alteration in cell growth *in vitro* were observed. Moreover, upon transplantation, these cells resulted in the formation of AML suggesting that E2A-PBX1 retains a leukaemic potential not limited to B cell malignancy (Kamps and Baltimore, 1993). Mice carrying *Eu-E2a-Pbx1* construct died in 5 months developing lymphoma. In pre-malignant mice, a decrease in both B and T cells was observed together with an increase in apoptotic cells in the thymus (Dedera et al., 1993). Later on, Bijl and colleagues proposed another model for E2A-PBX1 where the expression of the fusion gene is under the lymphoid-specific promoter/enhancer elements. Only reduced numbers of T cells were observed in these mice at two months of age. At later stage, mice developed B-ALL. When crossed with $CD3e^{-/-}$ to reduce T cell tumours, *E2a-Pbx1/*$CD3e^{-/-}$ mice developed B-ALL. The appearance of B-ALL was accelerated by pro-viral insertional mutagenesis that induced *Hoxa* genes over-expression suggesting the requirement of secondary mutation for the acceleration of leukaemia formation (Bijl et al., 2005).

The Philadelphia chromosome is the most common translocation observed in CML but it is also seen in acute leukaemia, and results from the fusion between the *BCR* and *ABL* genes. The *BCR* gene is located on chromosome 22 and encodes for a protein that has serine/threonine kinase activity, while the *ABL* gene is on chromosome 9 and encodes for a non-receptor tyrosine kinase. The N-terminal oligomerization domain of BCR is essential for the constitutive activation of the ABL tyrosine kinase (Pendergast et al., 1991; Zhao et al., 2002). Two BCR-ABL fusion proteins can be produced from different breakpoints: p210 and p190. $BCR\text{-}ABL^{p190}$ mutant mice, generated by inserting the *BCR-ABL* sequence into the *Bcr* locus, developed B-ALL

with more than 30% of blast cells in the bone marrow (Castellanos et al., 1997). Interestingly, BCR-ABL^{p190} and BCR-ABL^{p210} fusion proteins cause different leukaemias. The former induces only B-ALL while the latter induces B but also T and myeloid leukaemias. In addition, the latency period for the appearance of the disease is short for BCR-ABL^{p190} and long BCR-ABL^{p210} (Honda et al., 1998; Voncken et al., 1995). Another model for BCR-ABL^{p210} has been described where the expression of the fusion gene is under the control of the cytoplasmic tyrosine kinase *Tec* promoter. The generated mice developed B-ALL and died 4 months after birth. In contrast, the progeny of one of these mice developed myeloproliferative disease with features of CML after one-year (Honda et al., 1998).

All these studies highlighted the complexity of modelling human disease showing in most cases that the chromosomal translocations are not sufficient to induce the formation of leukaemia. Therefore, secondary mutations are thought to occur and to accelerate the process of leukaemogenesis. Among them, over-expression of the FLT3 tyrosine kinase receptor induces an uncontrolled cell growth in ALL. Cases of ALL also report mutations or deletion of genes involved in the tumour suppressor retinoblastoma protein (RB) or TP53 pathways (Pui et al., 2004).

The treatment of ALL normally consists of three different phases: remission-induction, consolidation and continuation therapy. The first phase aims to eradicate more than 99% of leukaemia cells and to restore a normal haematopoietic system. Drugs typically used in this phase are glucocorticoids, vincristine, asparaginase or anthracycline. Two weeks after the start of the induction treatment, the number of blasts is measured in the patient's bone marrow to analyse the minimal residual leukaemia. When more than 1 % of minimal residual leukaemia is measured, treatment is intensified. The use of imatinib is also included in this phase in patients positive for

BCR-ABL. The second phase of the treatment aims at reducing relapse eradicating drug-resistant residual leukaemic cells. Some of the drugs used in this phase are corticosteroids, vincastine and asparaginase. The last phase of the treatments is normally used to prevent the relapse. The therapy usually lasts for about two years with the use of mercaptopurine and methotrexate every week. Cranial irradiation and dexamethasone, highly penetrable, are also administrated to high-risk patients to prevent central nervous system (CNS) relapse. Patients with high-risk ALL and with poor initial response to treatment might require allogeneic haematopoietic stem cell transplantation (Pui et al., 2004; Pui et al., 2008).

The increase of drug-resistant patients in the population and difficulties in targeting transcription factors with drugs result in the necessary need of new treatment protocols and therefore in the generation of new therapeutic strategies. The use of small molecules targeting RNA and histone deacetylation or DNA methylation has been developed (Thomas et al., 2006).

Briefly, the RNase enzyme Dicer generates small interfering RNA of 21-28 base pair long (siRNA) by cleaving long double-strand RNA. Then, the RNA-induced silencing complex (RISC) removes one strand from the siRNA duplex. The remaining strand binds to the complementary RNA sequence inducing the silencing of the target gene (Heidenreich, 2009). Specific siRNA synthesized to bind the *BCR-ABL* fusion gene were used to down-regulate the mRNA and protein level of the fusion gene both in CML and ALL cell lines and primary CML samples. The reduction in the BCR-ABL expression correlated with a decrease in proliferation in cell lines (Scherr et al., 2003). Moreover, Wohlbold and colleagues showed that the reduction of BCR-ABL protein induced by anti-*BCR-ABL* siRNA enhanced the sensitivity to imatinib in imatinib-resistant cells (Wohlbold et al., 2003). Another example of use of siRNA in ALL was

performed using human ALL cell lines carrying the *MLL-AF4* translocation. The inhibition of *MLL-AF4* expression induced apoptosis, differentiation and increase in survival *in vivo* (Thomas et al., 2005b). Altogether, these studies demonstrate a possible use of siRNA for the treatment of ALL.

DNA methylation is a process regulated by methyltransferases (DNMT) that catalyse the addition of one methyl group to cytosine. This process occurs often in CpG islands and modulates transcriptional regulation (Robertson, 2002). Histone deacetylation occurs when histone deacetylases (HDAC) deacetylate lysine residues on histone tails and induces transcription inhibition (Johnstone, 2002). Analysis on paediatric ALL and AML patients showed that HDAC activity was increased in primary leukaemia cells compared to healthy peripheral blood and bone marrow samples (Sonnemann et al., 2012). Since many genes are hypermethylated or silenced by histone deacetylases in cancer, the use of DNA demethylating agents or HDAC inhibitors provides a new therapeutic approach (Claus et al., 2005).

Another way to treat leukaemia is through the development of new therapies targeting leukaemia stem cells. The term leukaemia stem cells refer to a population that retain the ability of self-renewal and to generate the tumour (Bomken et al., 2010; Chen et al., 2013; Guzman and Jordan, 2004).

In contrast to AML where the leukaemia stem cells reside in the immature $CD34^{+}CD38^{-}$ population, in the ALL was more difficult to identify cells with immuno-phenotype similar to HSC. The self-renewal property was not only observed in $CD34^{+}CD19^{-}$ cells but also in more differentiated B cell precursors such as $CD34^{-}CD19^{+}$ (Castor et al., 2005; Kong et al., 2008; Le Viseur et al., 2008). Therefore, the hierarchy that mimics the normal haematopoietic development is not suitable for describing ALL.

1.15. SOX and leukaemia

The expression of most SOX proteins is linked to several solid tumours as described in the previous section, but some of them, in particular members belonging to group C and F, are also reported to act in non-solid tumours such as leukaemia and lymphoma disease.

Pu.1 expression levels are reduced in murine bone marrow cells transduced with *Sox4* resulting in AML. *Sox4* regulates *Pu.1* expression through the binding to the TCF-binding site in *Pu.1* (Aue et al., 2011). When mouse bone marrow *Pml/Rara* cells are retroviral transduced with *Sox4* and transplanted *in vivo*, cells become more aggressive and mice developed AML faster than the control, *Sox4* or *Pml/Rara* alone (Omidvar et al., 2013). Another study reported the role of *Sox4* in AML. Knock-down of *Sox4* in *C/ebpα* deficient progenitor cells reduced their self-renewal activity while induced their differentiation toward macrophages. Moreover, inactivation of C/EBPα resulted in the over-expression of *Sox4* that induced leukaemogenesis in AML (Fung et al., 2013; Zhang et al., 2013).

As already reported, *Sox4* is important for early B cell development and therefore its mis-expression might be involved in B cell disease. It has been recently shown that *SOX4* is highly expressed in BCR-ABL1 ALL and correlated with poor prognosis. *Sox4* promoter was hypomethylated and *Sox4* expression induced the activation of PI3K/AKT and MAPK signalling (Ramezani-Rad et al., 2013). In addition, Ma and colleagues demonstrated that *Sox4* is important for the proliferation of leukaemia cells together with *Tcf7l1*. Deletion of *Sox4* in progenitor B cells reduced proliferation *in vitro* and *in vivo* with an increase in survival *in vivo* (Ma et al., 2014).

More than 90% of mantle cell lymphoma (MCL) cases expressed the transcription factor SOX11. Its expression is also associated with hairy cell leukaemia,

Burkitt lymphoma and immature lymphocytic neoplasms (Xu and Li, 2010). High expression level of *SOX11* are related to the presence of the t(11;14) translocation, *CCND1* expression and poor prognosis suggesting a possible use of *SOX11* expression as prognostic marker for distinguishing B-CLL from more aggressive MCL (Meggendorfer et al., 2013; Wasik et al., 2014).

In addition, some AML patients are characterized by the down-regulation of *SOX17*. In particular, low level of *SOX17* corresponds to short overall survival compared to *SOX17* high expressing patients suggesting the low *SOX17* expression as a risk factor (Tang et al., 2014). This down-regulation of *SOX17* in AML might be due to methylation of the *SOX17* promoter. The analysis of the WNT signalling inhibitor methylation performed by Griffiths and colleagues reported indeed *SOX17* methylation in 29% of primary AML samples (Griffiths et al., 2010).

Only one study reported *SOX7* expression in blood malignancy, in particular with myelodisplastic syndrome (MDS). This disease is characterized by a decrease in mature blood cells and often can evolve in AML. Fan and colleagues analysed the methylation status of CpG islands in MDS samples. They found that *SOX7* promoter was methylated and that the methylation status correlated with a decrease in survival suggesting the use of SOX7 methylation as a prognostic marker for MDS (Fan et al., 2012). In addition, Gandillet and colleagues showed, by using a Tet-on doxycycline-inducible expression system, that the enforced expression of *Sox7* in early committed haematopoietic precursors promoted their self-renewing capacity and blocked their differentiation to mature primitive and definitive lineages. Thus, the sustained expression of *Sox7* was sufficient to completely alter the balance between proliferation and differentiation at the onset of haematopoiesis (Gandillet et al., 2009).

1.16. Aim of the project

SOX7 is a transcription factor belonging to the SOX F group whose expression is modulated at the onset of blood specification. The blockage in differentiation and enhance in proliferation of haematopoietic precursors induced by *Sox7*-sustained expression (Gandillet et al., 2009; Serrano et al., 2010) has led to the hypothesis that the mis-expression of *Sox7* may confer a proliferative or survival advantage to adult cells and that *Sox7* may be implicated in the emergence or maintenance of cancer initiating cells.

Therefore, my project investigated whether *Sox7*-enforced expression confers a proliferative advantage to adult cells, whether this may lead to leukaemogenesis and whether this over-expression is associated with cases of human leukaemia. In particular in this project, I investigated in a mouse model, both *in vitro* and *in vivo*, the consequences of *Sox7*-enforced expression on the homeostasis of adult haematopoiesis and the induction of haematological malignancies. Then, since the expression of *SOX7* correlates with B-ALL, my project also aimed to analyse the consequences of *Sox7*-enforced expression on B cell differentiation in adult bone marrow cells. Finally, I analysed the requirement of *SOX7* for leukaemia formation in human leukaemia cell lines both *in vitro* and *in vivo*.

In summary, the work presented in this thesis demonstrates for the first time that *SOX7* is expressed in B-ALL and its enforced expression blocked B cell differentiation at an immature stage with a stem cell signature. Moreover, the down-regulation of *SOX7* in human leukaemia cell lines induced a decrease in proliferation both *in vitro* and *in vivo*. Understanding the mechanism by which *Sox7*-enforced expression blocks B cell differentiation might improve the development of new drugs for leukaemia treatment.

Chapter 2

2. Material and methods

2.1. DNA and RNA manipulation techniques

2.1.1. Genomic DNA isolation

For mouse genotyping analysis, genomic DNA was extracted starting from 20,000 cells from bone marrow or spleen. Cells were re-suspended in 25 µl of PBS and were incubated at 95°C for 8 minutes. Then, 1 µl of 20 mg/ml proteinase K and 5 µl of 0.1X PBS were added to the cells and were incubated at 55°C for 30 minutes. Proteinase K was heat-inactivated at 95°C for 8 minutes. 1 µl of extracted genomic DNA was used as template DNA for genotyping PCR using primers (Sigma) and GoTaq DNA polymerase mix (Promega).

For cloning procedure, genomic DNA was extracted using the Wizard genomic DNA Purification Kit (Promega) according to manufacturer's protocol. Briefly, cells were lysed using cell lysis solution and proteins were removed by the salt-precipitation step. The genomic DNA was concentrated by isopropanol precipitation and finally re-suspended in water.

Then, 100 ng of extracted genomic DNA was used as template DNA for cloning PCR using primers (Sigma) and GoTaq DNA polymerase mix (Promega) or Phusion DNA polymerase (Finnzymes).

2.1.2. RNA isolation and cDNA synthesis

Cells were collected by centrifugation and total RNAs were extracted using RNeasy Mini or Micro kit (Qiagen). Briefly, samples were lysed and homogenized. Then, the lysate was added onto the RNeasy silica membrane and washed to remove any contaminants. Finally, the RNA was eluted in water and measured using a Nanodrop ND 1000 spectrophotometer (Thermo Scientific).

Up to 2 µg of RNA was reverse transcribed with random hexamers primer (Roche) to generate cDNA using the Omniscript Reverse Transcriptase kit (Qiagen), according to manufacturer's protocol. In summary, primers, RNase inhibitor, buffer, dNTPs and Omniscript reverse transcriptase were added to the template RNA and incubated for 60 minutes at 37°C.

For Affymetrix analysis, total RNAs were treated with DNase I to remove any DNA contamination using RNase-Free DNase (Qiagen) and further processed as described in section 2.1.3.

2.1.3. Polymerase Chain Reaction (PCR)

For genotyping, 1 µl of gDNA (~100ng) was incubated in a final volume of 10 µl with GoTaq DNA polymerase mix (Promega) and 0.5 mM of target-specific primers (Sigma) (Table 2.1). The thermal cycling conditions were as follows: 3 minutes at 95°C, followed by a three-step PCR for 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 1 minutes and a final extension at 72°C for 5 minutes.

For cloning, ~100 ng of gDNA was incubated in a final volume of 20 µl with 0.2 µl Phusion DNA polymerase (Finnzymes), 4 µl of 5X Phusion HF Buffer, 0.4 µl 10 mM dNTP and 0.5 mM of each primers (Sigma) (Table 2.2). The thermal cycling conditions were as follows: 30 seconds at 98°C, followed by a three-step PCR for 35

cycles of 98°C for 10 seconds, 64°C to 68°C for 30 seconds, 72°C for 1 minutes per Kb and a final extension at 72°C for 10 minutes. The presence of amplification was evaluated by running the samples on a 0.8-1% agarose gel.

Modification	Database number	Primer sequence	Product size (bp)
Cre	AW53 AW54	gcctgcattaccggtcgatgcaacga gtggcagatggcgcggcaacaccatt	band: 726
Sox7 $^{WT/flox}$	AX42 AV31 AX42 AV31	gggttaccgcacttaagagaca ggaagtcctacccgacctaatc gggttaccgcacttaagagaca ggaagtcctacccgacctaatc	WT band: 195 KI band: 345
Sox7 $^{WT/KO}$	AX42 AV26 AX42 AV26	gggttaccgcacttaagagaca agaatagtcagggctggtcaag gggttaccgcacttaagagaca agaatagtcagggctggtcaag	WT band: 1832 KO band: 357

Table 2.1. List of primers used for genotyping

Primer name	Database number	Primer sequence	Use
pGIPZ insert_F	AU35	gtatataaaattattca	cloning of shRNA-sffv-GFP-Luciferase vector
pGIPZ insert_R	AU36	agaggggcggaatttg	cloning of shRNA-sffv-GFP-Luciferase vector
pGIPZ-sFFV promoter_F	AU37	acagggacagcagatcca	cloning of shRNA-sffv-GFP-Luciferase vector
pGIPZ-sFFV promoter_R	AU38	ttgtacaaacttgtgattc	cloning of shRNA-sffv-GFP-Luciferase vector
pGIPZ-GFP_F	AU39	gctctttaaaggaaccaa	cloning of shRNA-sffv-GFP-Luciferase vector
pGIPZ-GFP_R	AU40	cagatcgaagttcagggt	cloning of shRNA-sffv-GFP-Luciferase vector
pGIPZ-mCherry_F	AU41	aagcaggctctttaaagg	cloning of shRNA-sffv-GFP-Luciferase vector
pGIPZ-mCherry_R	AU42	ttgctttccacatcgccggc	cloning of shRNA-sffv-GFP-Luciferase vector
pGIPZ-Luciferase1 F	AU43	acttcgatctgctgaaact	cloning of shRNA-sffv-GFP-Luciferase vector
pGIPZ-Luciferase1 R	AU44	atgatcggtagcttcttttg	cloning of shRNA-sffv-GFP-Luciferase vector
pGIPZ-Luciferase2 F	AU45	catcatcatggatagcaaga	cloning of shRNA-sffv-GFP-Luciferase vector
pGIPZ-Luciferase2 R	AU46	ggtgtccaagtccaccacct	cloning of shRNA-sffv-GFP-Luciferase vector
SalI-EBF1prox short_F	BJ13	gtcgacgctctcccgtctcctcc	cloning of EBF1 promoter
EcoRV-EBF1prox short_R	BJ14	gatatctgtcattttccacaaccag	cloning of EBF1 promoter
XhoI-Ef1-Tomato FW	BB48	ctcgagggctccggtgcccgtc	cloning of MLLAF4-EF1-Tomato vector
RsrII-Ef1-Tomato REV	BB49	cggtccgttacttgtacagctcg	cloning of MLLAF4-EF1-Tomato vector

Table 2.2. List of primers used for cloning. F stands for forward and R for reverse.

2.1.4. Quantitative real-time Polymerase Chain Reaction

Quantitative real-time PCR (qRT-PCR) reactions were performed in triplicate on a ABI 7900 system (Applied Biosystems) using the Exiqon Universal Probe Library (Roche) and Universal PCR Master Mix TaqMan (Applied Biosystems). The sequences of the primers used for quantitative RT-PCR are listed in Table 2.3.

For each sample, 1 µl cDNA was incubated in a final volume of 10 µl with Universal PCR Master Mix TaqMan (Applied Biosystems), 0.1 µl Exiqon Universal Probe Library (Roche) and 0.2 µl of both forward and reverse target-specific primers (Sigma). The thermal cycling conditions used to validate gene expression changes were as follows: hold for 10 minutes at 95°C, followed by a three-step PCR for 40 cycles of 95°C for 15 seconds, 55°C to 60°C for 25 seconds, and 72°C for 30 seconds. The expression of each target gene was evaluated using a relative quantification approach (-ΔCT method) with β-2microglobulin or β-actin as the internal reference for human and mouse genes, respectively. The reactions were performed in triplicates.

2.1.5. Plasmid construction

pGIPz vectors containing shRNA sequence against SOX7 were bought from Openbiosystem. The CMV promoter was replaced by the sFFv promoter. Two vectors containing shRNA sequence against SOX7 (SCBR1079 and SCBR1080) and the vector containing the sFFv promoter-GFP-2A-Luciferase construct (SCBR1207) were digested using SanDI (Fermentas) and NotI-HF (Biolabs) restriction enzymes. The digested samples were run on a 0.8% agarose gel for bands extraction. The bands (~ 9608 bp for the vector SCBR1079/1080 and ~ 3313 bp for the insert extracted from SCBR 1207) were extracted and purified using the QIAquick Gel Extraction Kit (Qiagen). Ligation was then performed using T4 ligase (Roche). Finally, bacteria were transformed and

Gene name	Database number	Primer sequence	Probe	Species
Sox7	Forward: AI25 Reverse: AI26	cagcaagatgctgggaaag tgcatcatccacatagggtct	#97	Mouse
VE-cadherin	Forward: AN34 Reverse: AN35	gttcaagtttgccctgaagaa gtgatgttggcggtgttgt	#56	Mouse
Gfi1	Forward: AA35 Reverse: AA36	tccgagttcgaggacttttg agcggcacagtgacttctc	#7	Mouse
C/EBPε	Forward: BD24 Reverse: BD25	tcccctgcagtaccaagtg gtgccttgagaaggggact	#74	Mouse
IRF8	Forward: BD26 Reverse: BD27	ccggcaagcaggattaca gctttgtctccctctttaaacttc	#10	Mouse
Pax5	Forward: BD16 Reverse: BD17	acgctgacagggatggtg ggggaacctccaagaatcat	#83	Mouse
E2A	Forward: AF26 Reverse: AF27	ctgagcagcccgtacctct aggggccatttcatctgtag	#22	Mouse
β-actin	Forward: H66 Reverse: H67	tgacaggatgcagaaggaga cgctcaggaggagcaatg	#106	Mouse
Sox7	Forward: AK18 Reverse: AK19	ctcagggcagggaggtct gcactcggataaggagagtcc	#57	Human
Sox17	Forward: AK10 Reverse: AK11	acgccgagttgagcaaga tctgcctcctccacgaag	#61	Human
Sox18	Forward: AK06 Reverse: AK07	gtgtgggcaaaggacgag agctccttccacgctttg	#28	Human
B2M	Forward: AK33 Reverse: AK34	atctgagcaggttgctccac gaccaagatgttgatgttggataa	#32	Human

Table 2.3. List of primers used for quantitative real-time polymerase chain reaction

colonies were processed for DNA extraction following the plasmid purification kit protocol (Qiagen). DNA was digested using AgeI and PshAI restriction enzymes and sequenced to confirm that the cloning was performed correctly and no mutation had been introduced. The same approach was used to generate the control vector. The pGIPz control vector containing the scramble sequence (SCBR1245) was generated by ligating the ~ 9608 bp fragment from the vector containing EF1-Tomato-Scramble sequence (SCBR1247) and the ~ 3313 bp fragment from the vector containing sFFv promoter-GFP-2A-Luciferase construct (SCBR1207). The new vectors, SCBR1201, SCBR1202 and SCBR1245 (thereafter referred to as shSOX7_2, shSOX7_3 and Scramble) were used for transfecting leukaemia cell lines.

For the generation of the MSCV vector containing the human MLL-AF4 sequence, the EF1-Tomato sequence was amplified by PCR from the SCBR1247 vector with primers containing XhoI and RsrII restriction sites. Then, the amplified fragment was inserted in the TOPO vector using the TOPO-TA cloning kit (Invitrogen) in order to simplify the cloning process. The pMSCV-flag-MLL-plAF4-ESC vector (SCBR1316) was cut with RsrII and XhoI restriction enzymes and ligated with the EF1-Tomato fragment. The correct expression of the protein from the MSCV-MLL-AF4-EF1-Tomato vector was validated in transduced HEK293 cell line.

In order to generate a Luciferase reporter vector, the human EBF1 promoter region was amplified by PCR from RS4;11 gDNA using specific primers (Table 2.2). The amplified fragment was inserted in the pCR 2.1-TOPO vector using the TOPO-TA cloning kit (Invitrogen). Briefly, the amplified fragment was incubated together with salt solution and pCR 2.1-TOPO vector for 5 minutes at room temperature and then transformed into competent *E.Coli* cells. Successively, the EBF1 promoter fragment and the PGL4 luciferase vector (SCBR1064) were cut with EcoRV and KpnI and the

Right fragments were ligated together.

All constructs were sequenced on a 3130 Genetic Analyzer (Applied Biosystems).

2.1.6. Gene expression microarray, Ingenuity pathway analysis, GSEA analysis

Total RNA was extracted using the RNeasy Mini kit (Qiagen) and tested for RNA quality on an Agilent Bioanalyzer 2100 (Agilent Technologies). The RNA was incubated at 65ºC for 2 minutes to denature prior to loading onto the chip. Amplified cDNA was produced from the total RNA samples using the Ovation Pico WTA System V2 (NuGEN Technologies; 3302-60). The amplified cDNA was then fragmented and labelled using the Encore Biotin Module (NuGEN Technologies; 4200-60). Then, the cDNA library was hybridized to Affymetrix mouse exon 1.0 ST array. Background adjustment and data normalization was performed with the Robust Multi-array Average (RMA) algorithm. The probesets were subsequently filtered by Affymetrix detection above background (DABG) scores. A probeset was kept if its DABG p-value was less than 0.01 in at least 2 samples. After these filtering steps, 189,956 probesets (representing 22,140 unique genes) remained for differential gene expression analysis. Annotations and cross-mappings between probesets, exons and genes were obtained from Ensembl mouse genome build 70 using the R/BioConductor package Annmap (Bioconductor 2.14). The microarray analysis was performed by the Molecular Biology core facility and the data normalization was performed by Hui Sun Leong, member of the Bioinformatics group at the CRUK-Manchester Institute.

The computational method, Gene Set Enrichment Analysis (GSEA), was used to determine a priori defined set of genes statistically significant and concordant

differences between the two biological samples: $Sox7^{low}$ and $Sox7^{high}$ (Subramanian, Tamayo, 2005; Mootha, Lindgren, 2003). Ingenuity IPA software was utilized for analysing and accurately interpreting the biological meaning in our genomic data (Ingenuity).

2.2. Protein expression and manipulation assay

2.2.1. Protein extraction and Western blotting

Protein samples were isolated using the home-made lysis buffer (50 mM, NaCl 150 mM, 0.1% SDS, 0.5% NaDeoxyclorate, 1% NP40, 1:1000 Protease inhibitor cocktail (Sigma)) and quantify using Bradford assay (BioRad). Electrophoresis was carried out using 10% NuPAGE Novex Bis-Tris Mini Gels and NuPAGE MOPS SDS Running Buffers (Invitrogen). 50-100 mg of protein extracts were loaded into the polyacrylamide gel together with Loading Buffer 4X and Reducing Agent 10X and electrophoresed for 45 minutes at 50V, then for 1 hour at 120V. The SeeBluePlus2 1X marker was used as molecular weight indicator. Gels were blotted onto nitrocellulose membranes using Mini iBlot Gel Transfer Stacks Nitrocellulose (Invitrogen). The nitrocellulose membranes were incubated for 5 minutes with Ponceau S solution (Sigma) for rapid reversible detection of protein bands. Membranes were washed with PBS with 0.1% Tween-20 and non-specific binding was blocked by overnight incubation at 4°C in blocking buffer (PBS with 0.1% Tween-20 and 5% fat-free powdered milk (Marvel)).

Subsequently, the membranes were incubated with an anti-SOX7 sheep polyclonal antibody (home-made) overnight at 4°C. After six washes of 10 minutes each with PBS-0.1% Tween-20, the membranes were incubated with a secondary HRP

labelled donkey anti-sheep antibody (Table 2.5). After six washes of 10 minutes each with PBS-0.1% Tween-20, signal was developed using the ECL Western Blotting Detection Reagent (GE-Healthcare Bio-Sciences). Films were developed using a MAS automated developing machine.

2.2.2. Immunohistochemistry staining

Organs were fixed in 4% paraformaldehyde overnight at 4°C. Paraffin sections were cut at a thickness of 10 μm, de-waxed in xylene twice for 5 minutes, re-hydrated through graded alcohol solutions for 1 minute in each and then washed with running water for 2 minutes. Reticulin staining was performed using the Gordon and Sweet's stain. To this end, sections were incubated for 5 minutes in a potassium permanganate (3% sulphuric acid) solution followed by washes in tap water. Next, 1.5% oxalic acid was applied until clear. After washes in tap water, sections were incubated with 2% ferric ammonium sulphate for 15 minutes. Multiple washes in distilled water were performed before applying the ammoniacal silver solution (10% silver nitrate, ammonium hydroxide and 3% sodium hydroxide solution). Sections were washed in distilled water and then were fixed in 10% formalin for 5 minutes. After washes in tap water, slides were stained with 5% sodium thiosulphate for 2 minutes to remove unreduced silver. Sections were first rinsed in tap water and incubated in 0.2% gold chloride for 3 minutes and then after washes in tap water, 0.1% neutral red was applied for 10 second. Finally, slides were dehydrated, cleared and mounted on clear glass covered by a coverslip.

The CD45R (B220) staining was performed using the following steps. Sections were incubated with Dako target retrieval solution (S2369) at 125°C for 1 minute, and then cooled to 90°C for 10 seconds. Slides were brought back to room temperature by

placing them in running water for 3 minutes. Each section was surrounded with a pap pen (brand) to ensure sections did not dry. Endogenous peroxidase was blocked in 0.3% TBS for 10 minutes. After washing thoroughly in running water for 5 minutes, blocking solution (10% rabbit serum) was applied for 30 minutes. Blocking solution was removed and antibody anti-CD45R (BD Biosciences) was applied overnight at 4°C or Rat IgG2b antibody at equivalent concentration as isotype control. After washing twice in TBS for 5 minutes, a rabbit anti-rat biotinylated antibody (Dako) was applied for 30 minutes at room temperature (Table 2.4). Slides were washed twice in TBS for 5 minutes and were stained with Strep-AB-Complex–HRP (Dako K0377, 9 µl solution A, 9 µl solution B and 1 ml TBS). Two washes in TBS for 5 minutes were performed before applying the commercial DAB solution (K3468, 1 drop of chromagen to 1 ml buffer) to the sections for 5 minutes. Sections were well rinsed in running water for 3-5 minutes and then counterstained with 1X Gills Haematoxylin for 1 minute, followed by washing in water for 1 minute, dipped in alkaline water for 30 seconds, washed again in water for 1 minute. Finally, slides were dehydrated, cleared and mounted on clear glass covered by a coverslip. Pictures were taken using a slide scanner machine (Laica SCN400).

These immunohistochemistry staining were performed by members of the Histology facility at the CRUK-Manchester Institute.

Antibody	Host	Species	Manufacturer	cat. Number	Dilution	Incubation	Use
anti-Sox7	Sheep	Mouse	Home made	-	1/2000	ON 4°C	WB
Secondary HRP	Donkey	Sheep	Sigma	A3415	1/5000	1h RT	WB
anti B220	Rat	Mouse	BD Biosciences	553084	1/10	ON 4°C	IHC

Table 2.4. List of antibodies for western blotting (WB) and immunohistochemistry (IHC)

2.2.3. Luciferase reporter assay

HEK293 cells were seeded at a density of 50,000 cells per well onto 24-well plates. The next day cells were washed and transfected with the DNA mixes using GeneJuice kit (Novagen) according to protocol's manufacturer. Each DNA mixture contained 0.3 ng of Renilla vector (SCBR1066) and 50 ng of Firefly Luciferase reporter construct (SCBR EBF1promoter) and a total of 140 ng expression constructs in pcDNA vector. The DNA mixture was added to the GeneJuice transfection reagent and then, after 5-15 minutes incubation at room temperature, was dropped wise to cells. Forty-eight hours after the transfection, cells were lysed in 200 µl of lysis solution. 10 µl of lysate were used for measurement using the Dual-Luciferase reporter assay system (Promega). The Luciferase activity was acquired using the GloMax-Multi Detection System machine (Promega). All Firefly luciferase values were normalized to the respective Renilla internal control.

2.3. Cell culture and maintenance

2.3.1. Culture of Human Embryonal Kidney 293 T (HEK293), Platinum-Ecotropic retroviral packaging -human embryonal kidney (Plat-E) and human cervix carcinoma (HeLa)

HEK293, Plat-E and HeLa cell lines were cultured in Dulbecco's modified Eagle medium (DMEM, Lonza) supplemented with 10% foetal calf serum (FCS, PAA Laboratories), 2 mM L-Glutamine (Gibco) and 50 µg/ml penicillin/streptomycin (Gibco). All cell lines were cultured at 37°C in a humidified 5% CO_2 incubator.

2.3.2. Culture of mouse stromal cell line: OP9 and OP-DL1

OP9 and OP9-DL1 cells were expanded in alpha-minimum essential medium (α-MEM, Lonza) supplemented with 20% foetal calf serum (FCS, PAA Laboratories), 50 μg/ml penicillin/streptomycin (Gibco) and 2 mM L-Glutamine (Gibco) and cultured at 37°C in a 5% O_2%-5% CO_2 incubator.

2.3.3. B cell-OP9 co-culture

OP9 stromal cell line supports *in vitro* the differentiation of B cells. Prior to co-culture, OP9 cells were expanded as described above in section 2.3.2. Before the start of the co-culture, OP9 cells were seeded onto 6-well plates at a density of 40,000 cells per well and mitotically inactivated by irradiation at 30 Grays (Gy) for 21.9 minutes. Single cell suspensions obtained from bone marrow were plated on irradiated OP9 in RPMI (Roswell Park Memorial Institute, Lonza) supplemented with 20% foetal calf serum (FCS, PAA Laboratories), 2 mM L-Glutamine (Gibco), 5 μg/ml Kit Ligand (KL), 2 μg/ml Interloukin-7 (IL-7) and 5 μg/ml FLT3 (all from Prepotech) and cultured at 37°C in a humidified 5% CO_2 incubator. Twice a week cells were harvested, counted and re-plated onto fresh irradiated OP9 cells.

2.3.4. Culture of leukaemia cell lines

Sup-B15 (BCR-ABL (e1-a2)), REH (TEL-AML1), RS4;11 (MLL-AF4 (e10-e4)), BV173 (BCR-ABL), 697 (TCF3-PBX1), SD1 (BCR-ABL) and NALM6 (PDGFRβ-ETV6) are human B cell precursor leukaemia cell lines, MV4-11 (MLL-AF4 (e9-e5)) is a human acute monocytic leukaemia cell line. The Sup-B15, MV4-11 and REH cell lines were expanded in RPMI (Roswell Park Memorial Institute, Lonza) supplemented with 10% foetal calf serum (FCS, PAA Laboratories) and 1% L-

Glutamine (Gibco). The SD1 and NALM6 were cultured in the same medium with the addition of 50 µg/ml penicillin/streptomycin (Gibco). The BV173 and 697 were cultured in the medium described above with 20% foetal calf serum (FCS, PAA Laboratories), 1% L- Glutamine (Gibco) and 50 µg/ml penicillin/streptomycin (Gibco). The RS4;11 cell line was expanded in alpha-minimum essential medium (α-MEM, Lonza) supplemented with 10% foetal calf serum (FCS, PAA Laboratories) and 2 mM L-Glutamine (Gibco). All the cell lines were cultured at 37°C in a humidified 5% CO_2 incubator.

2.3.5. Culture of primo-graft B-ALL samples

CD13 5.2 (very high-risk sample-VHR03), CD13 11.2 (standard-risk sample-SR03), CD19 3.1 (high hyperdiploidy), CD20 1.1 (ETV6-RUNX1), 694E (high level of minimal residual disease-MDR hi) and TR (high hyperdiploidy) are primo-graft B-ALL samples. These samples have been engrafted and amplified *in vivo*. These human samples were expanded in StemSpan medium (Stem cell technology) supplemented with 10% foetal calf serum (FCS, PAA Laboratories), 2 mM L-Glutamine (Gibco), 25 µg/ml penicillin/streptomycin (Gibco), 50 ng/ml human stem cell factor (hSCF), 50 ng/ml mFLT3, 20 ng/ml hTPO, 20 ng/ml hIL3, 10 ng/ml hIL6 and 10 ng/ml mIL7 (All from Prepotech). All primo-graft samples were cultured at 37°C in a humidified 5% CO_2 incubator.

2.3.6. Proliferation assay

A total of 0.25-1 million cells were plated in a 6-well plate in the presence of their growth medium. Cells were counted every 3 days and re-seeded in their growth medium for about 10-12 days. Data are shown as the mean of absolute cell number from

two or three wells.

2.3.7. Mouse haematopoietic colony forming assay

Single cell suspensions obtained from bone marrow were plated at a density of 40,000 cells per 35 x 10 mm dishes (BD Falcon) in semi-solid medium supplemented with haematopoietic cytokines. The media contained 55% methylcellulose (10 g/L), 10% serum (Stem Cell Technology), 10% protein-free hybridoma medium (PFM, Gibco), 2 mM L-Glutamine (Gibco), 180 μg/ml transferrin, 0.5 mM ascorbic acid, $4.5x10^{-4}$ M MTG, 1% KL, 1% IL-3, 1% thrombopoietin conditioned medium, 1 ng/ml granulocyte-macrophage colony-stimulating factor (GM-CSF), 5 ng/ml IL-11, 2 U/ml erythropoietin (Ortho-Biotech), 5 ng/ml IL-6, 10 ng/ml macrophage colony-stimulating factor (M-CSF) (all from R&D system) and Iscove's Modified Dulbecco's Medium (IMDM, Lonza). When indicated, 1 ug/ml of doxycycline (Dox) was added to the semi-solid medium. Haematopoietic colonies were counted according to their morphology. For serial clonogenic replating assay colonies were dissociated with PBS and re-plated in fresh semi-solid medium every 3-4 days. For haematopoietic colony forming assay, colonies were counted after 10-12 days in culture. Data are shown as the mean number of colonies from three dishes and bars represent standard deviation of the mean.

2.3.8. Human colony forming assay

Human leukaemia cell lines were plated in the α-MEM (Lonza) containing 55% methylcellulose (10 g/L), 10% serum (Stem Cell Technology), 10% protein-free hybridoma medium (PFM, Gibco), 2 mM L-Glutamine (Gibco), 180 μg/ml transferrin, 0.5 mM ascorbic acid, $4.5x10^{-4}$ M MTG, 1% mouse KL, 10 ng/ml human IL-7, 200 ng/ml human IL-3, 3 ng/ml human IL-6 and 10 ng/ml hSCF (All from Peprotech).

Cultures were maintained at 37°C, 5% CO_2. Haematopoietic colonies were counted according to their size and morphology. Data are shown as the mean number of colonies from three dishes and bars represent standard deviation of the mean.

2.3.9. Flow cytometry analysis and cell sorting

Single cell suspensions from adult bone marrow, B cell culture, haematopoietic colonies, mice blood samples or human cell line and primary samples were stained and analysed with FACSCalibur or LSRII and sorted with Influx or Aria flow cytometers (all BD Biosciences).

Staining for sorting was performed in IMDM with 10% FCS, whereas cell surface staining for analysis was performed in PBS with 10% FCS. Cells were incubated with the primary antibodies for 30 minutes on ice or at 4°C, then washed in PBS with 10% FSC and stained with the secondary antibodies for 30 minutes on ice or at 4°C. After the secondary staining, cells were washed in PBS with 10% FSC and re-suspended in PBS with 10% FSC for cell surface staining or IMDM with 10% FCS for sorting. Monoclonal antibodies and streptavidin used for staining are listed in table 6. Data were analysed using the FlowJo software (TreeStar).

For magnetic activated cell sorting (MACS) of cKIT positive cells or lineage negative ($MAC1^-$, $GR1^-$, $B220^-$, $CD4^-$ and $TER119^-$) cells, cells were collected by centrifugation and washed in PBS with 5% FSC. Ten million cells of single cell suspension from adult bone marrow were incubated in 50ul of PBS with 5% FSC with biotin-labelled antibodies (eBioscience, Table 2.5) for 20 minutes on ice. Then, cells were washed in PBS with 5% FSC and incubated with anti-biotin magnetic beads (Miltenyi Biotec) for 20 minutes on ice. After washing, biotin-antibody labelled cells were isolated using LS Separation Columns (Miltenyi Biotec) following the

Antibody	Fluorochrome	Manufacturer	Dilution	Species
IgM	FITC	eBioscience	1/200	Mouse
AA4.1	FITC	eBioscience	1/200	Mouse
MAC1	PE	eBioscience	1/200	Mouse
GR1	PE	eBioscience	1/600	Mouse
B220	PE	eBioscience	1/200	Mouse
CD24	PE	eBioscience	1/200	Mouse
SCA1	PE	eBioscience	1/200	Mouse
TER119	PE	eBioscience	1/300	Mouse
CD4	PE	eBioscience	1/200	Mouse
CD23	PE	eBioscience	1/200	Mouse
CD45.2	PE	eBioscience	1/600	Mouse
BP-1	biotin	eBioscience	1/100	Mouse
MAC1	biotin	eBioscience	1/300	Mouse
GR1	biotin	eBioscience	1/300	Mouse
B220	biotin	eBioscience	1/300	Mouse
CD4	biotin	eBioscience	1/300	Mouse
TER119	biotin	eBioscience	1/300	Mouse
CD71	biotin	eBioscience	1/1000	Mouse
CD25	biotin	eBioscience	1/300	Mouse
IgD	biotin	eBioscience	1/100	Mouse
CD45	biotin	eBioscience	1/200	Human
CD19	PE-Cy7	eBioscience	1/300	Mouse
Strepavidin	PE-Cy7	eBioscience	1/300	Mouse
GR1	PE-Cy7	eBioscience	1/500	Mouse
cKIT	APC	eBioscience	1/100	Mouse
CD43	APC	eBioscience	1/200	Mouse
IgM	APC	eBioscience	1/100	Mouse
CD8	APC	eBioscience	1/300	Mouse
CD41	APC	eBioscience	1/200	Mouse
CD45.1	APC	eBioscience	1/300	Mouse
CD45.2	APC	eBioscience	1/200	Mouse
VE-Cadherin	Alexa-Fuor647	eBioscience	1/200	Mouse
CD19	Pacific blue	eBioscience	1/300	Mouse
CD19	Pacific blue	eBioscience	1/100	Human
CD21/CD35	eFluor 450	eBioscience	1/100	Mouse
B220	eFluor 450	eBioscience	1/100	Mouse

Table 2.5. List of antibodies used for flow cytometry analysis

manufacturer's protocol. Briefly, cell suspension was applied onto LS Column, previously inserted into an MACS Separator. LS column was washed two times with PBS buffer and all unlabelled cells were collected. When interested in labelled cells, 5 ml of PBS buffer was added to the LS column, previously removed from the MACS separator, and labelled cells were flushed out from the column.

2.3.10. Lentivirus production and transduction

Non-silencing and *Sox7* shRNA were purchased from Open Biosystems and expressed in a modified version of pGIPZ with a sFFv promoter replacing the CMV promoter as described in the plasmid construction section. For the second-generation lentivirus production, HEK293 cells were seeded onto 150 x 25 mm dishes (BD Falcon) at a density of about 40,000 cell/cm^2. Mix of 20 μg of pGIPZ vector containing non-silencing and *Sox7* shRNA, 13 μg packaging (pCMVΔ8.9) plasmid and 7 μg envelope (pMDG.2) plasmid were precipitated by adding 1/18 volume of 5M NaCl and two volumes of 100% ethanol. The next day, cells were transfected according to BD CalPhos Mammalian transfection kit protocol (Clontech). Briefly, 2X HBS solution was slowly vortexed while adding the solution composed by the mix of plasmids, 2 M Calcium solution and sterile water. The transfection solution was gently mixed and then added dropwise to plated HEK293 cells. Cells were incubated at 37°C in a humidified 5% CO_2 incubator for twenty-four hours. On day 3, medium was removed from the plates and replaced with fresh medium. On day 4, 5 and 6 virus supernatant was collected, centrifuged at 1000-2000 rpm for 10 minutes at 18 °C. Then, virus supernatant was filtered over a 0.45 μm filter (BD Falcon) into a sterile bottle (BD Falcon) and store at 4°C for no more than a week. Finally, produced viruses were concentrated by single or double ultra-centrifugation (Sorvall). First, all viruses were

pelleted at 20,000 rpm at 4°C for two hours. Supernatants were discarded by decanting; pellets were re-suspended in PBS or in appropriate cell medium and aliquots were stored at -80°C. Viral titres were expressed as HeLa-cell-transducing units per millilitre (TU/mL). Viruses were used to transduce cell line according to their titration.

The day before the transduction, leukaemia cell lines were split. On day 2, 150,000 cells were plated in one well of a 12-well plate and transduced at a MOI of 3-50 in culture media with the addition of polybrene (8 μg/ml) by spinoculation for 30 minutes at 1200 rpm at room temperature. Cells were incubated at 37°C in a humidified 5% CO_2 incubator for about two hours, then washed with PBS and re-plated into fresh medium. Seventy-two hours later, cells were analysed for GFP expression by FACS. When required, GFP positive cells were sorted or selected by addition of puromycin (0.4-0.8 μg/μl).

2.3.11. Cell cycle analysis

The cell cycle staining was performed according to Click-iT EdU Flow Cytometry Assay Kits manufacturer's protocol (Invitrogen; AI0202). One million cells were incubated with 10 μM EdU (5-ethynyl-2-deoxyuridine, nucleoside analogue to thymidine) for 10 minutes to 4 hours depending on the cell dividing time. Then, cells were harvested and fixed with Click-iT fixative solution for 15 minutes at room temperature, protected from light. After washing in 1% BSA with PBS, cells were permeabilized adding 1X saponin-based permeabilization and wash buffer. Click-iT reaction cocktail containing the fluorescent dye azide (Alexa Fluor 647) was prepared according to manufacturer's protocol and incubated for 30 minutes at room temperature, protected from light. Cells were washed with 1X saponin-based permeabilization and wash buffer and incubated for DNA content detection with Violet 406 for 30 minutes at

room temperature, protected from light. EdU detection is based on a reaction between the azide and the alkyne in the EdU. Acquisition analysis was performed on a LSRII cytometer.

2.3.12. Apoptosis analysis

The apoptosis staining was performed according to PE Annexin-V Apoptosis Detection Kit I manufacturer's protocol (BD Pharmingen). In summary, cells were washed twice with PBS and 100,000 cells were re-suspended in 100 µl of 1X Binding Buffer. Cells were stained with Annexin-V and 7-Amino-Actinomycin (7-AAD) antibodies for 15 min at room temperature in the dark. After adding 400ul of 1X Binding Buffer, cells were analysed by flow cytometry. Viable cells are 7-AAD$^-$ Annexin-V$^-$, early apoptotic cells are 7-AAD$^-$ Annexin-V$^+$ and apoptotic cells are 7-AAD$^+$ Annexin-V$^+$.

2.3.13. May-Grünwald-Giemsa staining

50,000-500,000 cells were centrifuged onto slides using a Cytospin 3 centrifuge (Shandon) for 5 minutes at 800 rpm at medium acceleration and left to dry. Slides were fixed with methanol for 10 minutes and stained with 5 volumes of O-Dianidisine, 1 volume of 3% Hydrogen Peroxide and 1 volume of Sodium Nitroferricyanide mix for 10 minutes protected from light. After rinsing with water, slides were stained with May-Grünwald solution (Merck) for 2-3 minutes, rinsed with water and stained with a 1 in 20 dilution of Giemsa's stain solution (VWR) for 20 minutes. Finally, slides were rinsed with water, left to dry and mounted with coverslip.

2.4. *In vivo* assays

2.4.1. Tet-on doxycycline-inducible expression system

The Doxycycline inducible *Sox7-2A-GFP* transgenic mice (*iSox7*) carry two transgenes allowing the inducible expression of *Sox7* as previously described (Gandillet et al., 2009). One transgene, inserted into the *Hprt* locus, encodes the *Sox7* sequence linked to GFP via a 2A viral peptide sequence under the control of a Tet responsible element. The second transgene inserted into the *Rosa* locus encodes a reverse tetracycline trans-activator. The stimulation with doxycycline initiates the transcription of both *Sox7* and GFP, allowing to monitoring the expression of *Sox7* through GFP expression. All animal work was performed under regulations governed by the Home Office Legislation under the Animal Scientific Procedures Act of 1986.

2.4.2. Cre recombinase transgenic mice

A previous member of the laboratory, Andrzej Mazan, generated the conditional lox P *Sox7*-targeted transgenic lines.

In *CD19-Cre* transgenic mice the deletion of *Sox7* gene is restricted to CD19 positive B cells. Blood, bone marrow and spleen were analysed by staining for all blood lineages. Unsorted and sorted $CD19^{+}$ cells from spleen and bone marrow were genotyped for *Sox7* deletion.

The conditional deletion of *Sox7* gene in *Mx1-Cre* transgenic mice was induced by Cre recombinase specifically in haematopoietic stem cells by intra-peritoneal injection of polyinosinic-polycytidylic acid (plpC). Mice were injected with 7 doses of plpC (600 μg per dose) every 2-3 days. Two, thirty and forty-four days after the final plpC dose, blood samples were analysed to assess the deletion level of the *Sox7* gene.

Once complete deletion was confirmed, blood, bone marrow and spleen cells were analysed by staining for all blood lineages.

2.4.3. Transplantation and tumourigenicity assay

One million bone marrow cells from *iSox7* mouse were re-suspended in PBS and transplanted intravenously into sub-lethally irradiated (125 cGy) NSG mice. After reconstitution of the haematopoietic system, these mice were fed with or without doxycycline diet pellet (Harlan). Every two months, one mouse per group (no Dox and Dox) was sacrificed and blood, bone marrow, spleen, liver and thymus were analysed.

For tumourigenicity assay, one or two million cells from *iSox7* colonies serially expanded in semi-solid culture were re-suspended in PBS and injected intravenously in NSG mice. Tumour growth was assessed every four weeks by blood analysis and every week by weight measuring and general health monitoring for signs of ill health (evidenced by piloerection, hunched posture, loss of response to environment). Mice were sacrificed around 10 months after the injections.

For transplantation of human leukaemia sample, RS4;11 and NALM6 cells or primary cells were transduced with lentiviruses as described above in section 3.10. A total of 150,000-2,000,000 transduced RS4;11 cells, 30,000-250,000 transduced NALM6 cells or 400-2000 transduced primary cells were sorted and transplanted intravenously into sub-lethally irradiated (125 cGy) NSG mice. Blood samples were analysed every 3 weeks and organs were analysed when required due to mouse sickness. Moreover, IVIS imaging was performed every 2-3 weeks in mice injected with transduced cells containing the Luciferase reporter construct as described in section 2.3.

All the procedures for injecting cells *in vivo* were performed by members of the

BRU facilities at the CRUK-Manchester Institute. All animal work was performed under regulations governed by the Home Office Legislation under the Animal Scientific Procedures Act of 1986 and was approved by the British Home Office.

2.4.4. *In vivo* imaging

Mice were injected intra-peritoneally with a Luciferin solution (dose of 75 mg/kg; PerkinElmer). Then, mice were placed into a clear Plexiglas anaesthesia box (3% isofluorane). After the mice were fully anesthetized, they were transferred to the nose cones in the imaging chamber (IVIS Lumina, Caliper Life Sciences). The imaging time is 30 seconds-1 minute per side (dorsal/ventral). The intra-peritoneal injection of Luciferin solution was performed by members of the BRU facilities at the CRUK-Manchester Institute.

2.5. Statistical analyses

Data have been statistically analysed using a Student's t Test. Significant differences are indicates with * for $p<0.05$ and with ** for $p<0.01$.

Chapter 3

3. Investigation of the consequences of *Sox7*-enforced expression on adult haematopoiesis during myeloid differentiation

3.1. Introduction

During embryonic development, the specification of the haematopoietic programme is orchestrated by transcription factors in response to signals from environmental factors. Many transcriptional regulators, growth factors and signalling cascades have been identified as part of these complex networks (Dzierzak and Speck, 2008). To fully understand these developmental processes, key elements within these regulatory networks still remain to be identified.

In order to identify novel regulators of blood specification, a gene expression profiling was performed on sub-populations isolated at various stages of the haematopoietic differentiation of mouse embryonic stem cells. This analysis led to the identification of several novel genes, among which the transcription factor *Sox7*, whose expression is modulated at the onset of blood specification.

Previous work performed in our laboratory established that *Sox7*-sustained expression in haematopoietic precursors blocked further differentiation while promoting self-renewal and that *Sox7* knock-down greatly impaired the generation of endothelial and haematopoietic precursors. Altogether, this work demonstrated the importance of

SOX7, along with its close homologue SOX18, in controlling the balance between proliferation and differentiation at the onset of haematopoietic specification (Gandillet et al., 2009; Serrano et al., 2010).

Given the profound impact of *Sox7*-sustained expression on embryonic blood differentiation, my project has investigated how the enforced expression of *Sox7* might affect the adult haematopoietic system. The first hypothesis I tested was whether *Sox7*-enforced expression conferred a proliferative advantage to adult haematopoietic cells. Secondly, I have defined which adult haematopoietic cells were responsive to *Sox7*-enforced expression and finally, I have characterized the immuno-phenotype of cells proliferating upon *Sox7*-enforced expression.

To this end, an elegant model previously established in the laboratory was used to elucidate the above hypothesis. This model system carries two transgenes allowing the inducible expression of *Sox7* as previously described by Gandillet and colleagues (Gandillet et al., 2009) (Figure 3.1).

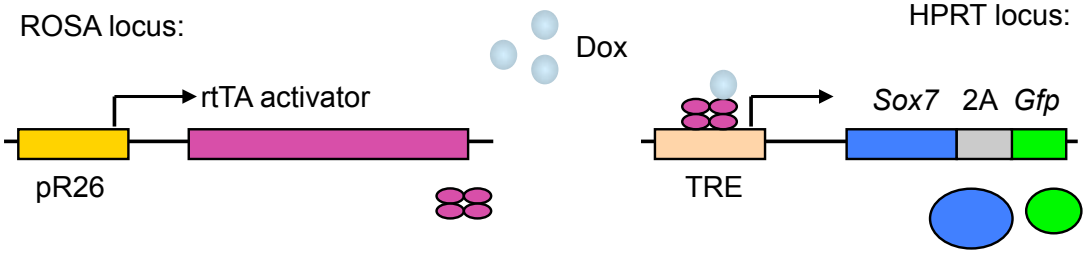


Figure 3.1. Doxycycline-inducible expression system.
Sox7 coding sequence, together with GFP sequence, is inserted into the HPRT locus. Adding doxycycline (Dox) to the cells results in the binding of the reverse tetracycline activator (rtTA) to the tetracycline responsive element (TRE) and drives expression of SOX7-2A-eGFP.

One transgene is inserted into the *Hprt* locus and encodes the *Sox7* coding sequence linked to a *Gfp-neomycin* cassette via a 2A viral peptide sequence under the control of a Tet responsible element. The 2A sequence mediates the co-translational cleavage of the poly-protein SOX7-2A-GFP (Ryan and Drew, 1994). The second transgene, inserted into the *Rosa* locus, encodes a reverse tetracycline trans-activator. The introduction of doxycycline in the culture conditions induces the transcription of *Sox7* and *Gfp* allowing monitoring the expression of SOX7 via GFP detection.

3.2. Results

3.2.1. *Sox7*-enforced expression induces the proliferation of bone marrow cells.

In order to investigate the consequences of *Sox7*-enforced expression on the homeostasis of adult haematopoiesis, the outcome of *Sox7* ectopic expression in cells derived from the doxycycline inducible *Sox7* transgenic mice (thereafter referred to as *iSox7*) has been analysed *in vitro*. The bone marrow was extracted from wild type (*wt*) and *iSox7* transgenic mice and serial clonogenic replating assays were performed in myeloid semi-solid culture in the presence or in the absence of doxycycline. Every three-four days, colonies were counted and the same number of cells was re-plated in fresh semi-solid medium to assess the clonogenic potential of those cells. Different assays were performed at each replating round (Figure 3.2 A). As shown in figure 3.2 B, the numbers of colonies steadily increased with each replating of *iSox7* bone marrow cells in the presence of doxycycline while a progressive decrease in colonies number was observed in the absence of doxycycline. In parallel, cells from *wt* mice were tested as control. In both conditions, a gradual decrease in colonies formation was observed

upon serial replating as reported in the *iSox7* sample cultured without doxycycline. Since no differences were identified between *wt* samples and *iSox7* cultured without doxycycline, only the latter was used as control in the subsequent experiments.

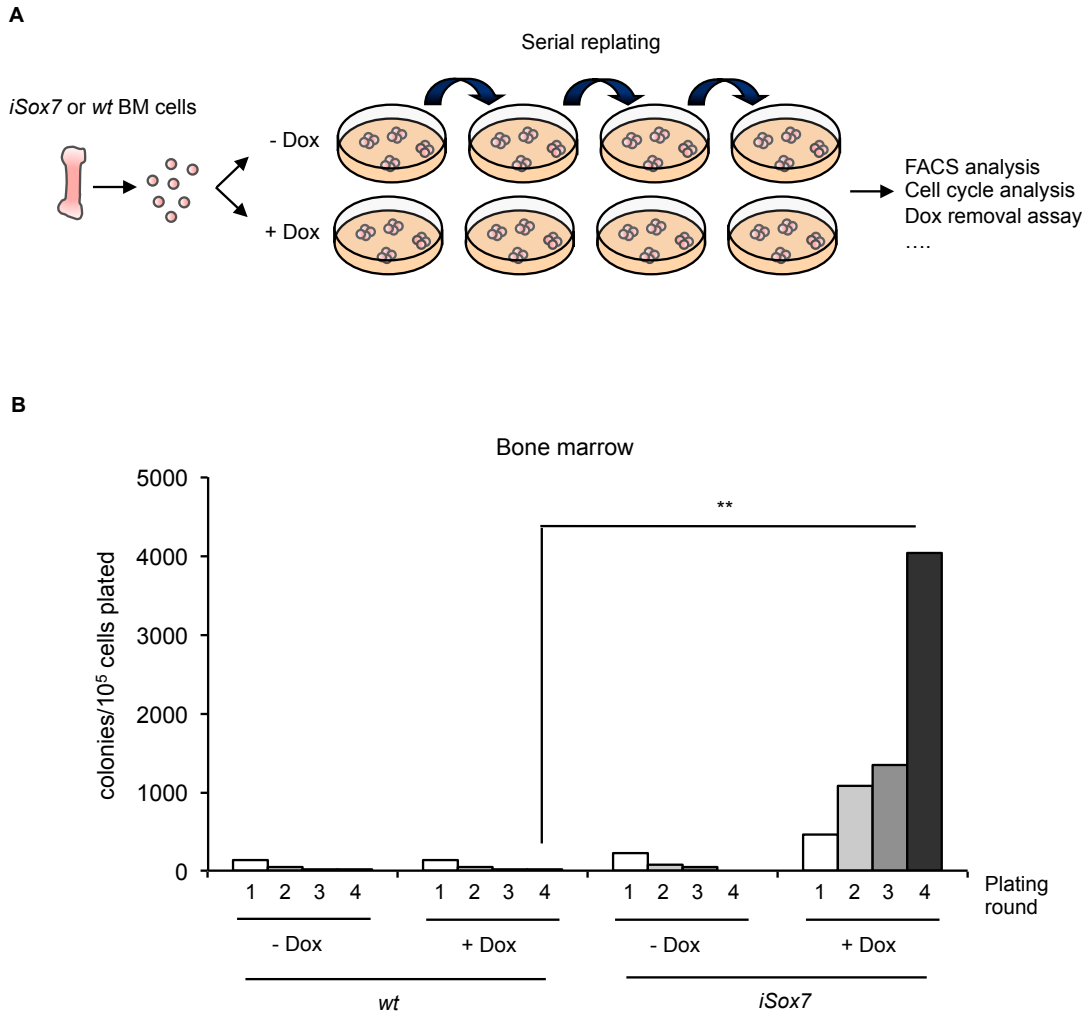


Figure 3.2. *Sox7*-enforced expression induces proliferation of bone marrow cells. A) Scheme of the protocol used for serial replating assay. The bone marrow was extracted from femurs of inducible *Sox7* transgenic or *wt* mice and serial replating assays were performed in semi-solid culture in the presence or in the absence of doxycycline. Different assays were performed at each replating. B) The *iSox7* and *wt* bone marrow cells were plated in semi-solid culture and colonies were counted every three-four days at each plating rounds. The bar chart shows the colony number for each plating rounds (n=3; $^{**}P<0.01$). Bar graph is representative of four independent experiments.

Large colonies, round and tight in appearance, were observed when *Sox7* expression was induced while only few small colonies were identified in the control cultures. In the presence of doxycycline, most colonies were GFP positive (Figure 3.3 A) and expressed increasing level of *Sox7* upon serial replating as indicated by the qRT-PCR analysis (Figure 3.3 B).

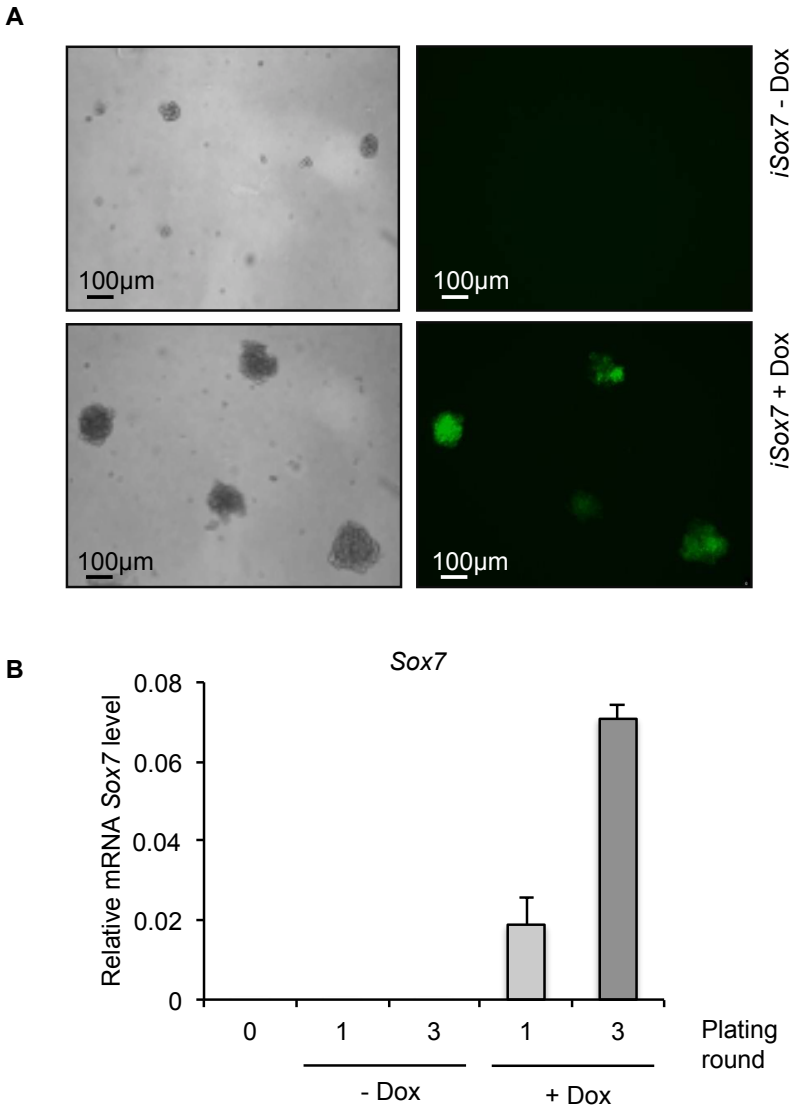


Figure 3.3. Colonies express increasing level of *Sox7* upon serial replating.
A) Colonies formed by *iSox7* bone marrow cells. Representative picture of *iSox7*$^{-}$ and *iSox7*$^{+}$ colonies obtained in clonogenic replating assay (Leica scan and software, scale bar: 100 µm). B) Quantitative RT-PCR analysis of *Sox7* transcript levels relative to βactin at 1st and 3rd plating round in *iSox7* bone marrow cells cultured with or without doxycycline. Error bars indicate mean ± standard deviation (n=3). Bar graph is representative of four independent experiments.

Next, analysis of the cell cycle was performed to further demonstrate the induction of proliferation by *Sox7*-enforced expression. The EdU incorporation together with the DNA content dye can easily and clearly separate cells into the different stages of the cell cycle. In particular, it is possible to distinguish cells in the G0-G1 state, in the DNA synthesis, S phase, and the G2 state (Schorl and Sedivy, 2007). In parallel to the increase in clonogenic potential, a 5-fold increase in the frequency of cells in the S-phase of the cell cycle analysis was also observed (Figure 3.4).

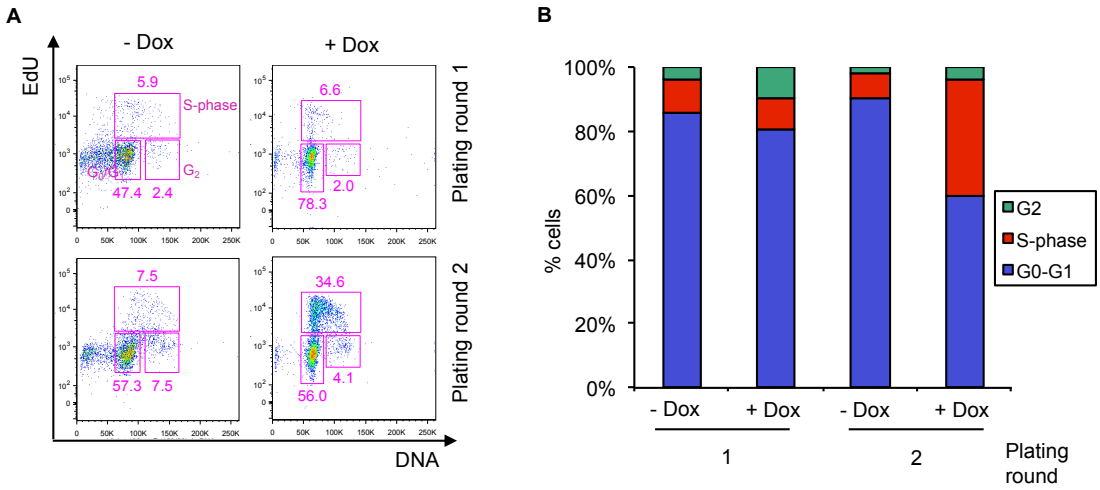


Figure 3.4. *Sox7*-enforced expression induces proliferation of bone marrow cells. A) The cell cycle status was assessed after 4 hours of EdU incubation of *iSox7* bone marrow cells cultured with or without doxycycline. Percentage of cells is shown in the indicated gates. FACS plots are representative of four independent experiments. B) The bar chart shows the percentage of cells in the G0/G1, S and G2 phases shown in A.

In addition, the morphology of cultured cells was analysed. Upon serial replating, the *iSox7*-induced colonies contained more immature blast-like cells and fewer macrophages compared to the control cultures as observed by May-Grünwald Giemsa staining (Figure 3.5).

Altogether, these data suggest that *Sox7* over-expression, to some extent, impaired haematopoietic differentiation and induces proliferation, a findings similar to what was observed in embryonic haematopoiesis (Gandillet et al., 2009).

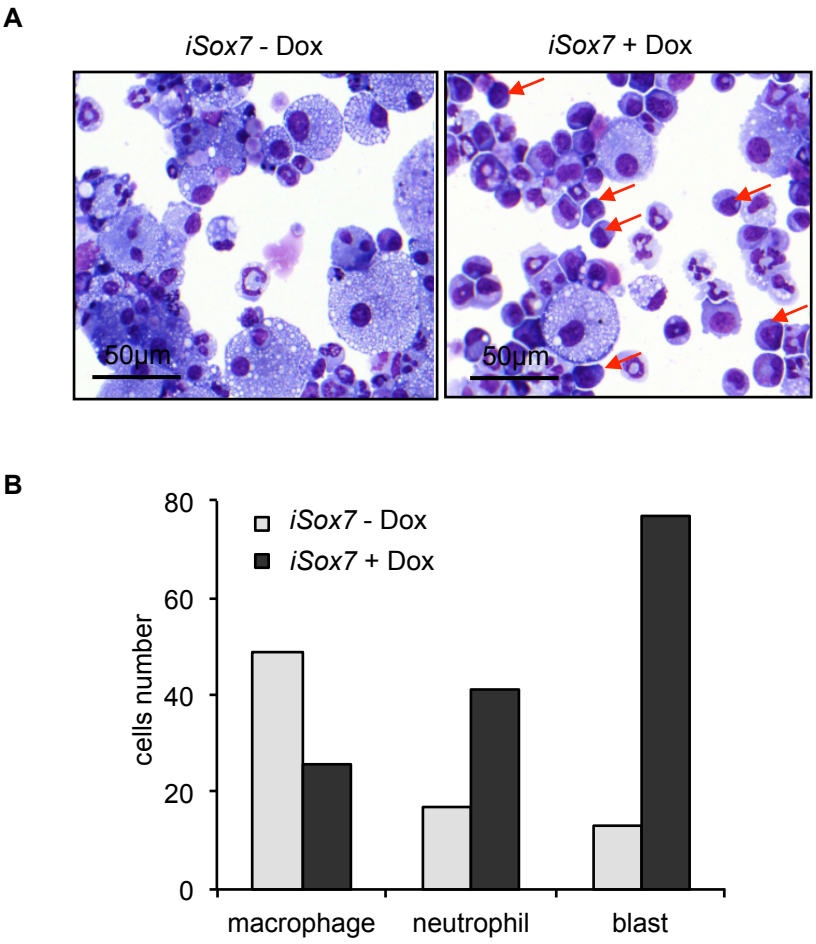


Figure 3.5. *Sox7*-enforced expression impairs haematopoietic differentiation of bone marrow cells.
A) May-Grünwald-Giemsa staining of *iSox7* bone marrow cells cultured with or without doxycycline at the plating round 2 (Leica scan and software, scale bar: 50 µm). Red arrows indicate immature blast-like cells. Pictures are representative of four independent experiments. B) The bar chart shows the number of macrophages, neutrophils and blast cells shown in A.

3.2.2. Characterization of *Sox7*-expressing cells

Next, the immuno-phenotype of *Sox7*-expressing cells was investigated using FACS analysis. During the serial replating, the increase in proliferation was paralleled with an increase in the frequency of SOX7::GFP-expressing cells as shown in figure 3.6.

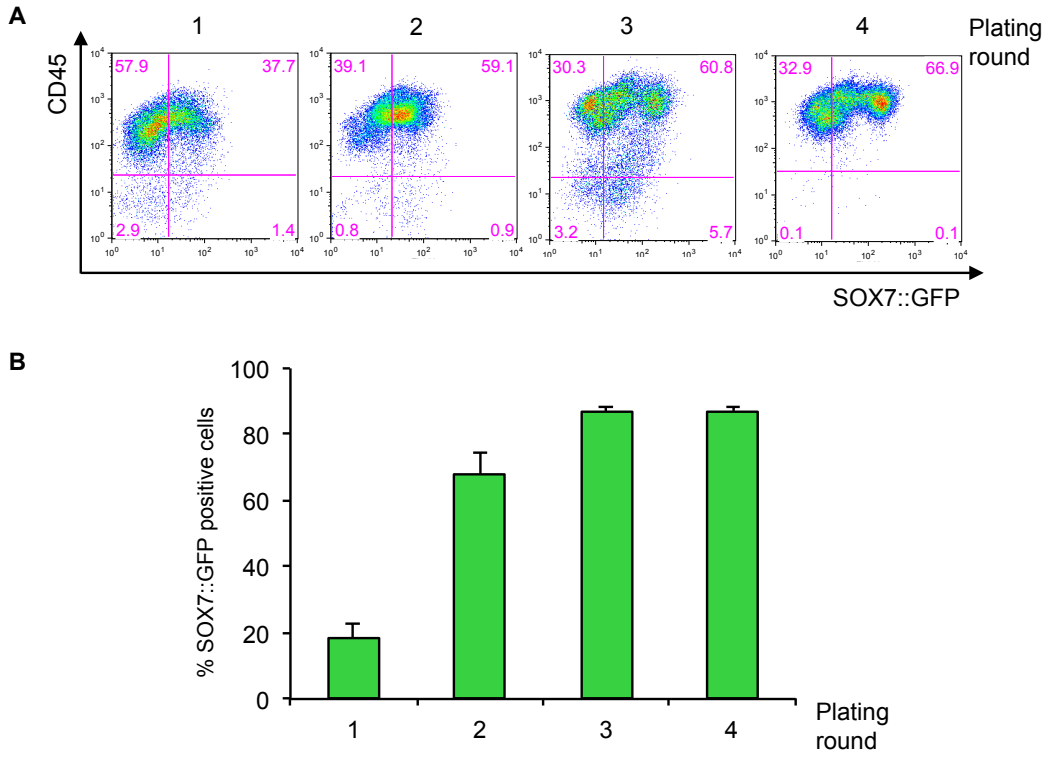


Figure 3.6. The increase in proliferation is reflected by an increase in the frequency of SOX7::GFP-expressing cells.
A) Representative FACS plot of SOX7::GFP and CD45 expression profiles of *iSox7* bone marrow cells cultured with doxycycline. Percentages are shown in the indicated quadrants. B) Percentage of SOX7::GFP positive cells are shown in the bar chart. Error bars indicate mean ± standard deviation (n=3). FACS plots and graph are representative of four independent experiments.

FACS analysis revealed two sub-populations based on their level of SOX7::GFP expression: SOX7::GFP^{low} and SOX7::GFP^{high}. Different markers were included in the analysis; in particular, MAC1 and GR1 were used for identifying myeloid cells, CD45 for haematopoietic cells and cKIT for progenitor cells. Moreover, the endothelial cell marker, VE-cadherin (CDH5), was included in the analysis since it was previously

shown to be a transcriptional target of SOX7 at the onset of haematopoietic development (Costa et al., 2012). The immuno-phenotypic characterisation of the two SOX7::GFP sub-populations revealed that most SOX7::GFP^{low} cells were MAC1^{high}, CD45^{high}, cKIT$^{-/low}$, CD34^{-}, GR1$^{low/high}$, VE-cadherin$^{-/low}$ and CD48^{low} while the SOX7::GFP^{high} cells were MAC1^{high}, MAC1^{low}, CD45^{high}, cKIT$^{+/-}$, CD34^{-}, GR1^{-}, VE-cadherin^{high} and CD48^{high} (Figure 3.7).

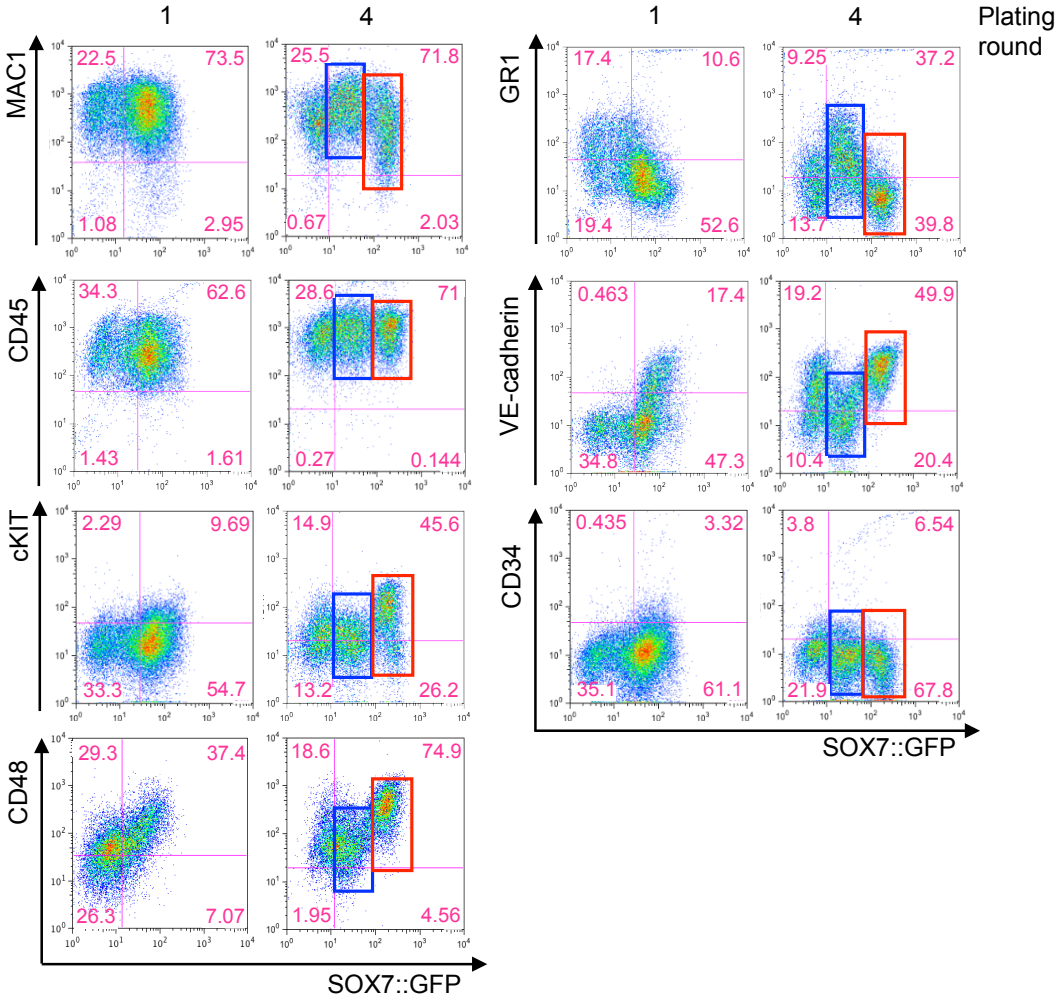


Figure 3.7. Characterization of *Sox7*-expressing cells.
Representative FACS plot of SOX7::GFP, MAC1, GR1, CD45, VE-cadherin, cKIT, CD34 and CD48 expression profiles of *iSox7* bone marrow cells cultured with doxycycline at the plating round 1 and 4. Percentages are shown in the indicated gate. SOX7::GFP^{low}-expressing cells are gated in blue while SOX7::GFP^{high}-expressing cells are in red. FACS plots are representative of four independent experiments.

In this analysis, we observed a progressive increase in the frequency of cells double positive for VE-cadherin and SOX7::GFP expression (Figure 3.8 A). Moreover, further analysis of the immuno-phenotype revealed an increase in the frequency of the SOX7::GFP^{+}cKIT^{+} population upon serial replatings suggesting that a SOX7::GFP^{+}cKIT^{+} blast-like population may maintain the clonogenic potential of *iSox7* bone marrow cells (Figure 3.8 B). Given the presence of this blast-like population, a detailed analysis of the expression of haematopoietic stem cell markers was performed. This analysis revealed an increase in the frequency of Lineage^{-} SCA1^{+} cKIT^{+} (LSK) fraction upon serial replating as well as an increase in the LSK CD48^{+} population (Figure 3.9) suggesting that *Sox7*-enforced expression may induce the proliferation of a progenitor subset.

Altogether these findings reveal that the enforced expression of *Sox7* in bone marrow cells not only enhances proliferation but also blocks lineage maturation.

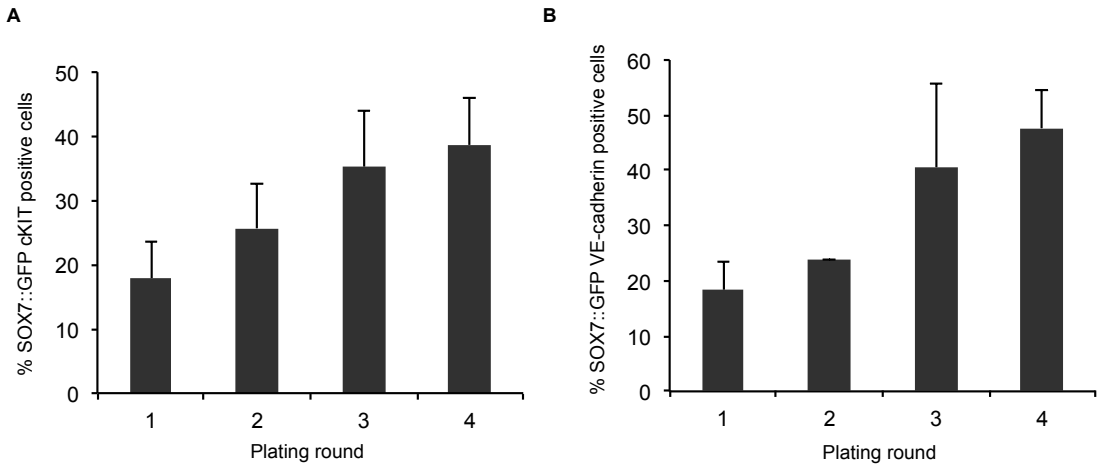


Figure 3.8. *Sox7*-enforced expression induces the expression of cKIT and VE-cadherin positive cells.
The percentage of SOX7::GFP-cKIT double positive (A) and SOX7::GFP-VE-cadherin double positive (B) cells is shown in the bar chart for *iSox7* positive colonies during the four plating rounds. Error bars indicate mean ± standard deviation (n=3).

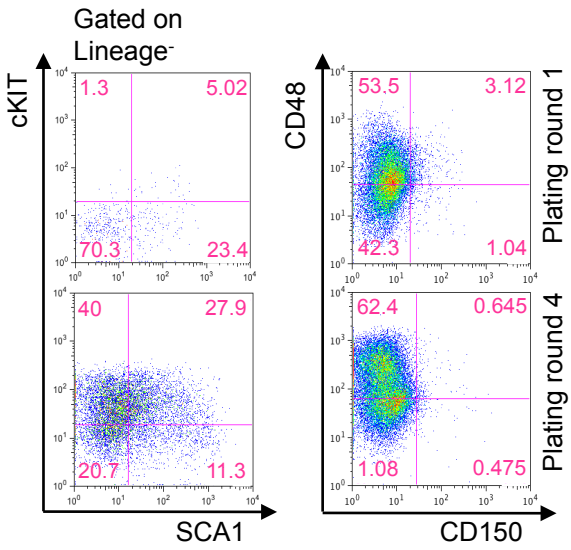


Figure 3.9. *Sox7*-enforced expression induces an increase in the frequency of LSK cells.
Representative FACS plot of cKIT, SCA1, CD150 and CD48 expression profiles of *iSox7* bone marrow cells cultured with doxycycline at the plating round 1 and 4. Percentages are shown in the indicated quadrants. SCA1 cKIT plot is gated on Lineage negative population. FACS plots are representative of four independent experiments.

3.2.3. The cKIT^{+}GFP^{high} sub-population contains the clonogenic potential and requires the sustained expression of *Sox7* to proliferate.

To define whether *Sox7* expression was essential for the maintenance of the clonogenic potential, doxycycline was removed from the culture and the colonies were counted after one week (Figure 3.10 A). As shown in the figure 3.10 B and C, *Sox7* expression was essential to maintain the formation of colonies since removal of doxycycline dramatically and significantly reduced both clonogenic and proliferative potential.

In order to address whether SOX7::GFP-expressing cells were able to maintain the clonogenic potential, SOX7::GFP^{+} and SOX7::GFP^{-} sub-populations were sorted from colonies obtained after 3 successive passages and replated in the presence or in the

absence of doxycycline (Figure 3.11 A). After one week in culture, only the SOX7::GFP^{+} cells replated in the presence of doxycycline were able to generate colonies and to maintain a high proliferative potential (Figure 3.11 B-C). Altogether, these data suggests that SOX7::GFP-expressing cells maintain the clonogenic potential upon *Sox7*-enforced expression.

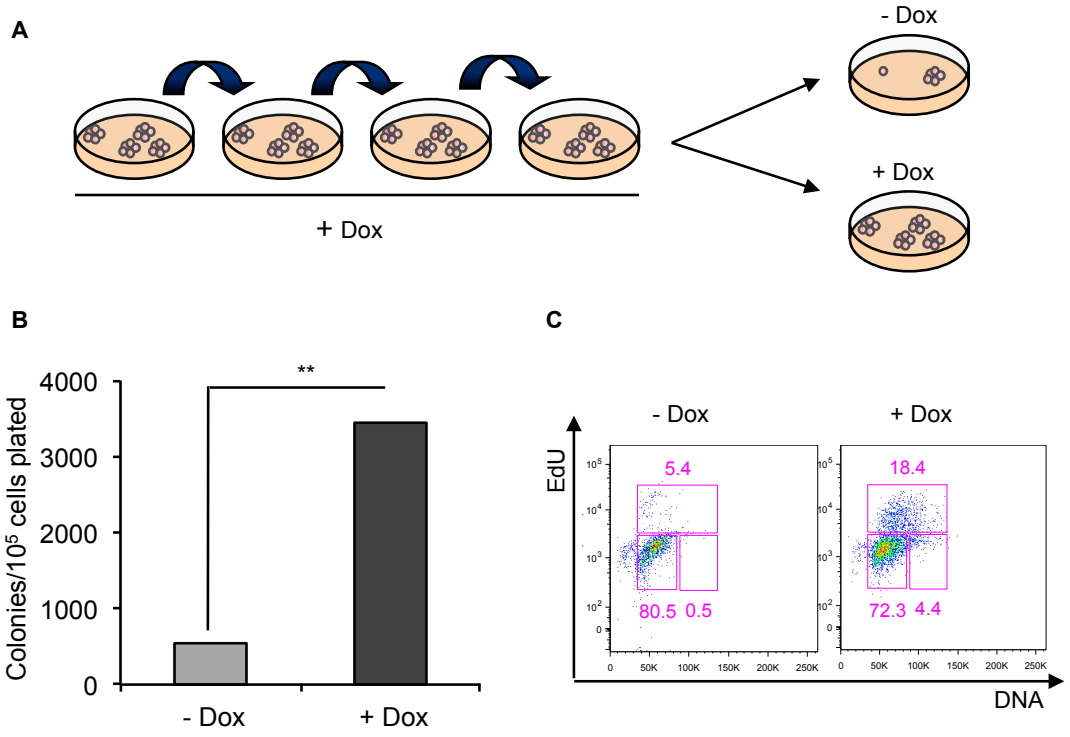


Figure 3.10. *Sox7* expression is essential to maintain the formation of colonies.
A) Scheme of the protocol used for doxycycline removal assay. Serial replating assays were performed in semi-solid culture in the presence of doxycycline. After the plating round 4, 100,000 *iSox7*$^{+}$ bone marrow cells were cultured in semi-solid culture with or without doxycycline. After one week in culture, cells were analysed. B) The bar chart shows the colony number for each sample (n=3; $^{**}P<0.01$). C) The cell cycle status was assessed after 4 hours of EdU incubation of *iSox7*$^{+}$ bone marrow cells cultured with or without doxycycline. Percentages of cells are shown in the indicated gates. Graph and FACS plots are representative of two independent experiments.

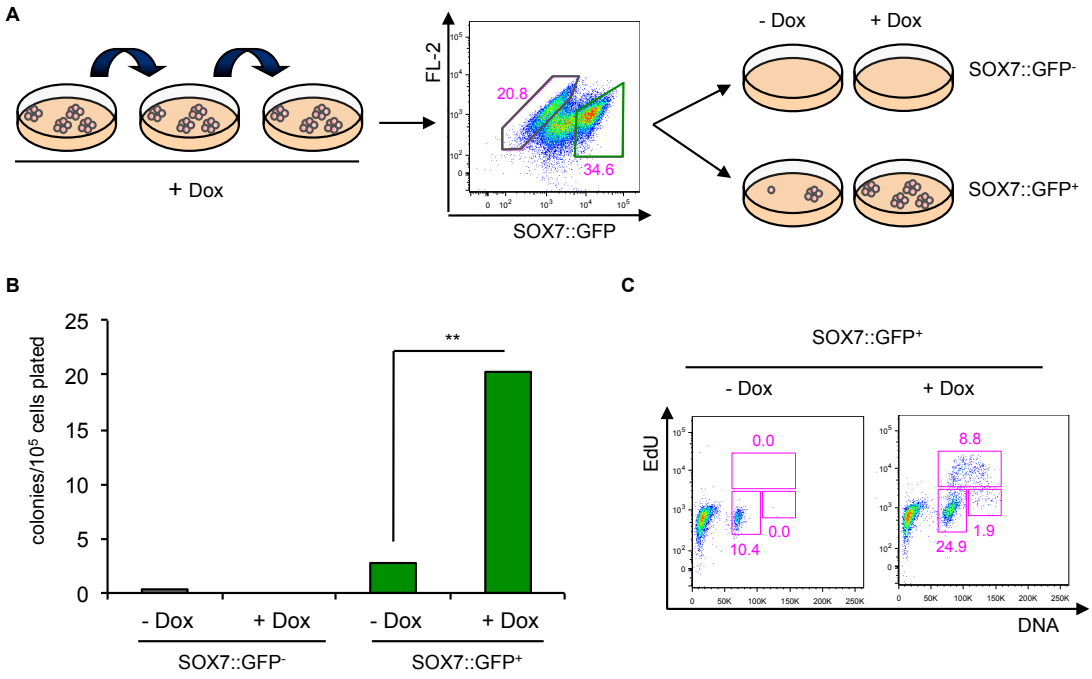


Figure 3.11. SOX7::GFP-expressing cells are able to maintain the clonogenic potential.
A) Scheme of the protocol used to address whether SOX7::GFP-expressing cells are able to maintain the clonogenic potential. Serial replating assays were performed in semi-solid culture in the presence of doxycycline. After the plating round 3, *iSox7*$^{+}$ bone marrow cells were sorted based on SOX7::GFP expression and 100,000 sorted cells were cultured in semi-solid culture with or without doxycycline. After one week in culture, cells were analysed. B) The bar chart shows the colony number for each sample (n=3; ***P*<0.01). C) The cell cycle status was assessed after 4 hours of EdU incubation of SOX7::GFP^{-} *iSox7*$^{+}$ bone marrow cells cultured with or without doxycycline. Percentages of cells are shown in the indicated gates. Graph and FACS plots are representative of two independent experiments.

A detailed immuno-phenotypic analysis performed in these experiments revealed the presence of three different sub-populations within the *iSox7*$^{+}$ replating cultures based on SOX::GFP and cKIT staining: SOX7::GFP^{low}cKIT^{-}, SOX7::GFP^{high}cKIT^{-} and SOX7::GFP^{high}cKIT^{+}. To define their nature and clonogenic potential, these three sub-populations were sorted and replated in semi-solid culture with or without doxycycline (Figure 3.12 A). When compared to the other sub-populations, the SOX7::GFP^{high}cKIT^{+} sub-population gave rise to more colonies after one week in culture with doxycycline (Figure 3.12 B) and showed a high proliferative potential as determined by EdU incorporation (Figure 3.12 C). In addition, analysis of the cell morphology revealed that the SOX7::GFP^{high}cKIT^{+} sub-population in the presence of doxycycline contained fewer macrophages and monocytes but more small immature blast-like cells (Figure 3.12 D).

Altogether, these data revealed that the SOX7::GFP^{high} cells expressing cKIT seem to best maintain the capacity to produce colonies and appears to be the most responsive to enforced *Sox7* expression.

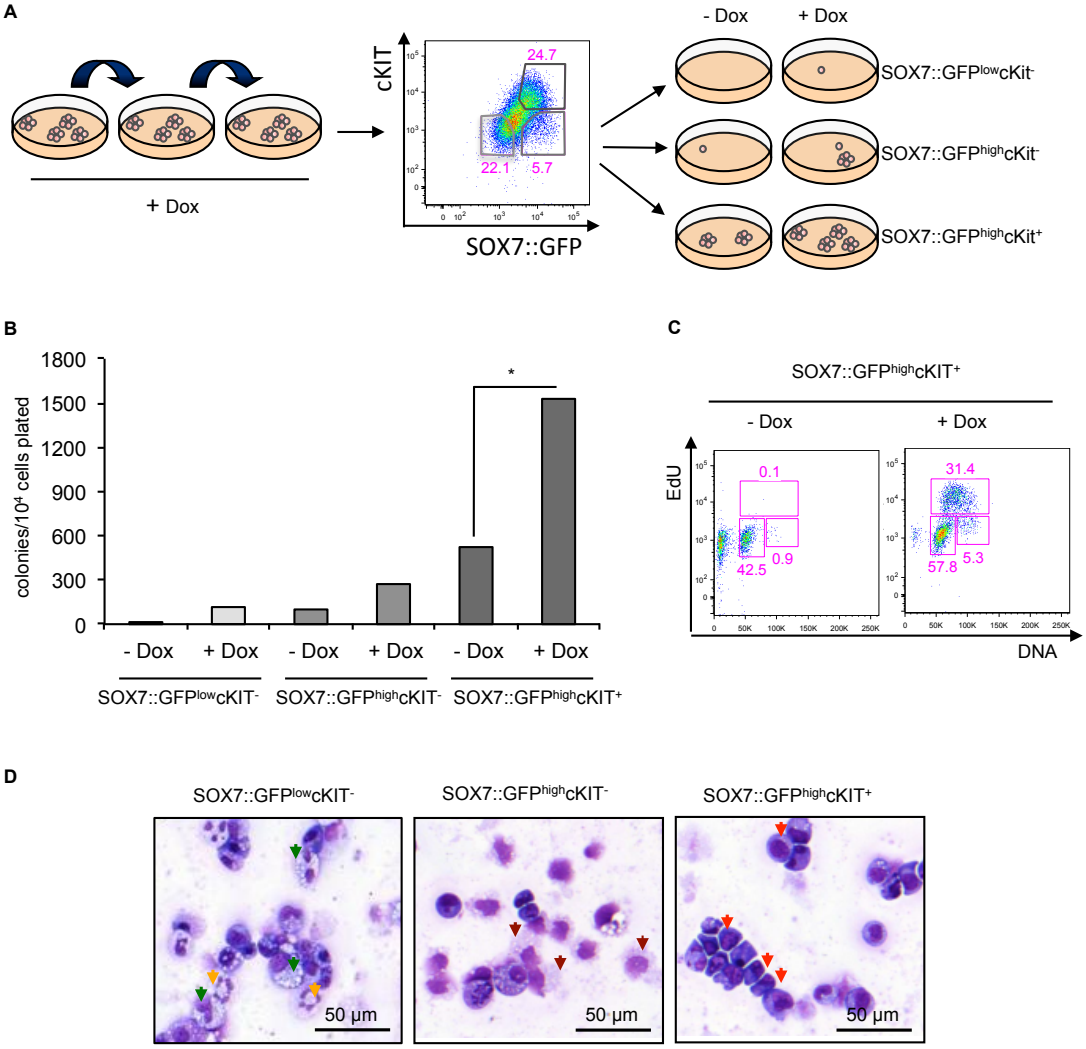


Figure 3.12. SOX7::GFP^{high}cKIT^{+}-expressing cells are able to maintain the clonogenic potential and show less sign of haematopoietic differentiation.
Scheme of the protocol used to address whether SOX7::GFP-expressing cells are able to maintain the clonogenic potential. Serial replating assays were performed in semi-solid culture in the presence of doxycycline. After the plating round 3, *iSox7*$^{+}$ bone marrow cells were sorted based on SOX7::GFP and cKIT expression and 100,000 sorted cells were cultured in semi-solid culture with or without doxycycline. After one week in culture, cells were analysed. B) The bar chart shows the colony number for each sample (n=3; *P<0.05). C) The cell cycle status was assessed after 4 hours of EdU incubation of SOX7::GFP^{high}cKIT^{+} bone marrow cells cultured with or without doxycycline. Percentages of cells are shown in the indicated gates. Graph and FACS plots are representative of three independent experiments. D) May-Grünwald-Giemsa staining of *iSox7* bone marrow cells cultured with doxycycline after one week (Leica scan and software, scale bar: 50 µm). Yellow arrows indicate neutrophils, green arrows indicate macrophages, dark red arrows indicate monocytes and red arrows indicate immature blast-like cells.

3.2.4. Differential responsiveness of splenic and thymic cells to *Sox7*-enforced expression.

To define which cell types could be targeted by *Sox7*-enforced expression the above experiments were repeated starting from spleen and thymus cells. The spleen and thymus were extracted from *iSox7* transgenic and *wt* mice. During the following weeks serial replatings of *iSox7* and *wt* cells were performed in semi-solid culture in the absent or presence of doxycycline and colonies were counted and analysed (Figure 3.13 A). As presented in figure 3.13 B, the number of colony significantly increased, during the successive serial replatings, only in the *iSox7* spleen cells in the presence of doxycycline. On the contrary, thymus cells were not able to response to *Sox7*-enforced expression.

To define whether *Sox7* was essential for the maintenance of the clonogenic potential of spleen cells, doxycycline was removed from the culture and the colonies were counted after one week. As shown in the figure 3.13 C and D, *Sox7* expression was essential to maintain the formation of colonies since removal of doxycycline dramatically and significantly reduced the clonogenic and proliferative potential.

In addition, analysis of the immuno-phenotype of SOX7::GFP^{+} cells revealed that most cells within the colonies were MAC1^{+}, GR1^{+} and cKIT^{low} (Figure 3.13 E) as observed in the bone marrow cell-derived cultures.

Hence, these data revealed that, in addition to bone marrow cells, a subset of cells within the spleen is also responsive to *Sox7*-enforced expression.

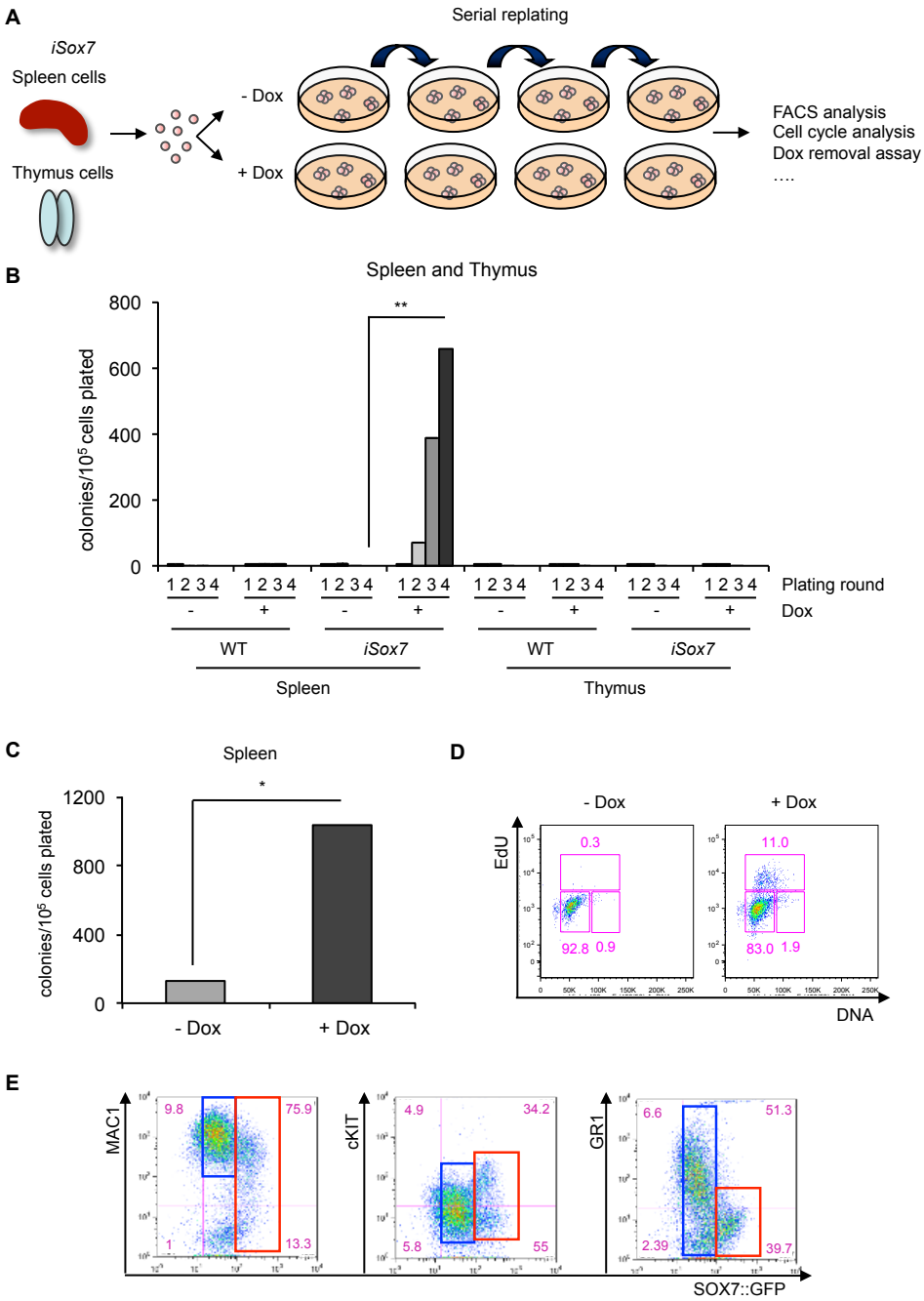


Figure 3.13. A subset of cells within the spleen is responsive to enforced *Sox7* expression.

A) Spleen and thymus cells were extracted from inducible *Sox7* transgenic or *wt* mice and serial replating assays were performed in semi-solid culture in the presence or in the absence of doxycycline. B) The bar chart shows the colony number for each plating rounds (n=3; $**P<0.01$). C) After the plating round 4, *iSox7*$^{+}$ spleen cells were cultured in semi-solid culture with or without doxycycline for one week. The bar chart shows the colony number for each sample (n=3; $*P<0.05$). D) The cell cycle status was assessed after 4 hours of EdU incubation of *iSox7*$^{+}$ spleen cells cultured with or without doxycycline. E) Representative FACS plot of SOX7::GFP, MAC1, GR1 and cKIT expression profiles of *iSox7* spleen cells cultured with doxycycline at the plating round 4. SOX7::GFP^{low}-expressing cells are gated in blue while SOX7::GFP^{high}-expressing cells are in red. Percentages are shown in the indicated gates. Graphs and FACS plots are representative of three independent experiments.

3.2.5. *Sox7*-enforced expression induces the proliferation of progenitor cells derived from adult bone marrow.

To extend those findings and further define which subset of cells within the bone marrow were responsive to *Sox7*-enforced expression, specific subset of haematopoietic progenitors (HSC, MPP, LSK $CD48^+$, LMPP) were isolated from *iSox7* bone marrow and tested in serial clonogenic replating assays (Figure 3.14 A).

A preliminary experiment was performed sorting cells only based on cKIT expression (Figure 3.14 B). All haematopoietic progenitor cells are included in the cKIT positive sub-population while the cKIT negative sub-population contains more mature and differentiated haematopoietic cells. As expected, colonies were obtained only from the cKIT positive sub-population. The induction of *Sox7* expression seemed not to increase the formation of colony in this particular condition (Figure 3.14 C). Next, a more detailed analysis was carried out to determine which subsets of progenitors within the bone marrow were responsive to *Sox7*-enforced expression. Based on specific immuno-phenotypes, subsets of haematopoietic progenitors were sorted using well-known markers and plated in the semi-solid culture. The HSC, MPP and LSK $CD48^+$ populations ware isolated gating first on $lineage^-$ $SCA1^+$ and $cKIT^+$ cells; then, CD48 and CD150 were used to further define the three sub-populations. HSCs are $CD150^+$ $CD48^-$, MPPs are $CD150^-$ $CD48^-$ while LSK $CD48^+$ cells are positive for both CD150 and CD48 (Figure 3.15 A).

As shown in figure 3.15 B, all sorted sub-populations of progenitors were able to generate colonies upon *Sox7*-enforced expression in serial replating. Colonies were maintained upon serial replating when *Sox7* expression was induced while clonogenic potential was lost in the control cultures.

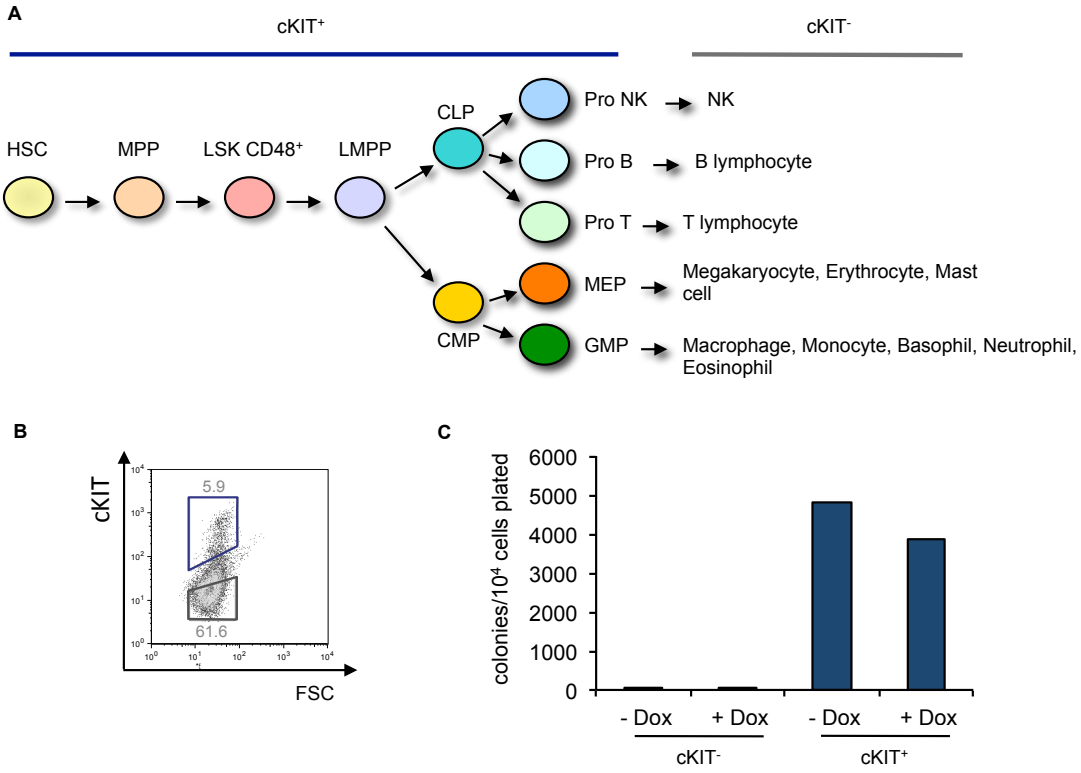


Figure 3.14. *Sox7*-enforced expression induces the proliferation of cKIT$^+$ cells.
A) Schematic representation of haematopoietic hierarchy. B) cKIT$^+$ and cKIT$^-$ cells were sorted from *iSox7* bone marrow cells. C) cKIT$^+$ and cKIT$^-$ cells were cultured in semi-solid culture with and without doxycycline for clonogenic replating assays. Bar chart shows the colony number for each sample (n=3; ** P< 0.01). Graph and FACS plot are representative of two independent experiments. HSC: haematopoietic stem cells, MPP: multi-potent progenitors, LSK: Lineage$^-$ SCA1$^+$ cKIT$^+$, LMPP: lympho-primed multi-potent progenitors, CLP: common lymphoid progenitors, CMP: common myeloid progenitors.

The more differentiated progenitors LMPP were sorted using the expression of the FLT3 receptor (Figure 3.15 C). Similarly, LMPP cells were able to generate colonies only in the presence of doxycycline while no colonies were observed in the control. Moreover, a gradual increase in colony number was observed upon *Sox7*-enforced expression (Figure 3.15 D).

These data reveal that *Sox7*-enforced expression is able to induce the proliferation of progenitors at several levels of the blood lineage hierarchy.

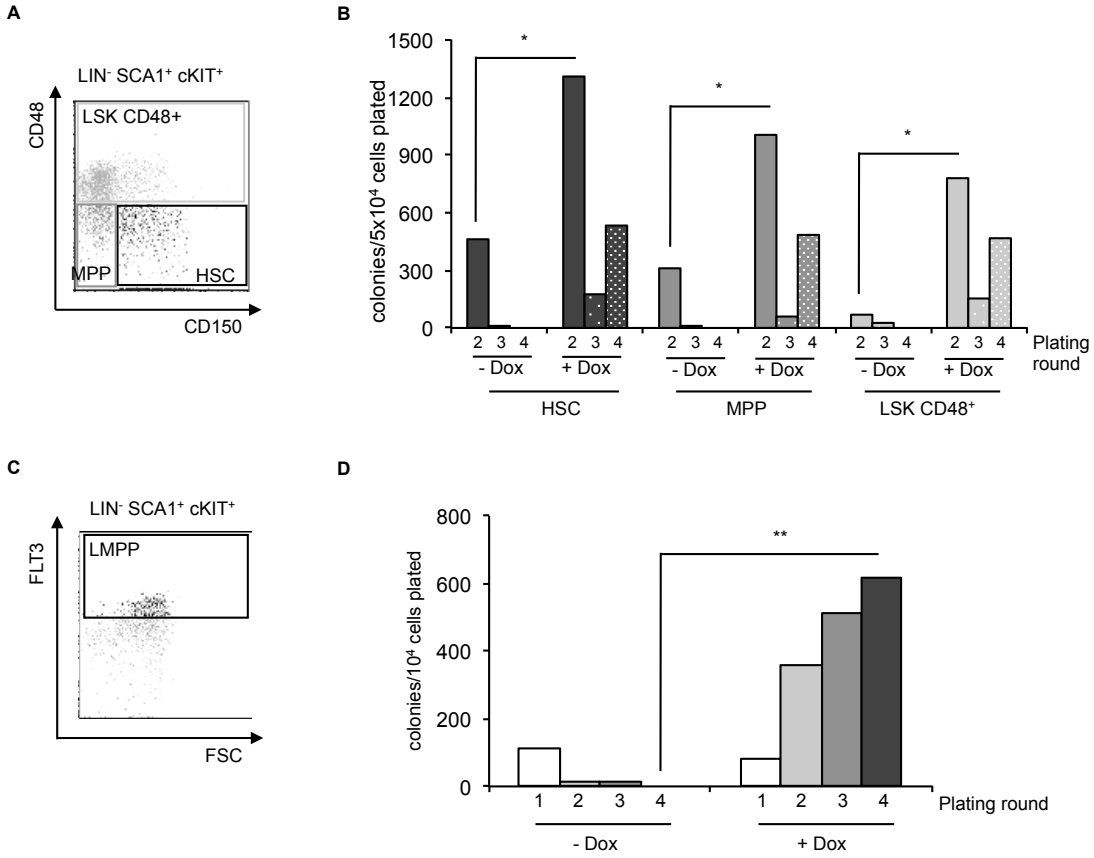


Figure 3.15. *Sox7*-enforced expression induces the proliferation of a subtype of cells with blast-progenitor characteristic at different level of blood lineage specification.

A) HSC, MPP and LSK CD48$^+$ cells were sorted from *iSox7* bone marrow cells gated on LIN$^-$SCA1$^+$cKIT$^+$ population. B) HSC, MPP and LSK CD48$^+$ cells were cultured in semi-solid culture with and without doxycycline for clonogenic replating assays. C) LMPP cells were sorted from *iSox7* bone marrow cells gated on LIN$^-$SCA1$^+$cKIT$^+$ population. D) LMPP cells were cultured in semi-solid culture with and without doxycycline for clonogenic replating assay. Bar charts show the colony number for each plating rounds (n=3; ** P< 0.01). Bar graphs are representative of one independent experiment.

3.2.6. Impact of *Sox7*-enforced expression on the haematopoietic differentiation of *p53*$^{-/-}$ bone marrow cells

Sox7-enforced expression might not be sufficient for the development of overt haematological malignancy. If this hypothesis is true, the cells responsive to enforced *Sox7* could represent a pre-leukaemic population and therefore require a second hit to induce overt malignancy. As a model system, the deletion of *p53* was used since it has been reported that in cases of acute myeloid leukaemia the loss of the *p53* response pathway is associated with disease progression (Krejci et al., 2008). Bone marrow was extracted from *p53*$^{-/-}$*iSox7* transgenic mice. Serial clonogenic replating assays of *p53*$^{-/-}$ *iSox7* cells were performed in semi-solid culture in the absent or presence of doxycycline and the colonies were counted and analysed by flow cytometry.

As shown in the graph presented in figure 3.16 A, the colony numbers increased more significantly during the successive replatings of *p53*$^{-/-}$*iSox7* bone marrow-derived cells in the presence of doxycycline. Colonies were tight and round in appearance and were expressing GFP when *Sox7* was induced (Figure 3.16 B). Level of SOX7::GFP expression gradually increased upon *Sox7*-enforced expression (Figure 3.16 C). In addition, analysis of the cell morphology in *p53*$^{-/-}$*iSox7* culture with doxycycline showed an increased frequency of immature blast-like cells and a decrease in macrophages compared to the control population (Figure 3.16 D-E).

To further define whether *Sox7* is essential for the maintenance of the clonogenic potential, doxycycline was removed from the culture and the colonies were counted after one week. As shown in the figure 3.17 A and B, *Sox7* expression was essential to maintain the formation of colonies since removal of doxycycline dramatically reduced the clonogenic and proliferative potential.

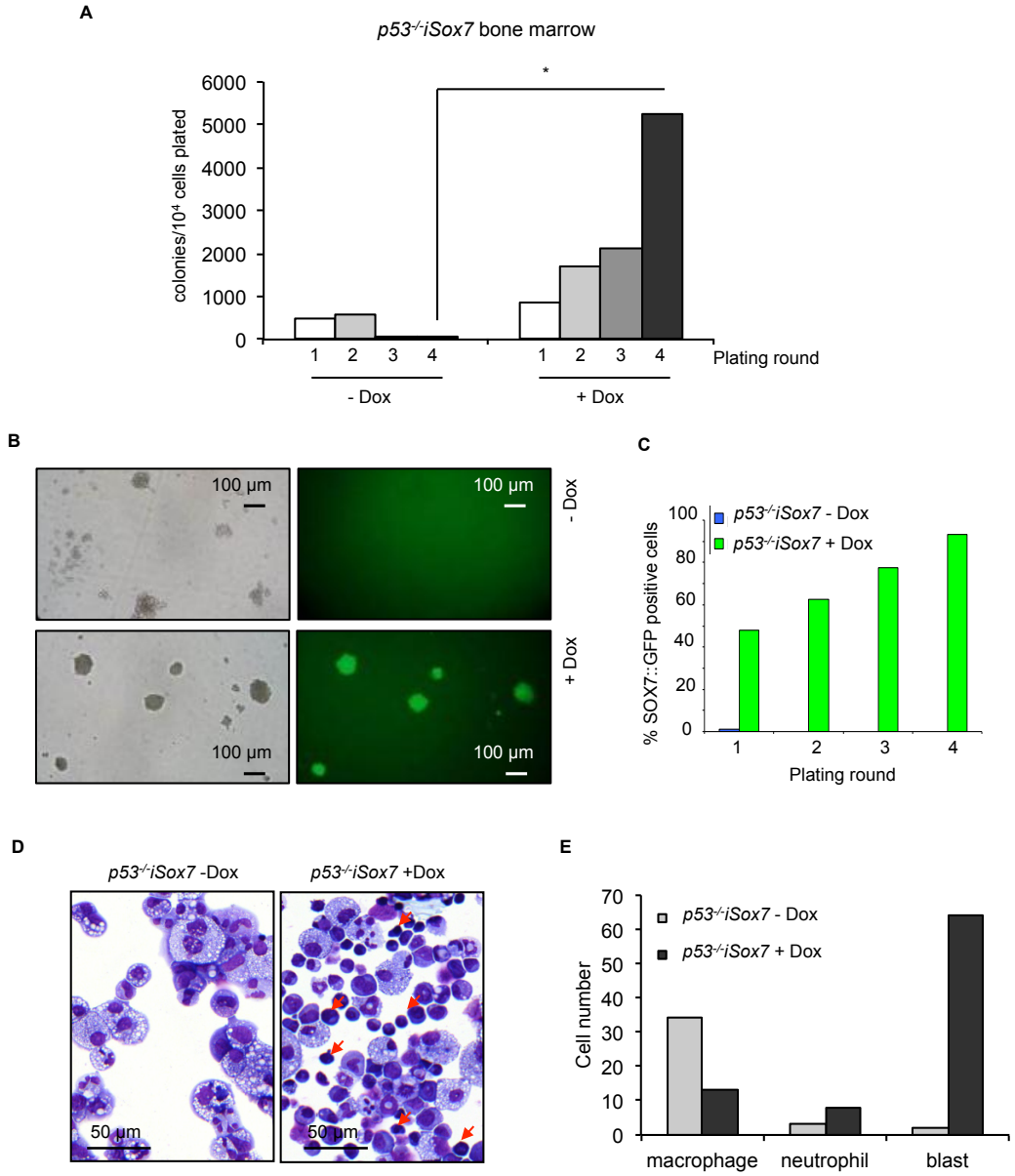


Figure 3.16. *Sox7*-enforced expression induces proliferation of *p53*$^{-/-}$*iSox7* bone marrow cells.
A) The *p53*$^{-/-}$*iSox7* bone marrow cells were plated in semi-solid culture and colonies were counted every three-four days at each plating rounds. The bar chart shows the colony number for each plating rounds (n=3; **P*<0.05). B) Representative picture of *p53*$^{-/-}$*iSox7*$^{-}$ and *p53*$^{-/-}$*iSox7*$^{+}$ colonies obtained in clonogenic replating assay (Leica scan and software, scale bar: 100 µm). C) Percentage of SOX7::GFP positive cells are shown in the bar chart. D) May-Grünwald-Giemsa staining of *p53*$^{-/-}$*iSox7* bone marrow cells cultured with or without doxycycline at the plating round 2 (Leica scan and software, scale bar: 50 µm). Red arrows indicate immature blast-like cells. E) The bar chart shows the number of macrophages, neutrophils and blast cells shown in D. Graphs are representative of three independent experiments.

In order to confirm that SOX7::GFP-expressing cells are able to maintain the clonogenic potential, SOX7::GFP$^+$ and SOX7::GFP$^-$ sub-populations were sorted from colonies obtained after 3 successive replatings and replated in the presence or in the absence of doxycycline as performed in the *p53 wt* background. After one week in culture, SOX7::GFP$^+$ cells replated in the presence of doxycycline were able to generate colonies (Figure 3.17 C-D) at a 3-fold increased frequency that in the control.

Altogether, these data demonstrate that SOX7::GFP-expressing cells maintain the clonogenic potential upon *Sox7*-enforced expression.

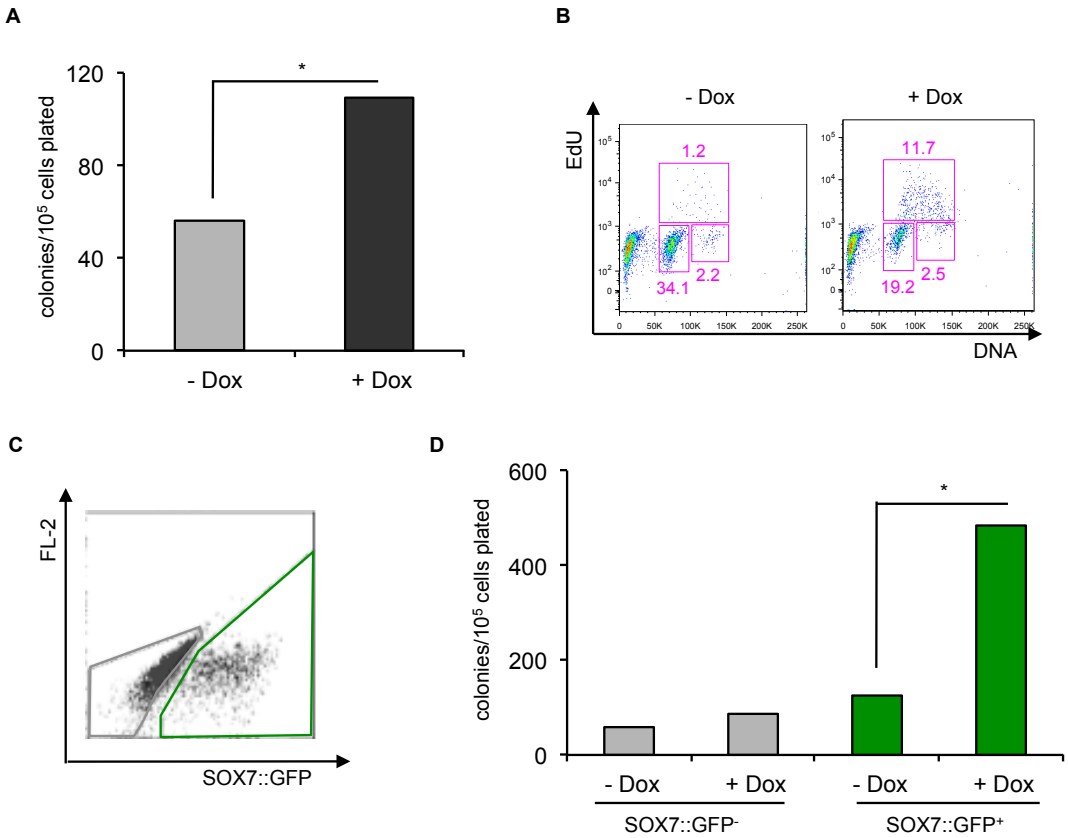


Figure 3.17. SOX7::GFP-expressing cells are able to maintain the clonogenic potential.
A) After the plating round 4, *p53*$^{-/-}$*iSox7*$^+$ bone marrow cells were cultured in semi-solid culture with or without doxycycline for one week. The bar chart shows the colony number for each sample (n=3; **P*<0.05). B) The cell cycle status was assessed after 4 hours of EdU incubation of *p53*$^{-/-}$*iSox7*$^+$ bone marrow cells cultured with or without doxycycline. Percentages of cells are shown in the indicated gates. C) After the plating round 3, *p53*$^{-/-}$*iSox7*$^+$ bone marrow cells were sorted based on SOX7::GFP expression and cultured in semi-solid culture with or without doxycycline. D) The bar chart shows the colony number for each sample after a week in culture (n=3; **P*<0.05). Graphs and FACS plots are representative of one independent experiment.

Then, the immuno-phenotype of those cells was investigated by FACS analysis. The analysis of the *p53*$^{-/-}$SOX7::GFP^{+} cells showed an immuno-phenotype similar to the one observed previously for *p53*$^{+/+}$SOX7::GFP^{+} cells with however a marked increase in the frequency of SOX7::GFP^{high} cells. In summary, *p53*$^{-/-}$SOX7::GFP^{low} cells are MAC1^{high}, cKIT$^{-/low}$ and VE-cadherin$^{-/low}$ while *p53*$^{-/-}$SOX7::GFP^{high} cells are MAC1$^{+/low}$, cKIT$^{-/+}$ and VE-cadherin^{+} (Figure 3.18). This flow cytometric analysis also revealed the presence of three sub-populations of cell based on SOX7::GFP and cKIT expression in the context of the *p53*$^{-/-}$*iSox7* bone marrow-derived cells similarly to the previous experiments performed on the *wt* background.

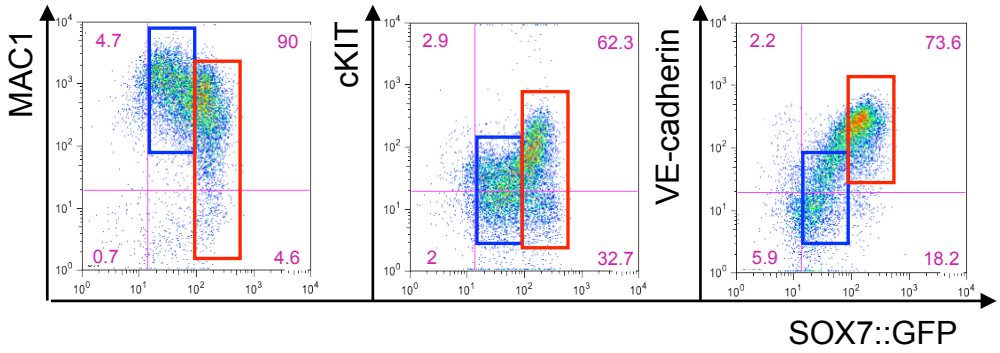


Figure 3.18. Characterization of *p53*$^{-/-}$*Sox7*-expressing cells.
Representative FACS plot of SOX7::GFP, MAC1, cKIT and VE-cadherin expression profiles of *p53*$^{-/-}$*iSox7* bone marrow cells cultured with doxycycline at the plating round 4. Percentages are shown in the indicated quadrants. SOX7::GFP^{low}-expressing cells are gated in blue while SOX7::GFP^{high}-expressing cells are in red. FACS plots are representative of three independent experiments.

To define their nature and clonogenic potential, these three sub-populations were sorted and replated in semi-solid culture. After one week in culture in the presence of doxycycline, the SOX7::GFP^{high}cKIT^{+} sub-population formed more colonies when compared to the other populations (Figure 3.19). Therefore, the double positive sub-population seems to best maintain the capacity to produce colonies and appears to be the most responsive to enforced *Sox7* expression as it was observed for the SOX7::GFP^{high}cKIT^{+} population on the *p53*$^{+/+}$ background.

Altogether, these data confirm the results previously obtained showing that bone marrow cells are responsive to enforced *Sox7* expression. Furthermore, the absence of *p53* seems to considerably increase the formation of colonies and frequency of SOX7::GFP^{high}cKIT^{+} cells compared to *p53 wt* cells.

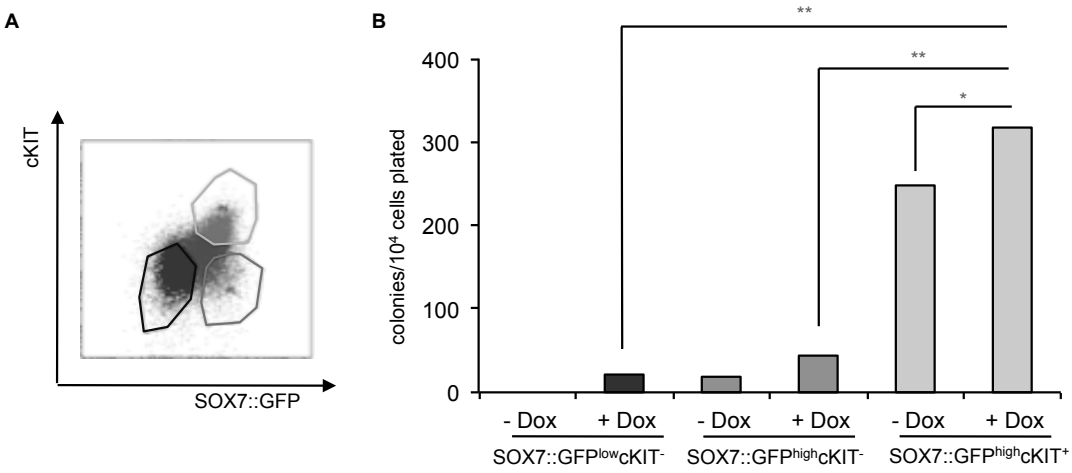


Figure 3.19. SOX7::GFP^{high}cKIT^{+}-expressing cells are able to maintain the clonogenic potential.
A) After the plating round 3, *p53*$^{-/-}$*iSox7*$^{+}$ bone marrow cells were sorted based on SOX::GFP and cKIT expression and 100,000 sorted cells were cultured in semi-solid culture with or without doxycycline. B) After one week in culture, colonies were counted. The bar chart shows the colony number for each sample (n=3; **P*<0.05; ***P*<0.01). Graph is representative of two independent experiments.

3.2.7. *In vivo* experiments

To assess *in vivo* whether *Sox7* expression may confer a leukaemic potential, mouse *iSox7*$^{+}$ clonogenic colonies derived from serial replatings were injected into NOD-SCID IL2Rγ$^{-/-}$ mice (NSG) mice fed or not with doxycycline. These recipient mice have impaired immunological functions, featuring an absence of mature T and B cells, a lack of functional NK cells and deficiencies in cytokine signalling. As a consequence, engraftments of injected cells are better than in other mice with little to no graft rejection.

3.2.7.1. *iSox7*$^{+}$ or *p53*$^{-/-}$ *iSox7*$^{+}$ colonies generated in semi-solid culture do not retain the ability to engraft *in vivo*

To assess *in vivo* whether *Sox7* expression may confer a leukaemic potential, *iSox7*$^{+}$ and *p53*$^{-/-}$ *iSox7*$^{+}$ clonogenic colonies derived from serial replatings were injected into immuno-compromised mice fed or not with doxycycline (Figure 3.20 A). Blood sample analysis and weight measurement was performed every 4 weeks to assess the impact of *Sox7* expression on the possible development of haematological malignancy. Following the weight of the *iSox7*$^{+}$ or *p53*$^{-/-}$ *iSox7*$^{+}$ mice until 24 weeks after the injection showed no significant changes (data not shown). Furthermore, analysis of the blood samples showed no engraftment of the injected cells; CD45.1 expression, which marks the recipient cells, was readily detected while CD45.2 expression, marking the donor cells was not observed (Figure 3.20 B).

A possible reason for the lack of CD45.2 detection, and therefore of engraftment, might be due to the high VE-cadherin expression observed on *iSox7* injected cells that does not allow homing of these cells to the bone marrow. Another interpretation for the lack of donor cells detection in blood samples might be that these

cells are blocked in the bone marrow and therefore difficult to detect in the peripheral blood. To address these hypotheses the analysis of spleen and bone marrow cells of the mice fed with and without doxycycline were performed. As shown in figure 3.20 C, cells derived from both mice did not express VE-cadherin and no SOX7::GFP expression was detected. This analysis revealed that the *iSox7*$^{+}$ or *p53*$^{-/-}$ *iSox7*$^{+}$ cells did not engraft. Therefore, based on these experiments it seems that colonies generated in semi-solid culture did not retained the ability to engraft and therefore, these cells do not seem to have the ability to confer leukaemia or other haematological malignancy.

3.3. Summary of findings

Given the striking impact of *Sox7*-expression on embryonic haematopoiesis, I hypothesized that the mis-expression of *Sox7* in adult cells may confer a proliferative or survival advantage while blocking lineage maturation. Hence, *Sox7* might be implicated in the emergence or maintenance of leukaemia initiating cells.

Based on this hypothesis, I have analysed the outcome of *Sox7* ectopic expression in cells derived from adult mice harbouring an inducible *Sox7* transgene. To define whether *Sox7* enforced expression induces proliferation in adult cells I used a serial clonogenic replating assay. Interestingly, I observed a progressive increase in colony numbers in replating assays performed with *iSox7* bone marrow cells in the presence of doxycycline while no significant changes were detected in the other conditions, *wild-type* bone marrow cells in the presence or absence of doxycycline and *iSox7* bone marrow cells without doxycycline. To extend this finding and further define the nature of the cell types amplified upon *Sox7*-enforced expression the serial clonogenic-replating assay was performed using spleen and thymus cells. In this case, I also observed a progressive increase in colony numbers upon serial replating in culture

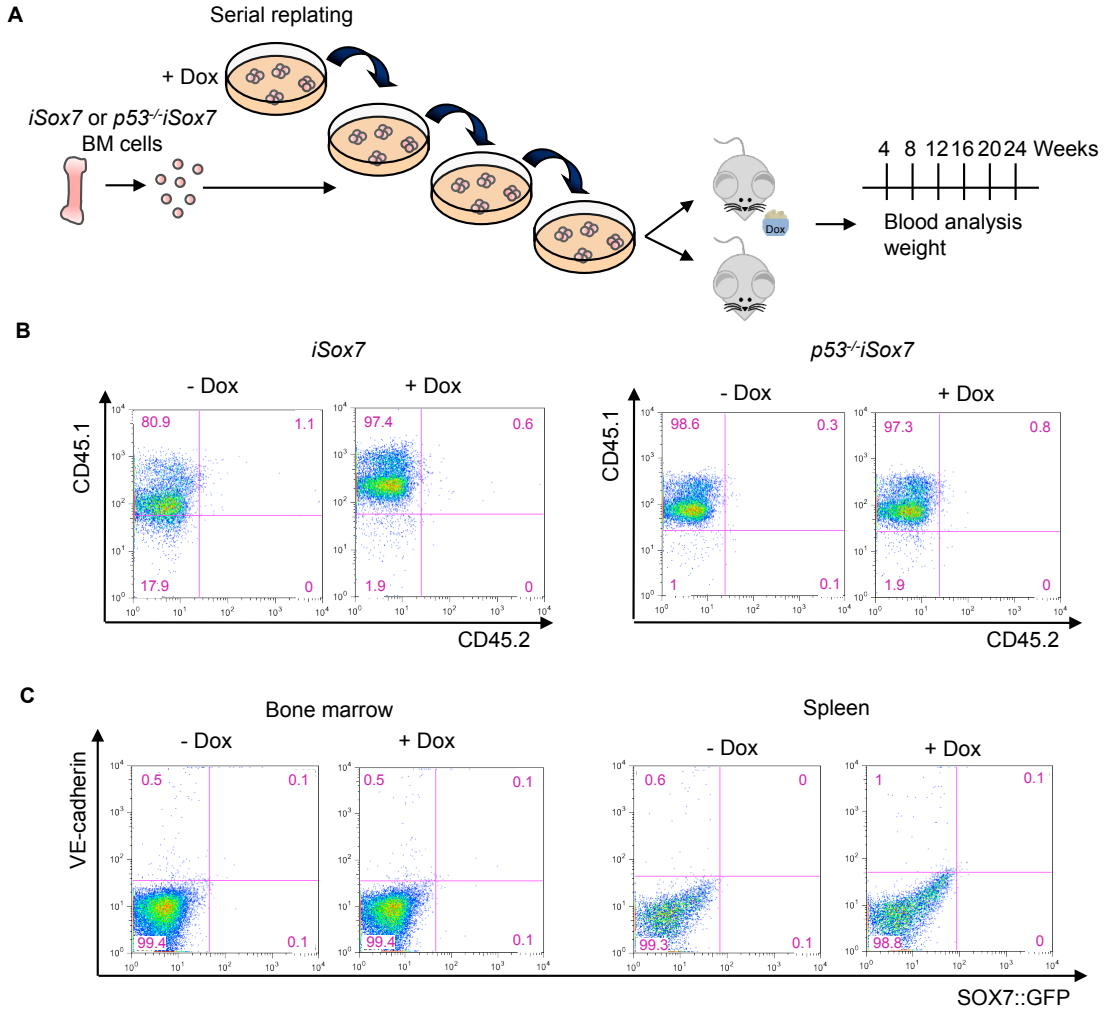


Figure 3.20. *Sox7*$^{+}$ or *p53*$^{-/-}$ *iSox7*$^{+}$ colonies generated in semi-solid culture do not retain the ability to engraft *in vivo*.
A) Schematic representation of the *in vivo* assays. After the plating round 4, *p53*$^{-/-}$-*iSox7*$^{+}$ or *iSox7*$^{+}$ bone marrow cells were injected into immuno-compromised mice fed or not with doxycycline. Every four weeks mice were weighted and blood samples were analyzed. B) Engraftment analysis performed at 16 weeks after injection. Representative FACS plots of CD45.1 (recipient cells) and CD45.2 (donor cells) expression profiles of blood samples from immuno-compromised mice fed with or without doxycycline. C) Representative FACS plots of SOX7::GFP and VE-cadherin expression profiles of bone marrow and spleen cells from immuno-compromised mice fed with or without with doxycycline. Percentages are shown in the indicated gates. FACS plots are representative of two independent experiments.

derived only from *iSox7* spleen cells in the presence of doxycycline suggesting that a subset of cells within the spleen are also responsive to enforced *Sox7* expression.

Next, I evaluated whether the increase in proliferation reflected an increase in the frequency of SOX7::GFP-expressing cells within the bone marrow and spleen cell replating cultures. During the serial replating FACS analysis was performed and indeed revealed an increase in the frequency of SOX7::GFP^{+} cells. I also characterized the immune-phenotype of cells induced to proliferate upon *Sox7* induction and the FACS analysis has shown two sub-populations based on the SOX7::GFP expression. The bulk of SOX7::GFP^{high} cells were MAC1$^{high/low}$, CD45^{high}, cKIT$^{+/-}$, CD34^{-}, GR1^{-}, VE-cadherin^{high} and CD48^{high}. The flow cytometry analysis has also revealed the presence of a sub-population of cells with different level of cKIT and SOX7::GFP expression. Further replating of the sorted populations revealed that the SOX7::GFP^{high}cKIT^{+} subpopulation contains most of the clonogenic potential as it best maintained the capacity to produce colonies.

To evaluate the extent of differentiation of cells amplified upon *Sox7*-enforced expression the morphology of the cells derived from these replatings has been analysed. The analysis revealed that *Sox7*-enforced expression impairs haematopoietic differentiation as was previously observed in embryonic haematopoiesis (Gandillet et al., 2009).

I also investigated the hypothesis that a second hit, such as the deletion of *p53*, might be necessary to induce haematological malignancy. This experiment has further validated the results obtained in the experiments carried out on *wt* background, that bone marrow cells are responsive to enforced *Sox7* expression that the SOX7::GFP^{high}cKIT^{+} sub-population contains most clonogenic potential, and that *Sox7*-enforced expression impairs haematopoietic differentiation.

To assess *in vivo* whether *Sox7* expression may confer a leukaemic potential, *iSox7*$^{+}$ and *p53*$^{-/-}$*iSox7*$^{+}$ clonogenic colonies derived from serial replating have been injected into immuno-compromised mice fed or not with doxycycline. The *iSox7*$^{+}$ and *p53*$^{-/-}$ *iSox7*$^{+}$ bone marrow cells did not show sign of engraftment and therefore did not appear to confer leukaemia.

Chapter 4

4. *SOX7* expression in human leukaemia samples

4.1. Introduction

Acute lymphoblastic leukaemia (ALL) is the most frequent type of cancer in children. Childhood ALL is curable in most cases and the survival rate is about 80 per cent. On the contrary, the rate of cure for adults is only 40 per cent (Pui et al., 2004; Pui et al., 2008). Progresses have been made in the treatment of leukaemia but there is a great demand for the development of new less toxic therapies. Therefore, a better understanding of the molecular mechanisms underlying the emergence and maintenance of ALL is fundamental.

Several SOX family members play an important role in tumour development. In particular, it has been reported that the expression of *SOX7* is frequently down-regulated in many solid tumours such as prostate, colon and endometrial cancers (Chan et al., 2012; Guo et al., 2008). However, the possible role of SOX7 in blood malignancy has not been investigated to date. Gandillet and college have reported using a murine model that *Sox7* knock-down *in vitro* significantly inhibits haematopoietic differentiation and that the enforced expression of *Sox7* in the earliest committed haematopoietic precursors promotes proliferation while blocks their differentiation (Gandillet et al., 2009).

Of interest, one of the downstream targets of SOX7 identified in embryonic

haematopoietic cells is EVI-1 (unpublished data). Over-expression of *Evi-1* occurs with high frequency in leukaemic patients and is an independent negative prognostic indicator. Furthermore, preliminary data indicate the over-expression of *SOX7* in few cases of childhood ALL (Trentin et al., 2009). Based on these facts, it was tempting to speculate that *SOX7* might be implicated in leukaemia development.

Therefore, analysis of the expression of *SOX7* transcripts in human cases of leukaemia was performed to confirm and extend this preliminary report. My aim here was to determine whether *SOX7* was up-regulated in haematological malignancies, in both paediatric and adult samples and whether the possible increased expression correlated with specific type of leukaemia.

4.2. Results

4.2.1. *SOX7* is not expressed in healthy bone marrow samples

Since preliminary data indicated the over-expression of *SOX7* in few cases of childhood ALL (Trentin et al., 2009), one important question to address was whether *SOX7* is expressed in healthy non-leukaemic samples.

To address this question, different existing datasets were used in the analysis. The gene expression database from the University of Stanford was consulted for the expression of *Sox7* in mouse haematopoiesis. As shown in figure 4.1 A *Sox7* transcripts were not detected in any haematopoietic compartment. Moreover, in the Oncomine database *SOX7* is reported not to be expressed in both healthy human bone marrow and blood samples (Figure 4.1 B).

To confirm and validate those data, *Sox7* expression was analysed in healthy *wt* mouse bone marrow samples since healthy human bone marrow or blood samples were

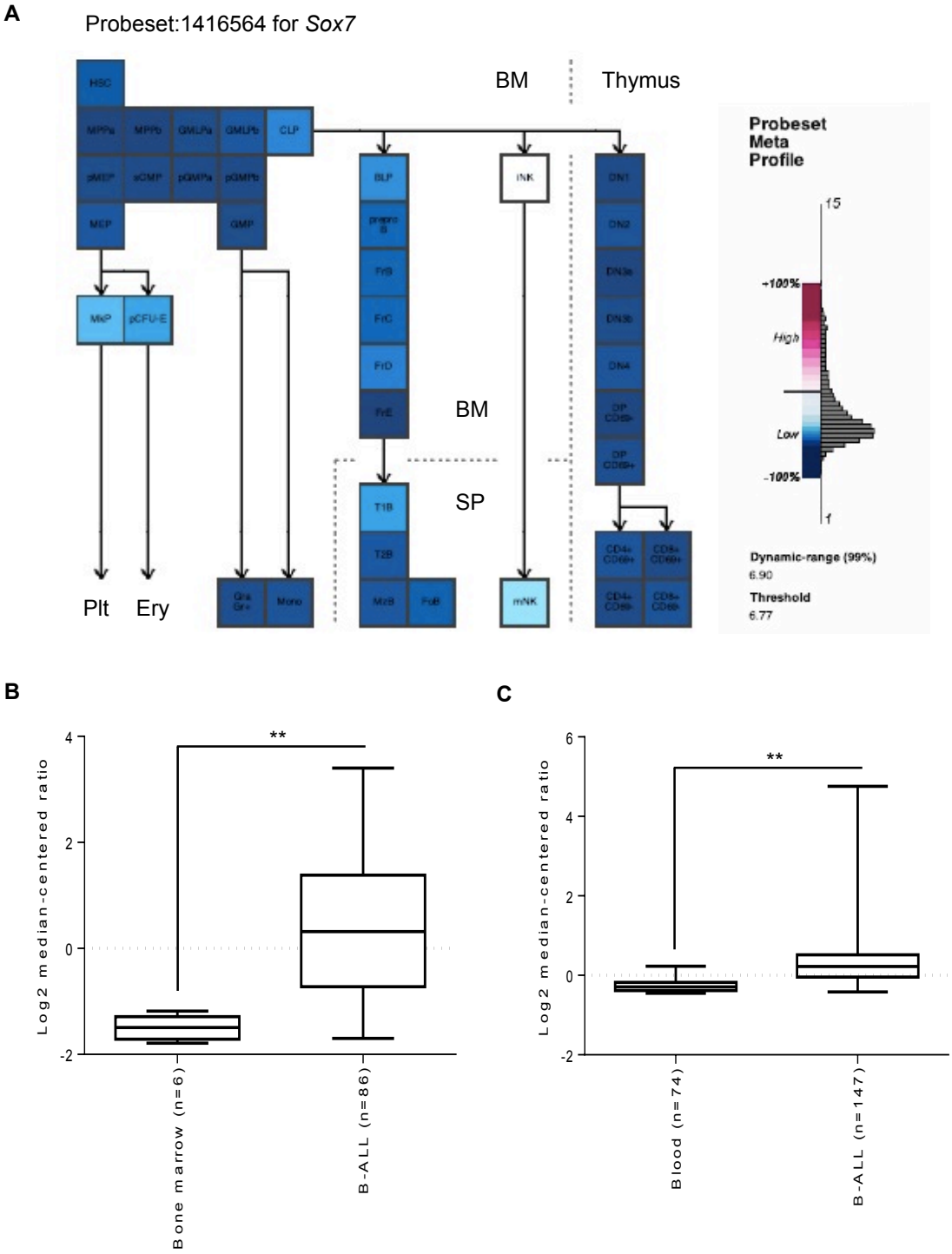


Figure 4.1. *Sox7* is not expressed in healthy samples.
A) Gene expression commons (https://gexc.stanford.edu/) on mouse haematopoiesis model submitted by Jun Seita (Stanford University) Reference: PubMed ID:, GEO GSE ID: 34723. BM: bone marrow, SP: spleen. B-C) *SOX7* expression in human healthy and leukaemia samples (Anderson's study on the graph B and Haferlach's study on the graph C, modified from Oncomine database, ** P<0.01).

no available. In particular, since *SOX7* appears to be expressed mainly in ALL, mouse bone marrow cells were sorted based on the expression of B lymphoid markers. Different populations were sorted starting from the common lymphoid progenitors to mature B cells as shown in figure 4.2. In particular, CLP population was sorted based on the absence of Lineage markers and the presence of cKIT and SCA1 markers. The three immature B cell sub-populations (Pre-pro B, Pro B and Late-pro B) were sorted according to their specific expression of B220, CD43, BP-1 and CD24 markers. The more mature B cell sub-populations (Small pre B, Immature B and Mature B) were sorted based on their expression of B220, IgD and IgM. All the sub-populations were gated on live cells, $7AAD^-$. Then, qRT-PCR was performed on the cDNA prepared from these sorted sub-populations and the expression of *Sox7* was analysed. In addition, *Pax5*, a well-known transcription factor involved in B cell development, was also included in the analysis as a positive control. As expected, *Sox7* was not detectable in any sub-populations. On the contrary, *Pax5* was detectable at good level in most B cell sub-populations except for the most immature CLP (Figure 4.3).

Altogether, these results suggest that *SOX7* is not expressed in both mouse and human healthy bone marrow samples and indicate that *SOX7* expression detected in cases of human ALL might be directly linked to the malignancy.

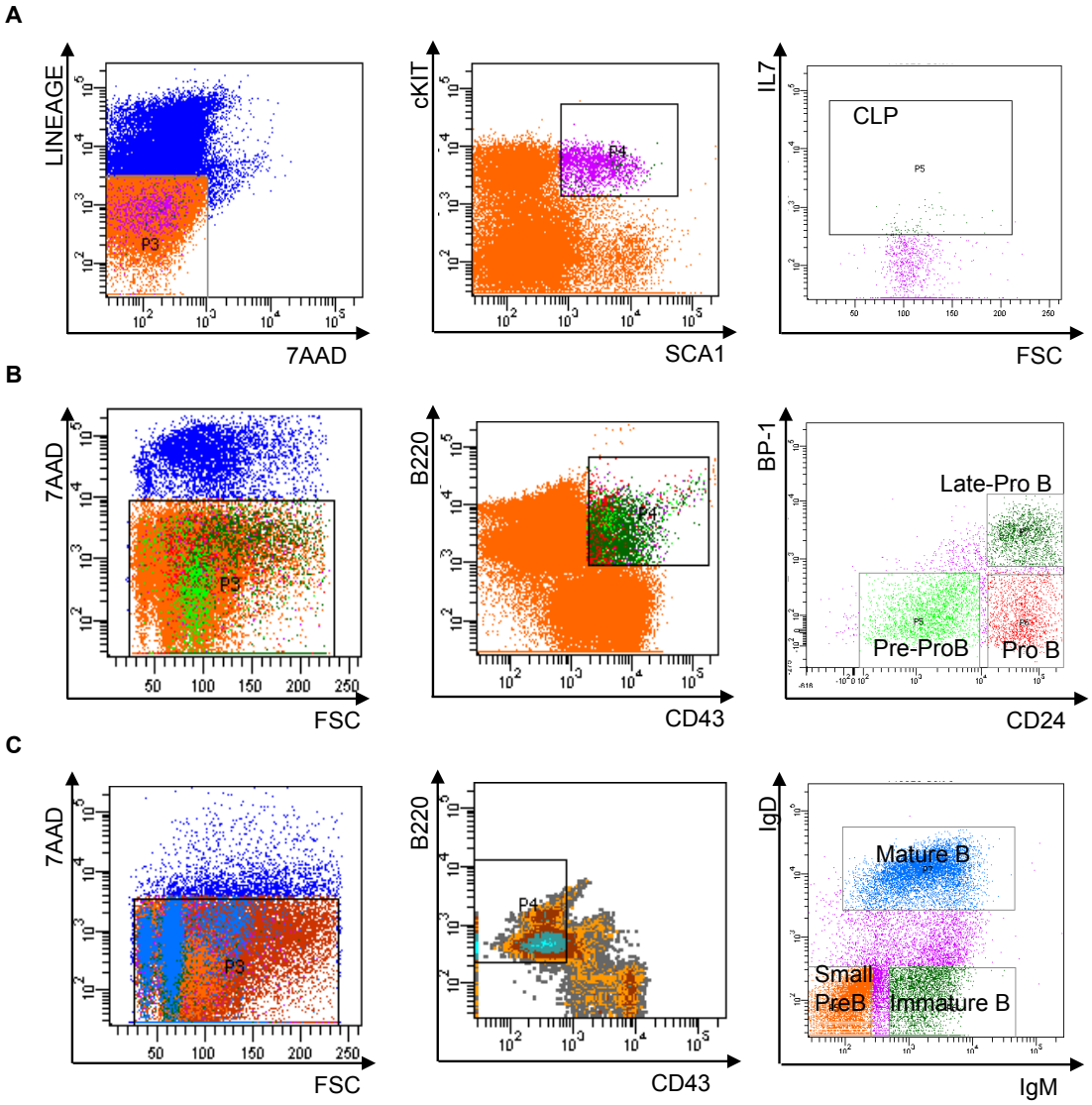


Figure 4.2. Sorting strategy of B cell sub-populations.
Representative FACS plots of B cell markers expression profile of bone marrow cells. A) CLP population was sorted based on Lineage, SCA1, cKIT and IL7 expression. B) Immature B cells were sorted based on B220, CD43, CD24 and BP1 expression. C) Mature B cells were sorted based on B220, CD43, IgD and IgM. Sorted populations were gated on live cells 7AAD$^-$. Percentages are shown in the indicated gates. FACS plots are representative of one independent experiment.

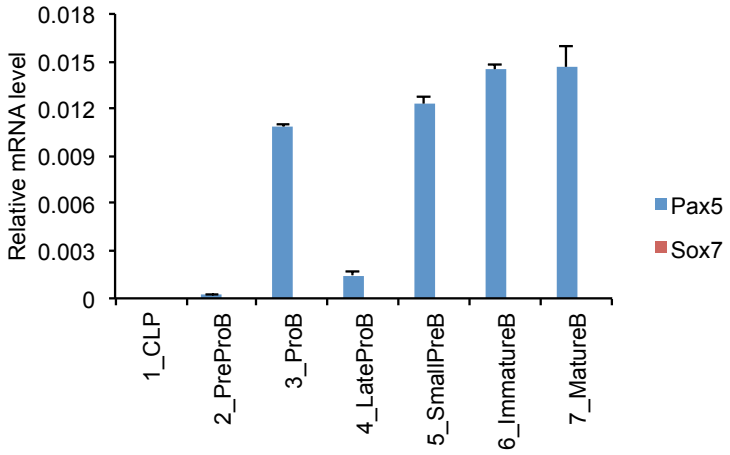


Figure 4.3. *Sox7* is not expressed in mouse B cell sub-populations.
Quantitative RT-PCR analysis of *Sox7* and *Pax5* transcript levels relative to βactin in sorted B cell sub-populations from mouse bone marrow. Error bars indicate mean ± standard deviation (n=3). Graph is representative of one independent experiment.

4.2.2. *SOX7* is over-expressed in human B-Acute Lymphoblastic Leukaemia samples

To determine whether *SOX7* was up-regulated in haematological malignancies and whether the possible increased expression correlated with specific type of leukaemia, qRT-PCR was performed on cDNA isolated from several childhood and adult acute lymphoblastic leukaemia samples.

Interestingly, in both childhood and adult human leukaemia samples tested, the detection level for *SOX7* transcripts was relatively high when compared to *SOX17* or *SOX18* transcripts that were not detected or at very low levels (Figure 4.4 A-C). *SOX7* expression appeared heterogeneous across all leukaemia samples tested but in some of these genotypes the expression was much higher than in others, in particular in some of the childhood ALL samples such as ETV6-RUNX1, MLL and progenitor B. Moreover, childhood bone marrow AML samples were tested for *SOX7* expression as well.

One out of three patients expressed high level of *SOX7* and this expression was much higher than any of those observed in AML samples (Figure 4.4 D).

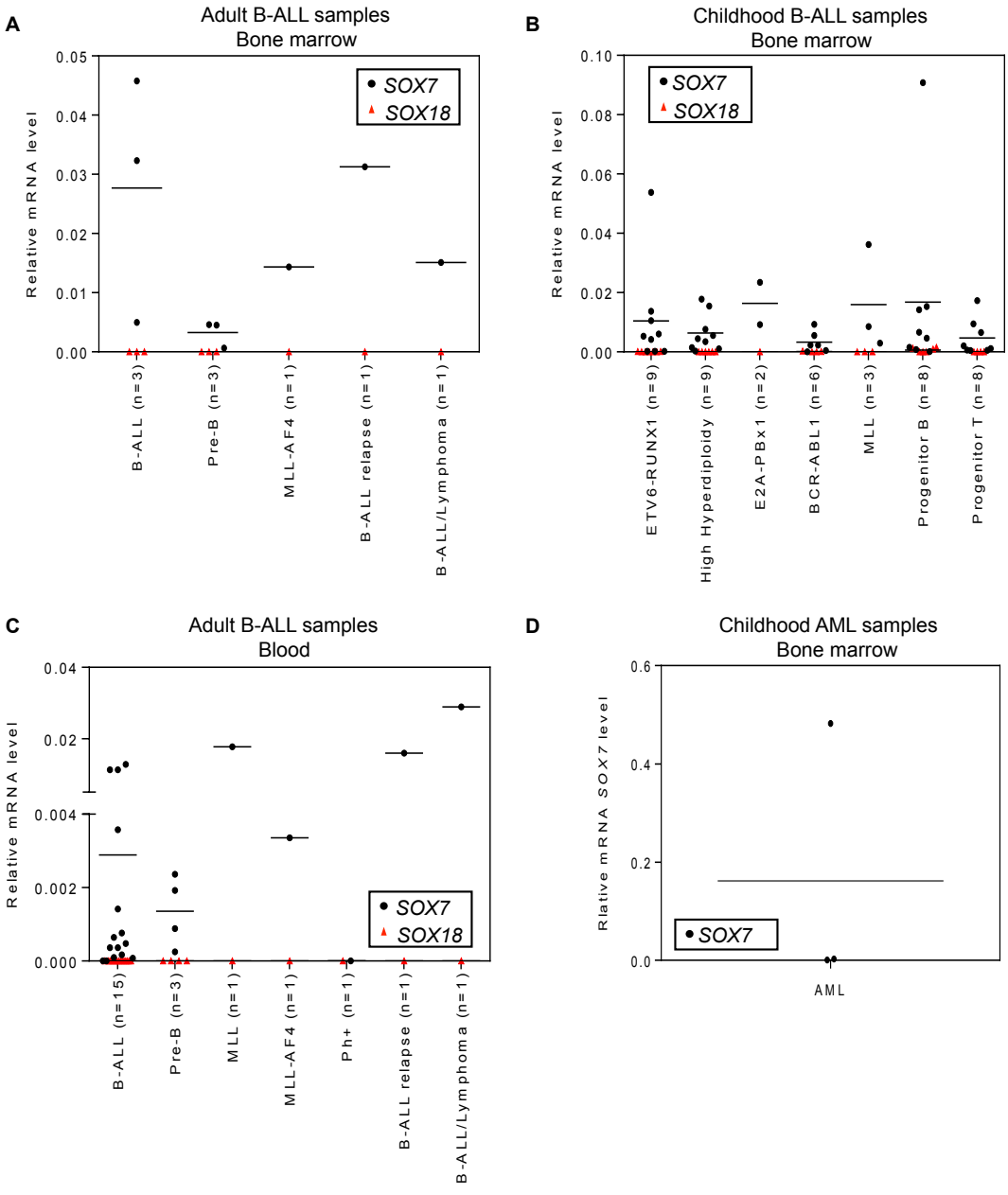


Figure 4.4. *SOX7* is expressed in B-ALL.
Quantitative RT-PCR analysis of *SOX7* and *SOX18* transcript levels relative to β2microglobulin in childhood and adult human B-ALL samples (A-C). Human Childhood AML samples are shown in panel D. Average between samples is indicated in the graphs.

To define whether *SOX7* is significantly expressed at higher levels in specific leukaemias more samples should be tested; however, I did not have access to more patient samples. To overcome this limitation, the Oncomine microarray database was consulted (Rhodes et al., 2004); analysis of leukaemia datasets supported my qRT-PCR data revealing the expression of *SOX7* in large cohorts of B-ALL samples but not in other types of haematological malignancies (Andersson Leukaemia study, 127 samples and Haferlach Leukaemia study, 2096 samples) (Figure 4.5).

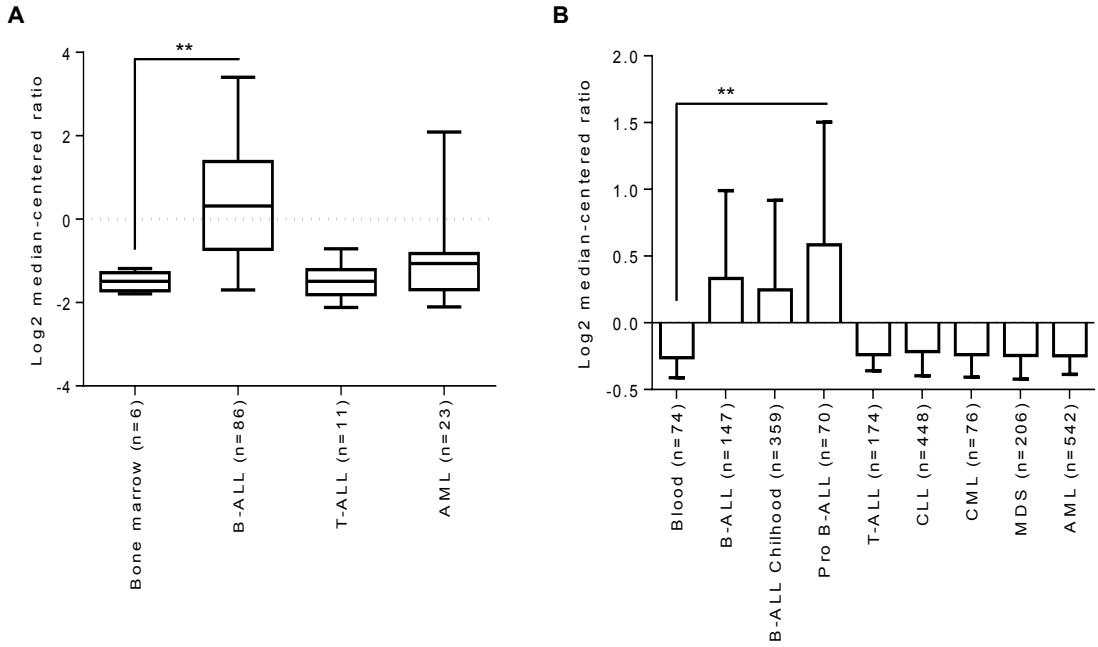


Figure 4.5. *SOX7* is specifically expressed in human B-ALL.
SOX7 expression in human healthy samples and in different types of leukaemias (Anderson's study on the bar graph A and Haferlach's study on the bar graph B, modified from Oncomine database; ** P<0.01). AML: Acute Myeloid Leukaemia, B-ALL: B Cell- Acute Lymphoblastic Leukaemia, B-ALL Childhood: B-Cell childhood Acute Lymphoblastic Leukaemia, CLL: Chronic Lymphocytic Leukaemia, CML: Chronic Myelogenous Leukaemia, MDS: Myelodysplastic Syndrome, Pro B-ALL: Pro-B Acute Lymphoblastic Leukaemia, T-ALL: T-Cell Acute Lymphoblastic Leukemia.

Moreover, the analysis clearly showed a specific increase in *SOX7* expression in B-ALL when compared to the expression of its close homologues, *SOX17* and *SOX18,* which are more heterogeneously expressed across all types of leukaemias (Figure 4.6). This analysis was performed for datasets from three independent studies analysed from the Oncomine database as shown by the heatmap representations in figure 4.6.

In addition, to investigate whether *SOX7* expression correlated with a specific type of mutation within B cell subtypes of leukaemias, the Oncomine database was consulted once again. Two independent studies were analysed; the first study showed an increased in *SOX7* expression in samples carrying the ETV6-RUNX1 translocation while a decrease in *SOX7* expression was observed in samples with the BCR-ABL fusion gene (Figure 4.7 A). These differences were largely attenuated in the second study where more samples were included in the analysis (Figure 4.7 B). Therefore, these data suggest that *SOX7* expression does not correlate with a specific type of translocation or mutation involved in B-ALL.

Altogether, these results suggest that *SOX7* is specifically expressed in human B-ALL samples.

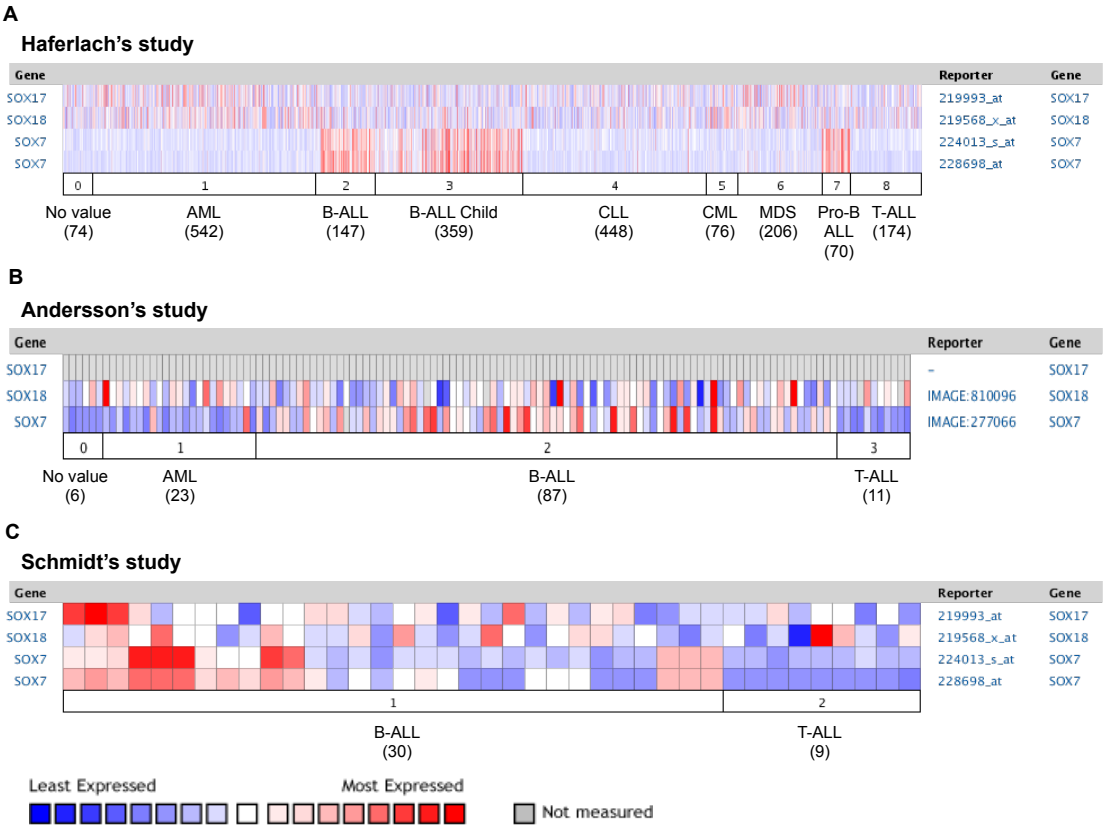


Figure 4.6. *SOX7* is specifically expressed in human B-ALL.
SOX7 expression dataset of human leukaemia samples obtained by Oncomine microarray database. Three different studies are shown (A-C). AML: Acute Myeloid Leukaemia, B-ALL: B Cell- Acute Lymphoblastic Leukaemia, B-ALL Child: B-Cell childhood Acute Lymphoblastic Leukaemia, CLL: Chronic Lymphocytic Leukaemia, CML: Chronic Myelogenous Leukaemia, MDS: Myelodysplastic Syndrome, Pro B-ALL: Pro-B Acute Lymphoblastic Leukaemia, T-ALL: T-Cell Acute Lymphoblastic Leukemia.

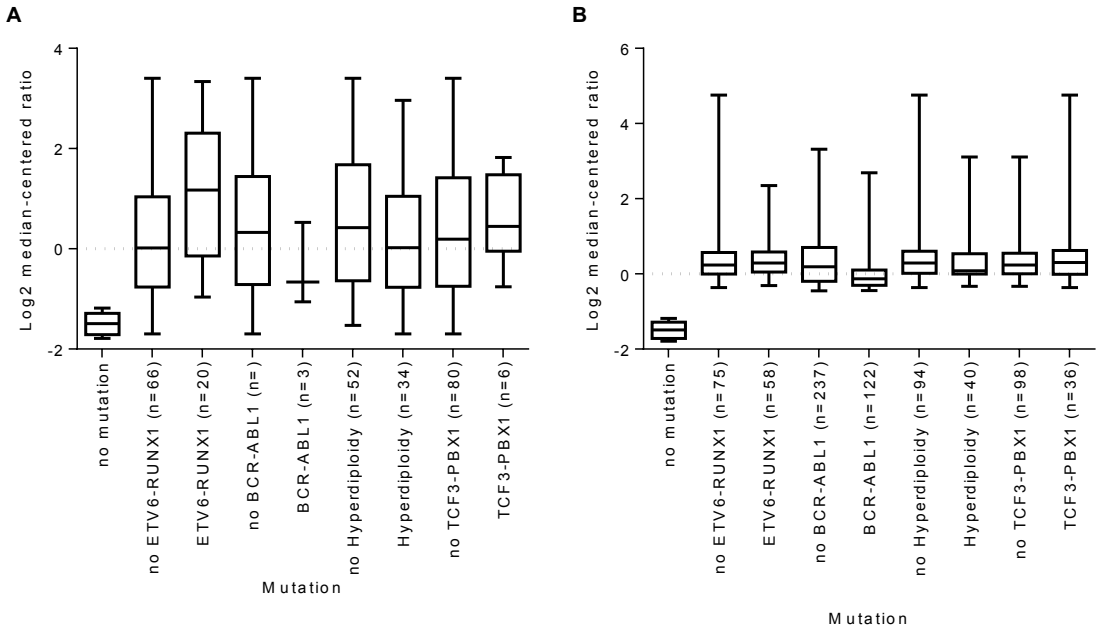


Figure 4.7. *SOX7* expression does not correlate with a specific type of mutation involved in B-ALL.
SOX7 expression in human B-ALL samples subdivided based on mutations (Anderson's study on the panel A and Haferlach's study on the panel B, modified from Oncomine database).

4.2.3. *SOX7* is expressed in human B-ALL cell lines and primo-graft samples

To further establish a possible correlation between *SOX7* and leukaemia, the following cell lines were tested for *SOX7* expression: SupB15, REH, RS4;11, BV137, NALM6, SD1, 697, MV4;11. SupB15, REH, RS4:11, BV137, NALM6, SD1 and 697 which are all B-cell ALL cell lines while MV4;11 is an AML cell line. The molecular genetic of SupB15, BV137 and SD1 cell lines is characterized by the expression of the fusion gene BCR-ABL. The REH cell line contains a TEL-AML1 translocation while RS4:11 and MV4-11 cell lines both contain a MLL-AF4 chromosomal translocation (e10-e4) and MLL-AF4 (e9-e5). The NALM6 cell line contains a PDGFRβ-ETV6 translocation and finally, the 697 cell line is characterized by TCF3-PBX1 fusion gene.

All the B-ALL cell lines, except for SD1, expressed *SOX7* transcript at different levels, from highest expression in the SupB15 and NALM6 cell lines to lowest expression in BV173. In contrast, the AML cell line MV4;11 expressed very low level of *SOX7* (Figure 4.8 A) as expected from the human samples analyses described above. In addition, *SOX7* expression levels in all B-ALL cell lines tested were quite heterogeneous in agreement with *SOX7* expression in human primary samples.

Moreover, human B-ALL primo-graft samples were tested for *SOX7* expression. These human samples were amplified by engrafting in mice and represent a very good model for testing the role of *SOX7* in leukaemogenesis. The cytogenetic of TR and CD19 3.1 samples is high hyperdiploidy. CD13 5.2 is a very high-risk sample compared to CD13 11.2 that is a standard-risk sample. The 694E sample contains high level of minimal residual disease, MLL-AF4 contains the MLL-AF4 chromosomal translocation, and finally, CD20 1.1 sample contains the ETV6-RUNX1 fusion gene. As observed in both primary samples and cell lines, *SOX7* was expressed at different levels in all B-ALL

primo-graft tested (Figure 4.8 B).

In order to determine the protein expression level of SOX7 in these cell lines, a western blotting analysis was performed using the home-made SOX7 polyclonal antibody initially derived against mouse SOX7. Immuno-blotting of protein extracts from three B-ALL lines revealed that this SOX7 antibody cross-reacts with human SOX7. Protein extract from MS1, a mouse endothelial cell line, was used as a positive control for SOX7 detection. All cell lines tested expressed readily detectable level of SOX7 protein (Figure 4.8 C).

Altogether, these data confirm the expression of *SOX7* in multiple B-ALL subtypes of leukaemias in cell lines and primo-graft samples.

4.3. Summary of findings

Since the sustained expression of *Sox7* in early mouse haematopoietic precursors was sufficient to completely alter the balance between proliferation and differentiation at the onset of haematopoiesis and the over-expression of *SOX7* was observed in few cases of childhood ALL, it was tempting to speculate that *SOX7* might be implicated in the emergence and/or maintenance of leukaemia.

Based on these facts, I analysed the expression of *SOX7* transcripts in multiple human cases of leukaemia. A qRT-PCR analysis was performed on childhood and adult human samples and the Oncomine database was consulted for a more complete analysis. Data indicated high expression levels of *SOX7* transcripts in ALL, in particular in B cell subtypes of leukaemia. Interestingly, the relative expression of *SOX7* was high when compared to *SOX17* and *SOX18* transcripts that were only detected at minimal level.
In addition, *SOX7* expression was heterogeneous across B-ALL samples and did not correlate with a specific type of translocation or mutation.

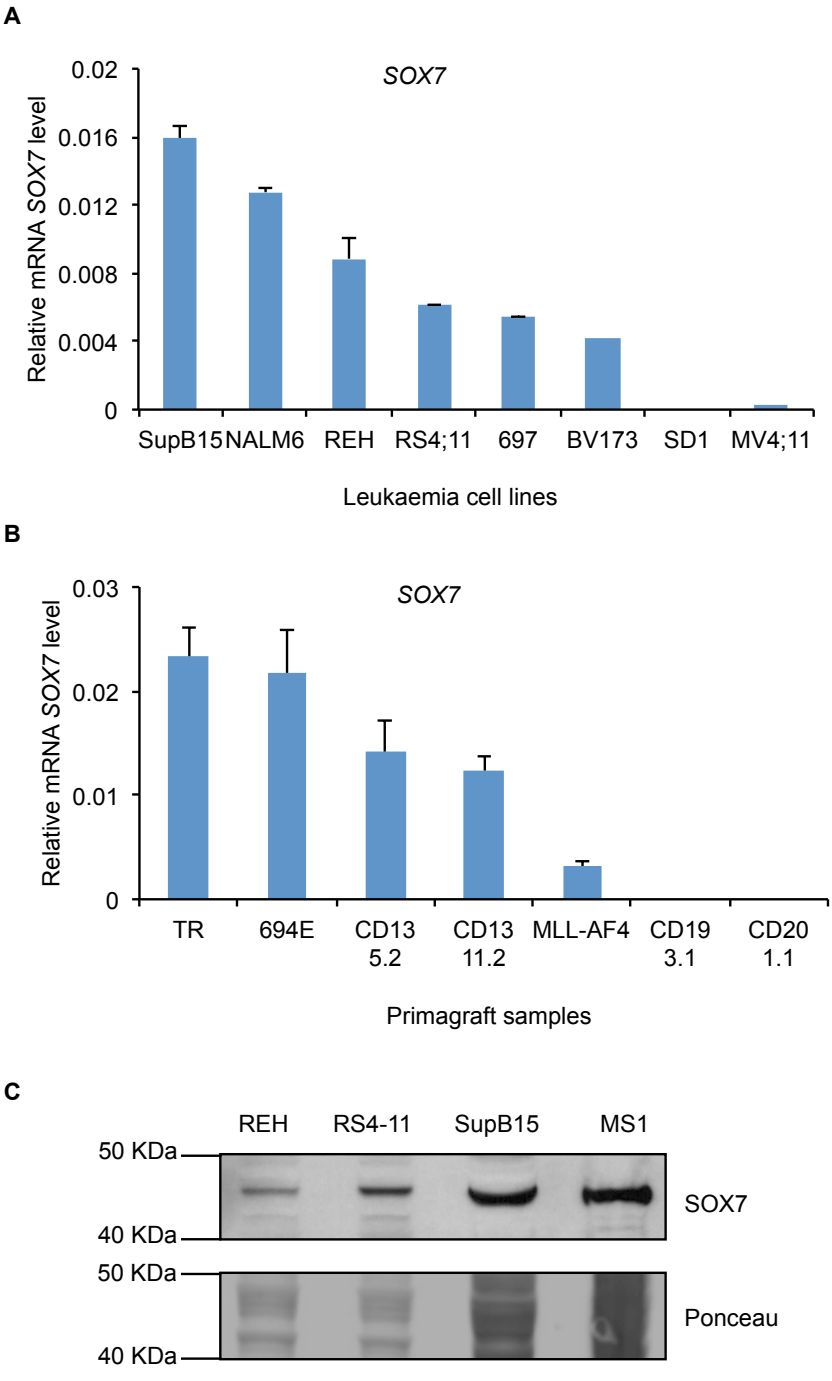


Figure 4.8. *SOX7* is expressed in human B-ALL cell lines and primo-graft samples. Quantitative RT-PCR analysis of *SOX7* transcript levels relative to β2microglobulin in human leukaemia cell lines (A) and primo-graft samples (B). Error bars indicate mean ± standard deviation (n=3). SupB15 (BCR-ABL), NALM6 (PDGFRB-ETV6), REH (TEL-AML1), RS4;11 (MLL-AF4(e10-e4)), 697 (TCF3-PBX1), BV173 (BCR-ABL) and SD1 (BCR-ABL) are B-ALL cell lines. MV4;11 (MLL-AF4(e9-e5)) is AML cell line. TR (high hyperdiploidy), 694E (high level of minimal residual disease), CD13 5.2 (very high-risk sample-VHR03), CD13 11.2 (standard risk- SR03), MLL-AF4 (MLL-AF4), CD19 3.1 (high hyperdiploidy) and CD20 1.1 (ETV6-RUNX1) are all B-ALL prima-graft samples. C) Western blotting analysis for SOX7 of whole protein extract from leukaemia cell lines. Ponceau was used as loading control.

To further document the correlation between *SOX7* expression and leukaemia, cell lines and primo-graft samples were also tested for the expression of *SOX7* transcripts. Most B-ALL cell lines tested expressed *SOX7* at high level compared to the AML cell line, used as a negative control. Only one B-ALL cell line, SD1, did not expressed *SOX7* confirming the heterogeneous expression of *SOX7* in B cell leukaemia samples. In addition, similar *SOX7* expression levels were observed in most primo-graft samples tested. The B-ALL cell lines and primo-graft samples represent very good models to test our hypothesis as they are able to induce leukaemia in mice upon transplantation and therefore will allow us to test the potential role of SOX7 in leukaemogenesis.

Overall, these data clearly demonstrated for the first time a tight correlation between *SOX7* expression and B-ALL suggesting a possible role for *SOX7* in leukaemogenesis.

Chapter 5

5. Analysis of *Sox7*-enforced expression during mouse B cell differentiation

5.1. Introduction

B cell development is a process mediated by stromal contact and is dependent on interleukin-7 (Cumano et al., 1990). The IL-7 cytokine is important for the proliferation of B cell precursors while its presence is no longer necessary for the complete maturation of B cells expressing the IgM and IgD immunoglobulins (Melamed et al., 1997). Another factor important for B cell growth is the contact with bone marrow derived-stromal cells, which mimic the matrix in organs where B cells develop (Clark et al., 2014; Cumano et al., 1990; Hardy and Hayakawa, 1991).

B lymphopoiesis is characterized by different progressive stages of differentiation. Briefly, B cell lymphocytes are generated by the common lymphocyte progenitor cells. Then, CLP cells progress to sequential immature stages encompassing pro B, pre B and immature B cell sub-populations. Finally B cells become mature when expressing both IgM and IgD immunoglobulins (Hagman and Lukin, 2006; Li et al., 1993; Miosge and Goodnow, 2005).

The analysis of *Sox7*-enforced expression in adult bone marrow cells, performed in semi-solid culture, showed that the mis-expression of *Sox7* in adult cells confers a proliferative advantage while blocking lineage maturation as was previously observed

in embryonic haematopoiesis (Gandillet et al., 2009). In particular, progenitor cells expressing the cKIT marker retained the clonogenic potential.

Moreover, the over-expression of *SOX7* was observed in multiple cases of ALL, in particular it correlated with the B cell subtype of leukaemia. Given these facts, I therefore investigated how enforced expression of *Sox7* might affect adult B cell differentiation. The first hypothesis I evaluated was whether *Sox7*-enforced expression impaired B cell development. Secondly, I defined which haematopoietic sub-population was responsive to *Sox7*-enforced expression and finally, I characterized cells proliferating upon *Sox7*-expression both in term of immuno-phenotype and genotype.

5.2. Results

5.2.1. *Sox7*-enforced expression induces the proliferation of bone marrow cells and impairs B cell differentiation

Since *SOX7* is mainly expressed in B-ALL, this part of the project was focused on investigating the consequences of *Sox7*-enforced expression during adult B cell differentiation. The bone marrow was extracted from inducible *Sox7* transgenic mice and serial replating assays were performed on irradiated OP9 culture in medium supplemented with IL-7 in the presence or in the absence of doxycycline (Figure 5.1). The IL-7 induces the proliferation of B cell progenitors while the OP9 stroma was used to support B cell development and maintenance, mimicking the bone marrow microenvironment (Vieira and Cumano, 2004). During successive passages without doxycycline, myeloid development was initially observed as measured by the expression of the surface marker MAC1 but disappeared with time (Figure 5.2 A and B upper panel).

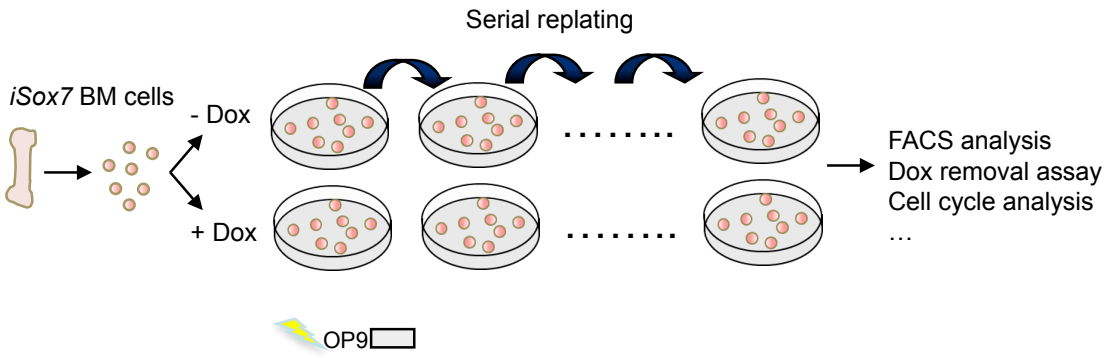


Figure 5.1. Scheme of the protocol used for serial replating assay.
The bone marrow was extracted from femurs of inducible *Sox7* transgenic mice and serial replating assays were performed on irradiated OP9 culture in the presence or in the absence of doxycycline. Different assays were performed at each replating.

In contrast, B lymphopoiesis progressively increased during the time course of the experiment as detected by the expression of B cell markers, B220 and CD19 (Figure 5.2 A and B lower panel). These data suggest that the culture condition used in the serial replating assay efficiently induces B cell maturation and maintenance. Interestingly, when *Sox7* expression was induced, the $MAC1^+$ population was maintained throughout the serial replatings while the $B220^+CD19^+$ population was severely impaired (Figure 5.2 A and B) suggesting a blockage in B cell maturation by *Sox7*-enforced expression. As shown in figure 5.3, the number of cells progressively increased upon serial replatings of *iSox7* bone marrow cells in the presence of doxycycline while no or only modest changes were observed in the absence of doxycycline. The cells cultured with doxycycline were GFP positive as expected (Figure 5.3 B) and expressed increasing level of the *Sox7* transcript upon serial replating (Figure 5.3 C). In addition, upon serial replatings, the *Sox7*-induced cultures contained more immature blast-like cells and fewer lymphocytes compared to the control cultures as observed by May-Grünwald Giemsa staining (Figure 5.3 D). Altogether, these data suggest that *Sox7* over-expression impaired B cell differentiation.

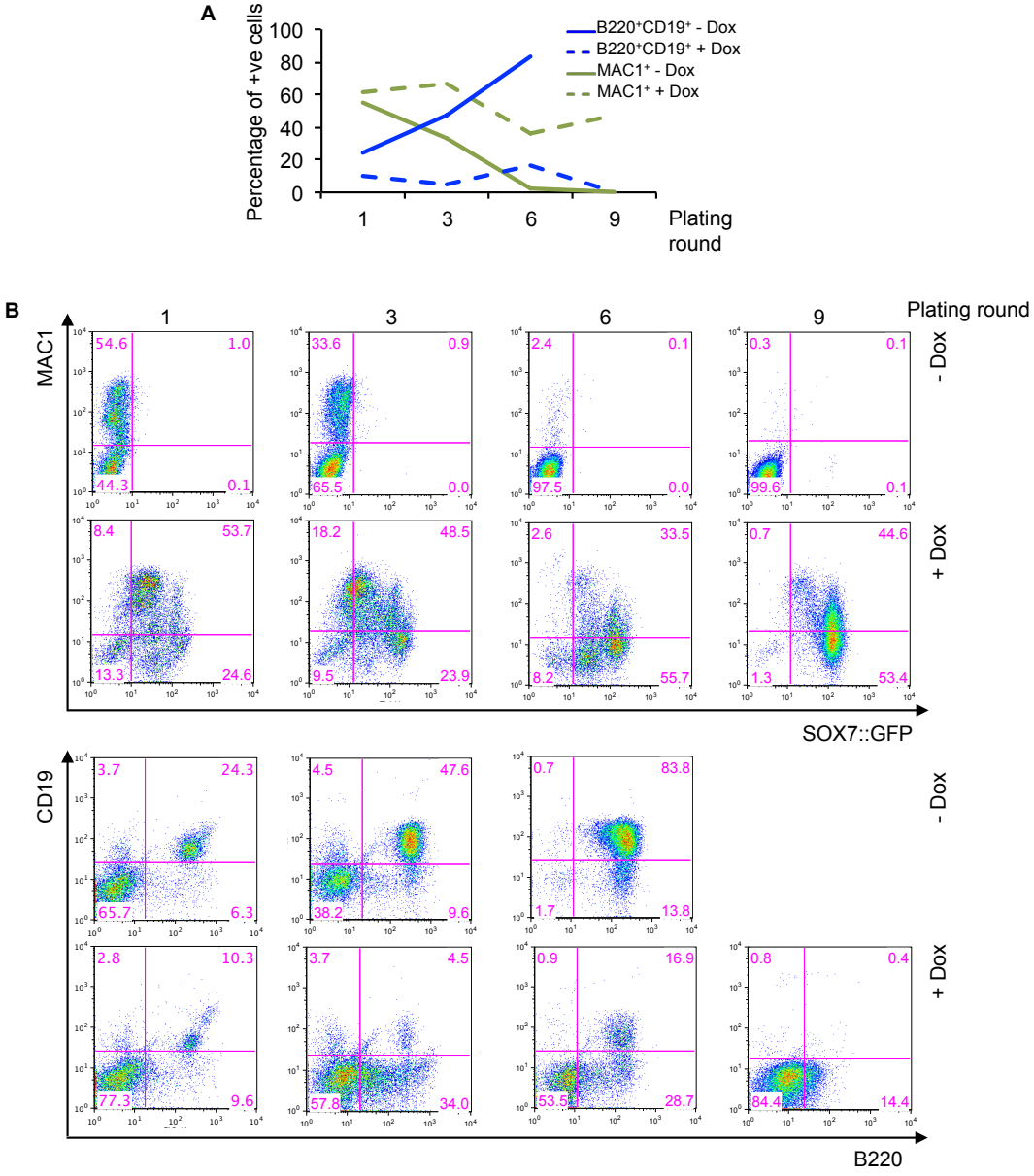


Figure 5.2. The culture condition, used in the serial replating assay, promotes the growth of B cells.
A) Representative line chart of MAC1$^+$ cells and B220$^+$CD19$^+$ cells in *iSox7* bone marrow cells cultured with and without doxycycline. B) Representative FACS plots of MAC1, SOX7::GFP, CD19 and B220 expression profiles of *iSox7* bone marrow cells cultured with and without doxycycline. Percentages of cells are shown in the indicated quadrants. Line chart and FACS plots are representative of four independent experiments.

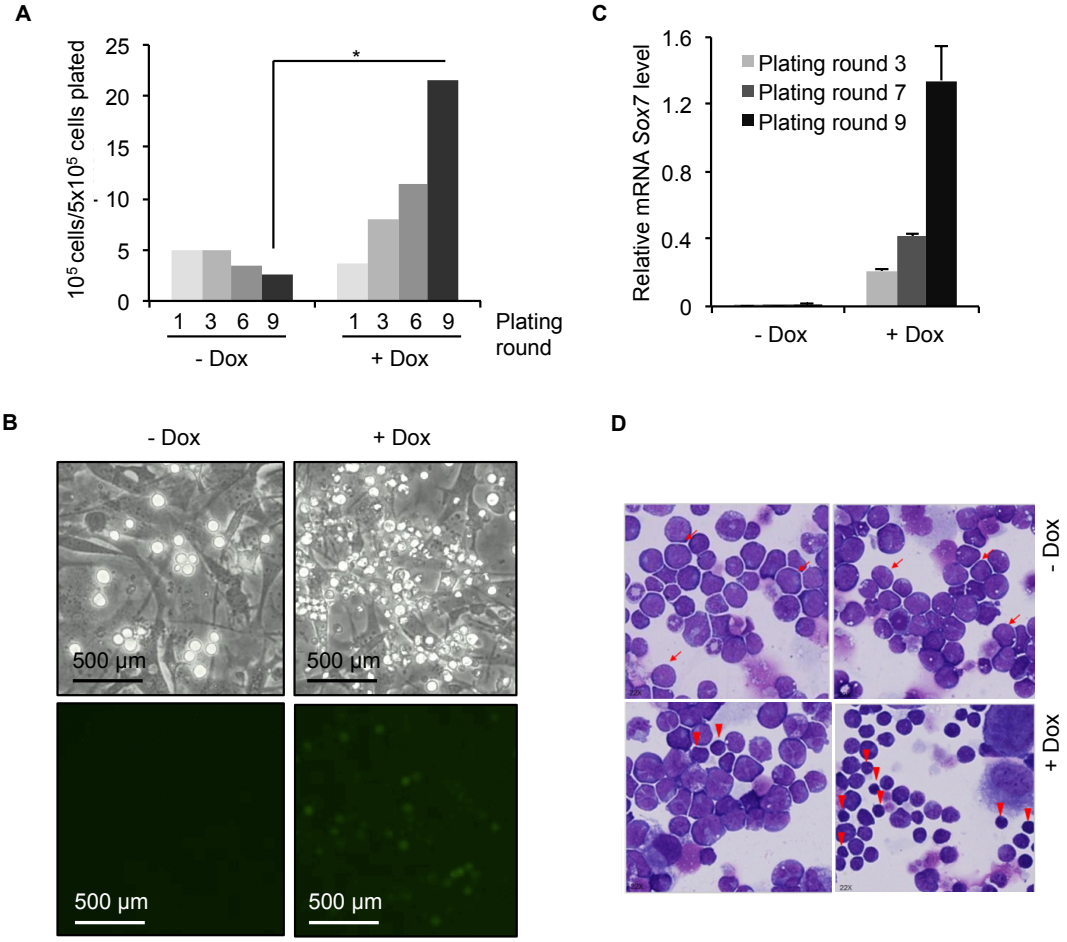


Figure 5.3. *Sox7*-enforced expression induces cell proliferation and impairs B cell differentiation.
A) The *iSox7* bone marrow cells were plated with or without doxycycline on irradiated OP9 culture for serial replating assay. The bar chart shows the cell number for each plating rounds (n=3; **P*<0.01). B) Representative pictures of *iSox7* bone marrow cells were plated with or without doxycycline on irradiated OP9 culture at plating round 7 (Leica scan and software, scale bar: 500 µm). C) qRT-PCR analysis of *Sox7* transcript levels relative to βactin in *iSox7* bone marrow cells cultured with or without doxycycline. Error bars indicate mean ± standard deviation (n=3). D) May-Grünwald-Giemsa staining of *iSox7* bone marrow cells cultured with or without doxycycline (Leica and scan software, 22X magnification). Top panels (from left to right): cells cultured without doxycycline at plating round 2 and 5. Bottom panels (from left to right): cells cultured with doxycycline at plating round 5 and 10. Red arrow indicates B lymphocytes, head arrow indicates immature blast-like cells. Graphs are representative of 4 independent experiments.

To define whether *Sox7* expression was essential for the maintenance of the observed phenotype once initiated, doxycycline was removed from the culture and cells were counted after three or four days. As shown in the figure 5.4, *Sox7* expression was essential to maintain cell growth since removal of doxycycline dramatically reduced the proliferative potential and accumulation of cells within the culture. These data suggests that maintenance of the high proliferative potential requires the persistent expression of *Sox7*.

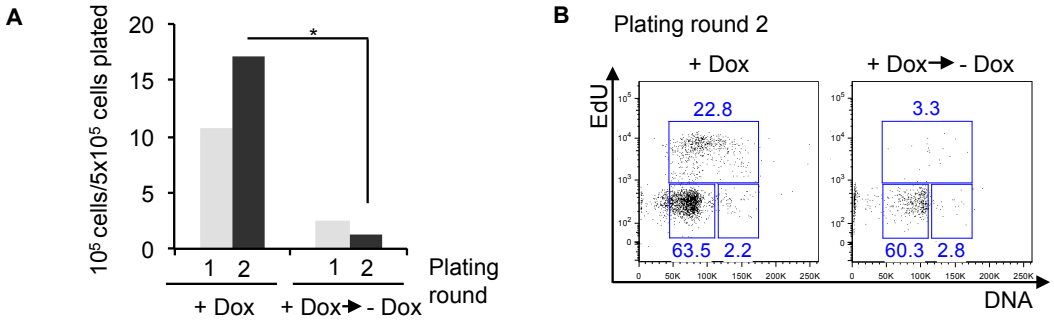


Figure 5.4. *Sox7* expression was essential to maintain the cell growth.
A) After the plating round 9, *iSox7*$^+$ bone marrow cells were cultured on irradiated OP9 with or without doxycycline. The bar chart shows the cell number for each plating rounds (n=3, **P*<0.01). B) Cell cycle status of *iSox7*$^+$ bone marrow cells cultured with or without doxycycline was assessed after 4 hrs of EdU incubation. Percentage of cells in the G0/G1, S and G2 phases are shown in the indicated gates. Graph and FACS plots are representative of two independent experiments.

Next, the immuno-phenotype of *Sox7*-expressing cells was investigated. During the serial replating, the increase in proliferation was paralleled with an increase in the frequency of SOX7::GFP-expressing cells as shown in figure 5.5. The B cell markers, B220 and CD19 and the progenitor marker, cKIT, were included in the flow cytometry analysis. The expression of VE-cadherin, a direct target of SOX7, was also analysed in order to further validate the functionality of the induced protein (Costa et al., 2012).

The FACS analysis revealed two different sub-populations based on their level

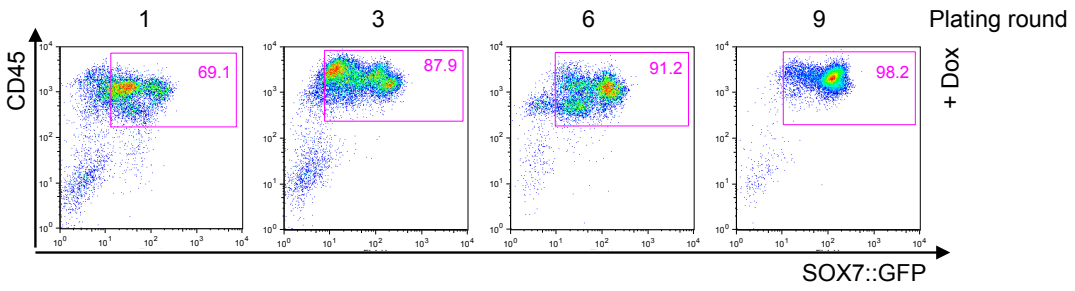


Figure 5.5. The increase in proliferation correlates with an increase in the frequency of SOX7::GFP-expressing cells.
Representative FACS plots of CD45 and SOX7::GFP expression profiles of *iSox7* bone marrow cells cultured with doxycycline. Percentages of cells are shown in the indicated gates. FACS plots are representative of 4 independent experiments.

of SOX7::GFP expression. In the SOX7::GFP^{low} sub-population, most cells were cKIT$^{-/low}$, VE-cadherin$^{-/low}$, B220^{+} and CD19^{+} while the SOX7::GFP^{high} sub-population contained mostly cKIT^{high}, VE-cadherin^{high}, B220^{-} and CD19^{-} cells (Figure 5.6 A). After 11 passages, most of the cells were within the SOX7::GFP^{high} sub-population and were enriched in progenitor cells expressing high level of cKIT and VE-cadherin (Figure 5.6 A bottom panels). In this analysis I also observed a progressive increase in the frequency of cKIT^{+}SOX7::GFP^{high} sub-population upon serial replatings suggesting that a cKIT^{+}SOX7::GFP^{high} blast-like population may maintain the proliferative potential within those cultures (Figure 5.6 B).

To better characterize at which stage of B cell differentiation the B220^{+} SOX7::GFP^{low} were blocked, a thorough analysis of the expression of B cell differentiation markers was performed. In particular, early B cells can be sub-divided into two sub-populations: less mature cells express B220 and CD43 while the more mature express only B220 (Li et al., 1993). Then, both sub-populations can be sub-divided into three fractions. The less mature one distinguishes, based on the expression of BP-1 and CD24, the three fractions: pre-pro B, pro B and late-pro B; whereas the

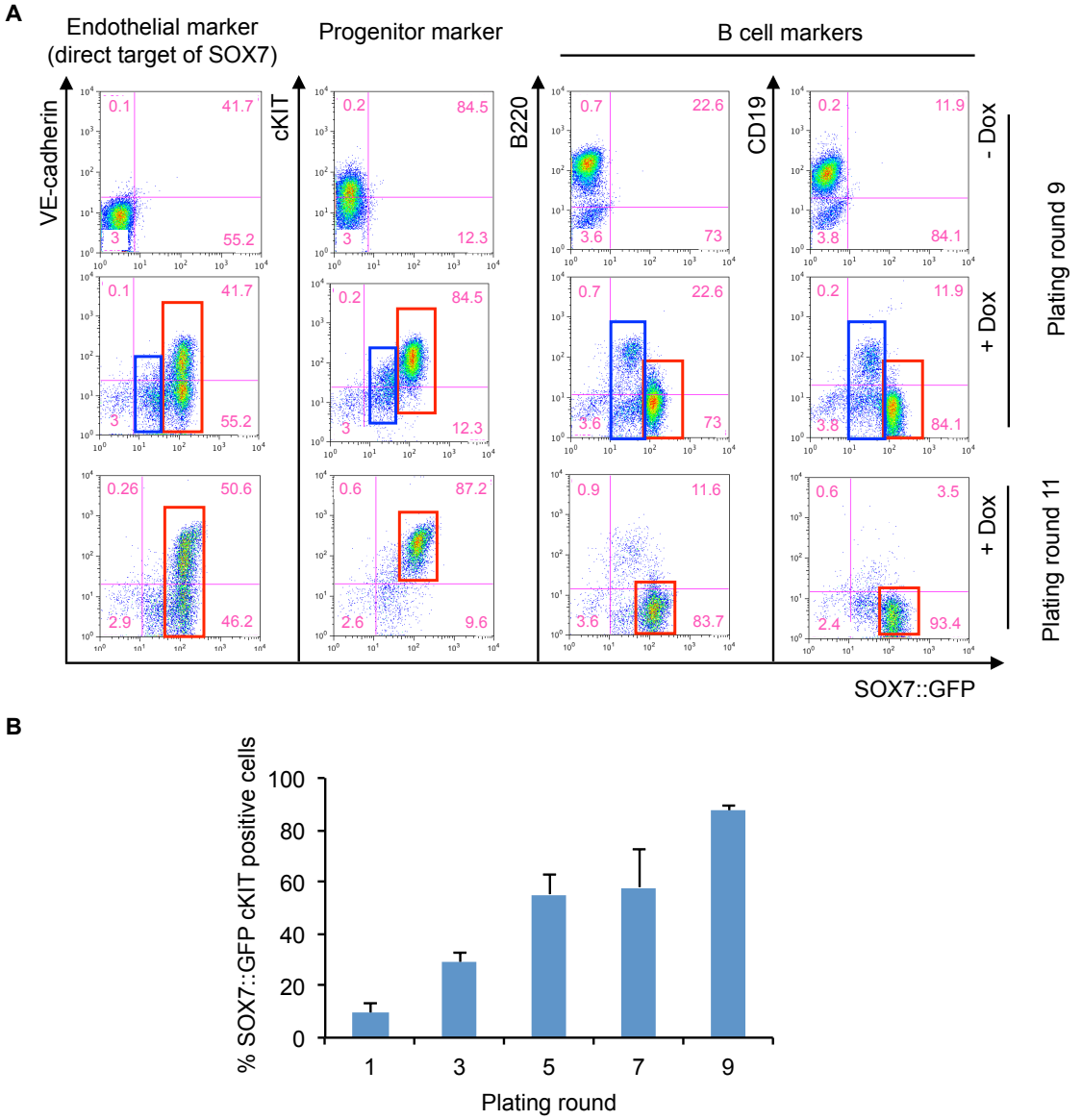


Figure 5.6. Characterization of *Sox7*-expressing cells cultured on OP9.
Representative FACS plots of VE-cadherin, B220, CD19, cKIT and SOX7::GFP expression profiles of *iSox7* bone marrow cells cultured with and without doxycycline at plating round 9 and 11. Percentages of cells are shown in the indicated gates. SOX7::GFP^{low} and SOX7::GFP^{high} population are gated in blue and red, respectively. FACS plots are representative of three independent experiments. B) Percentage of SOX7::GFP-cKIT positive cells are shown in the bar chart. Error bars indicated mean ±standard deviation (n=3). Graph and FACS plots are representative of three independent experiments.

more mature one identifies small pre B, immature B and mature B fractions based on the expression of IgM and IgD (Hardy and Hayakawa, 1991) (Figure 5.7 A-B). The analysis showed the absence of BP-1 and CD24 expression in *Sox7*-over-expressing $B220^{+}CD43^{+}$ cells suggesting that these cells were blocked at a pre-pro B differentiation stage or even earlier within the progenitor compartment (Figure 5.7 C).

Given the presence of blast-like cells as shown in figure 5.5, a detailed analysis of the expression of haematopoietic stem cell markers was performed. This analysis revealed an increase in the frequency of the $Lineage^{-}SCA1^{+}cKIT^{+}$ (LSK) fraction upon serial replatings (Figure 5.8 A) suggesting that *Sox7*-enforced expression may induce the maintenance and expansion of the LSK compartment which contains stem cells and early progenitors including HSCs, LMPP, CLP and CMP (Figure 5.8 B).

Altogether these findings reveal that the enforced expression of *Sox7* in bone marrow cells enhances the proliferation of an immature population while blocking B lineage maturation.

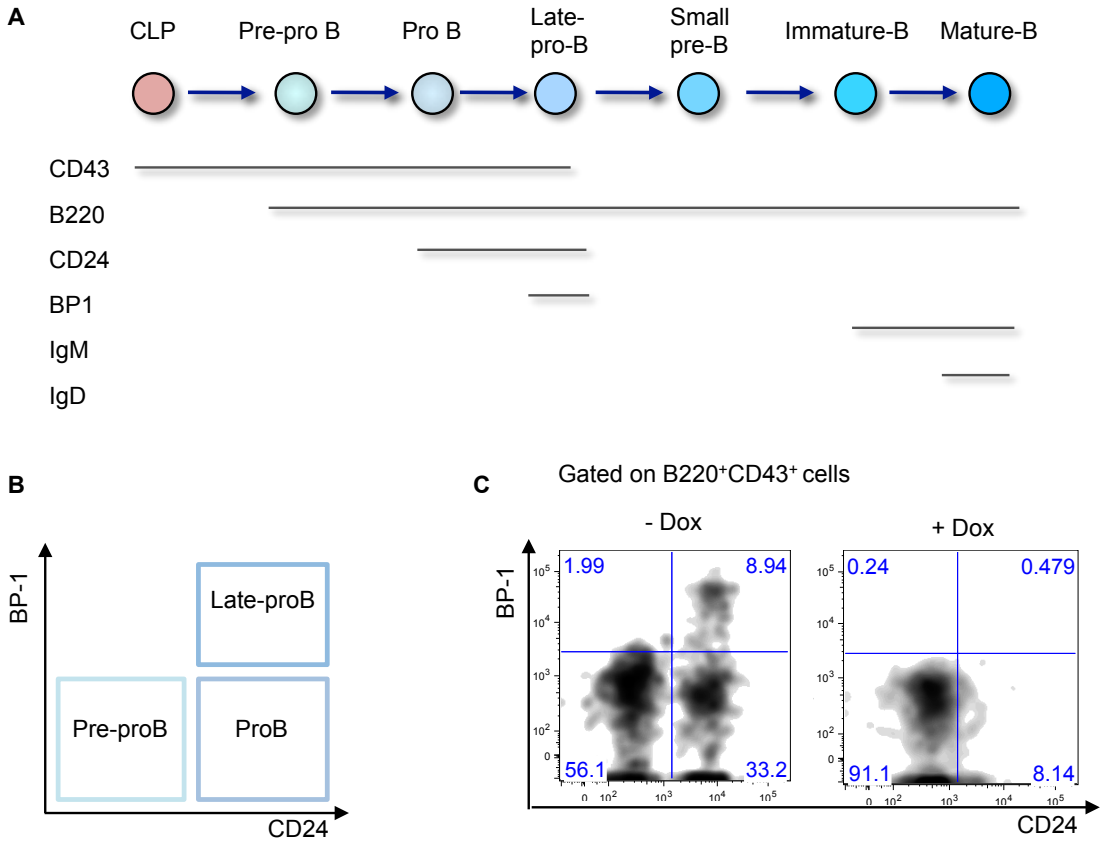


Figure 5.7. *Sox7*-enforced expression blocks bone marrow cells at a pre-pro B differentiation stage.
A) Schematic diagram of B cell differentiation. B) Schematic representation of the B cell sub-populations. C) Representative FACS plots of BP-1 and CD24 expression profiles of *iSox7* bone marrow cells cultured with and without doxycycline. BP-1 CD24 plots were gated on $B220^{+}CD43^{+}$ cells. Percentages of cells are shown in the indicated quadrants. FACS plots are representative of four independent experiments.

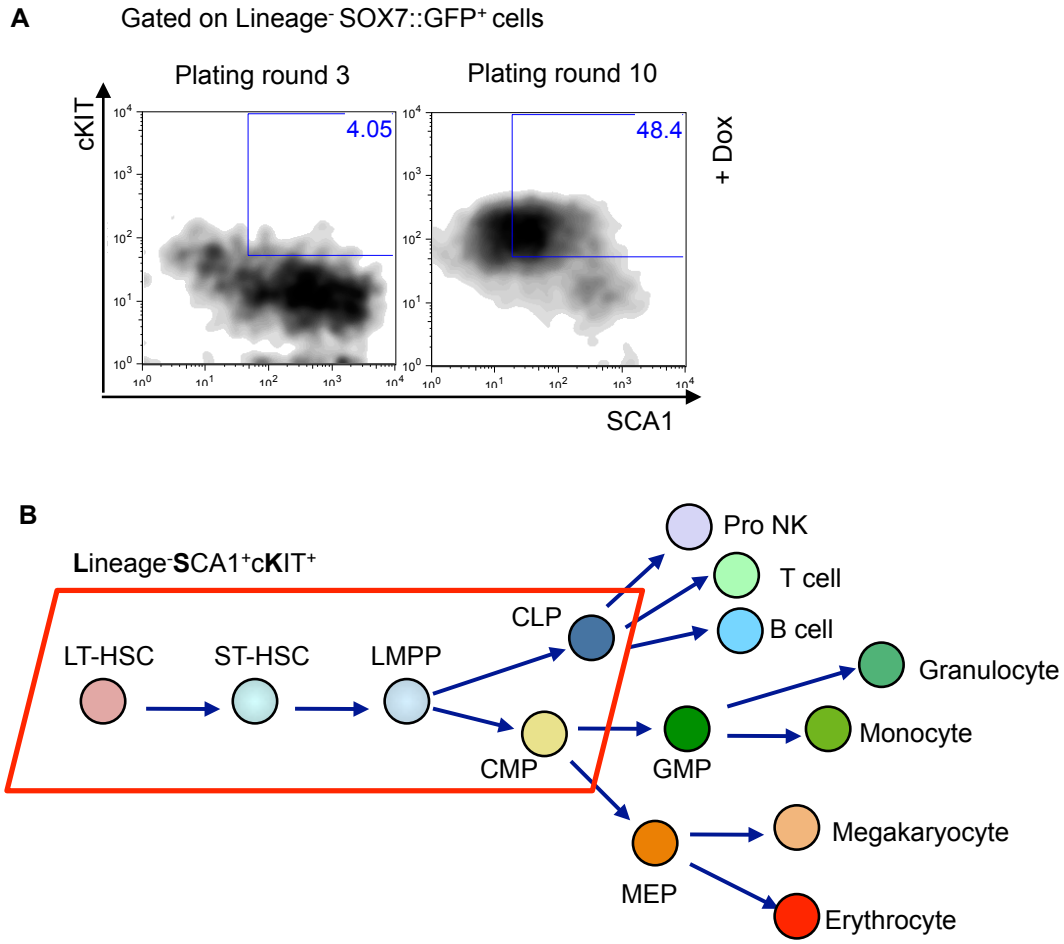


Figure 5.8. *Sox7*-enforced expression induces the expansion of a LSK subset.
A) Representative FACS plots of cKIT and SCA1 expression profiles of *iSox7*$^{+}$ bone marrow cells cultured with doxycycline at the plating round 3 and 10. cKIT SCA1 plots were gated on Lineage^{-} SOX7::GFP^{+} cells. Percentages of cells are shown in the indicated gates. FACS plots are representative of four independent experiments. B) Schematic diagram of haematopoietic hierarchy. LSK compartment contains HSC, LMPP, CMP and CLP populations.

5.2.2. The cKIT^{+}SOX7::GFP^{high} sub-population contains both myeloid and lymphoid potential

Next, we investigated the biological potential of the cKIT^{+}SOX7::GFP^{+} blast-like population and tested whether these cells were able to differentiate into lymphoid, erythroid and myeloid lineages in the absence of doxycycline. For this purpose, cKIT^{+}SOX7::GFP^{+} cells were sorted and plated in semi-solid culture to test their myeloid and erythroid potential and on irradiated OP9 or OP9-DL1 to test their B and T lymphoid potential, respectively (Figure 5.9 A). Analysis of the morphology of the cKIT^{+}SOX7::GFP^{+} sorted cells confirmed their blast-like phenotype as observed by May-Grünwald Giemsa staining (Figure 5.9 B).

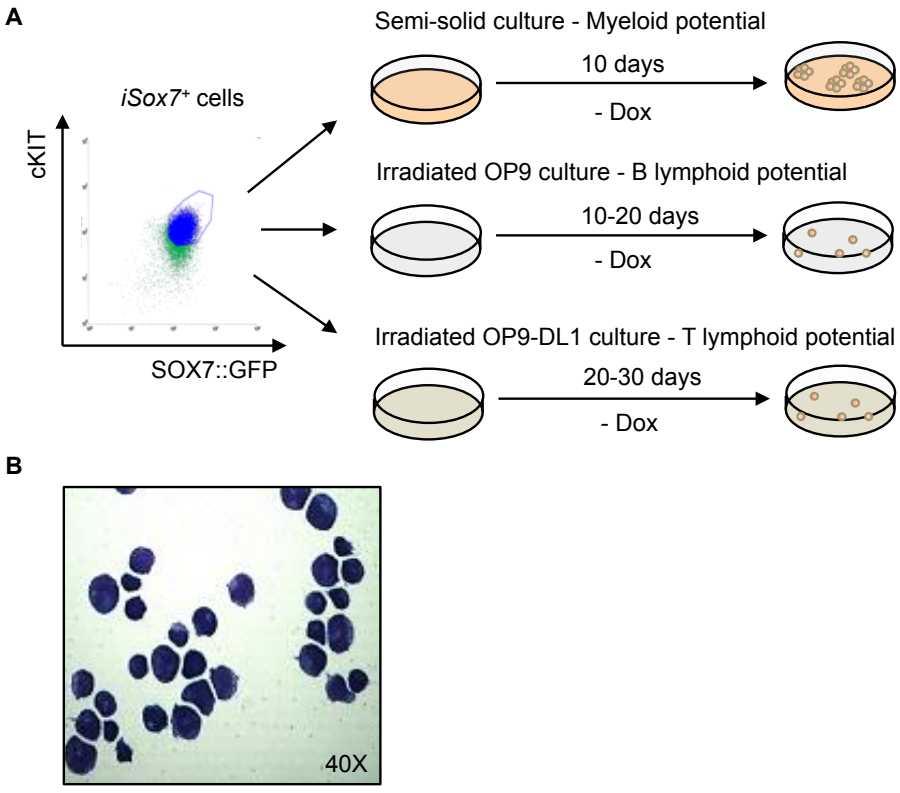


Figure 5.9. Scheme of the protocol used to assess the myeloid and lymphoid potential of SOX7::GFP^{high}cKIT^{+} cells.
A) Representative FACS plot of cKIT and SOX7::GFP expression profile. Sorted SOX7::GFP^{high}cKIT^{+} cells were cultured on semi-solid culture to assess the myeloid potential while on irradiated OP9/OP9-DL1 cells to assess the lymphoid potential. B) May-Grünwald-Giemsa staining of sorted SOX7::GFP^{high}cKIT^{+} cells (Leica scan software, 40X magnification).

As shown in Figure 5.10 A and B, cKIT$^+$SOX7::GFP$^+$ cells were able to generate erythroid colonies and all types of myeloid colonies such as megakaryocyte, granulocyte, mast and macrophage colonies. Gene expression analysis of three myeloid markers (*Gfi1, C/ebpε* and *Irf8*) was performed to further confirm myeloid identity (Figure 5.10 C). When plated on irradiated OP9, cKIT$^+$SOX7::GFP$^+$ cells were also able to generate B220$^+$CD19$^+$ B lymphocytes after about 20 days in culture (Figure 5.11 A); the expression of two lymphoid transcription factors, *E2a* and *Pax5*, further confirmed B cell identity (Figure 5.11 B). To test T lymphoid potential, cKIT$^+$SOX7::GFP$^+$ blast-like cells were plated on irradiated OP9-DL1 and cultured for about 10 days in the presence of high IL-7 concentration. Successively, cells were transferred to low IL-7 concentration medium for the acquisition of the mature T cell phenotype. The expression of the immature T cell marker, CD25, appeared after 10 days in culture suggesting that cKIT$^+$SOX7::GFP$^+$ cells were also able to differentiate toward T cell lineage and generate T lymphocytes at the DN2 or DN3 stage (Figure 5.11 C-D).

Next, in order to test whether the cKIT$^+$SOX7::GFP$^+$ population retained the ability to engraft *in vivo*, one million cells were injected into sub-lethally irradiated immuno-compromised recipient mice not fed with doxycycline. Every four weeks post engraftment, blood sample analysis was performed to assess for donor contribution. The detection of a small CD45.2$^+$ donor cell population was observed overtime indicating successful engraftment (Figure 5.12 A). However, a progressive decline in the contribution of donor cells to less than 1 per cent by 16 weeks was observed (Figure 5.12 B), suggesting that either the self-renewing ability of cKIT$^+$SOX7::GFP$^+$ progenitor cells was very limited *in vivo* or that the cKIT$^+$SOX7::GFP$^+$ population contained mainly short-term progenitors.

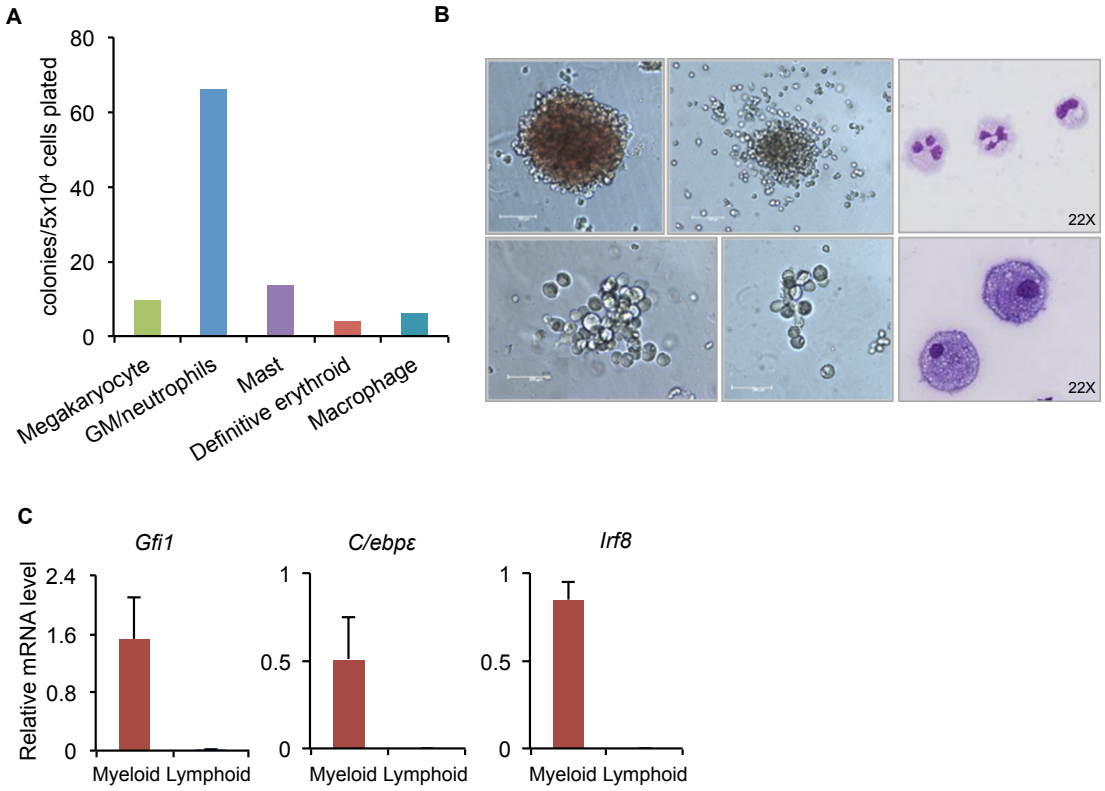


Figure 5.10. SOX7::GFP^{high}cKIT^{+} cells retain myeloid potential.
A) SOX7::GFP^{high}cKIT^{+} cells were sorted and cultured in semi-solid culture without doxycycline for colony assay. The bar chart shows the number of colonies for each colony types (n=3). B) Representative pictures of colonies obtained in colony assay and May-Grünwald-Giemsa staining of cells derived from colonies (Leica scan and software, 22X magnification, scale bar: 500 µm. Top panels (from left to right): erythroid colony, mast colony and granulocytes. Bottom panels (from left to right): macrophage colony, megakaryocyte colony and macrophages. C) qRT-PCR analysis of *Gfi1*,*C/ebpε* and *Irf8* transcript levels relative to β-actin in cells cultured in myeloid and lymphoid conditions. Error bars indicate mean ± standard deviation (n=3). Graphs are representative of three independent experiments.

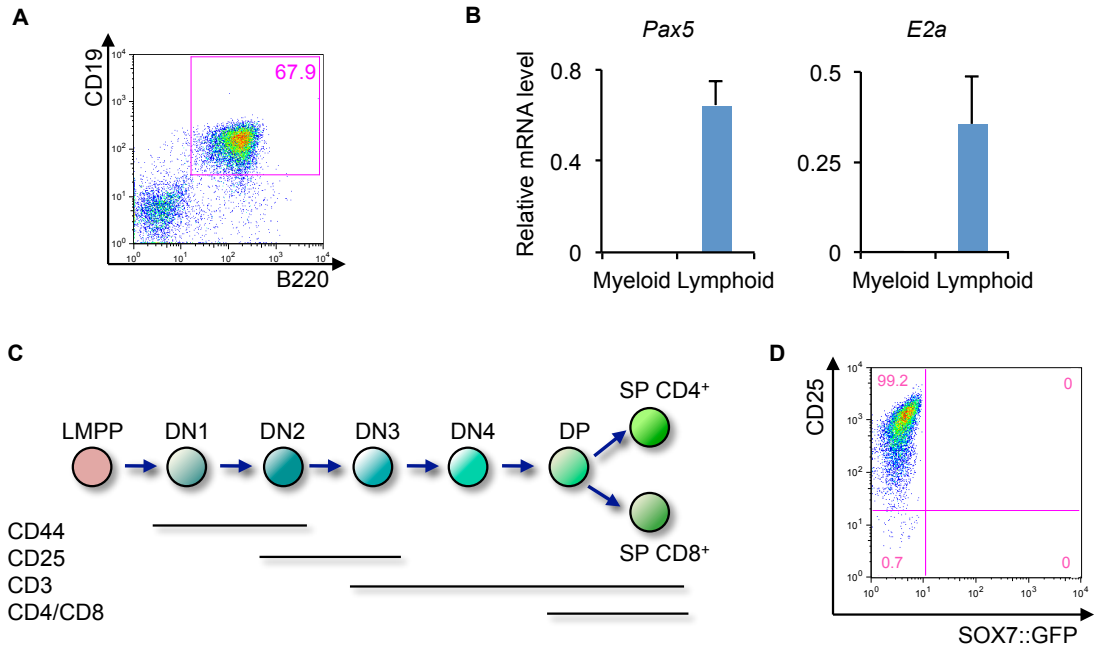


Figure 5.11. SOX7::GFP^{high}cKIT^{+} cells retain lymphoid potential.
A) Representative FACS plots of CD19 and B220 expression profiles of SOX7::GFP^{high}cKIT^{+} cells cultured without doxycycline. Percentages of cells are shown in the indicated quadrant. B) qRT-PCR analysis of *Pax5* and *E2a* transcript levels relative to β-actin in cells cultured in myeloid and lymphoid conditions. Error bars indicate mean ± standard deviation (n=3). Graphs and FACS plot are representative of three independent experiments. C) Schematic representation of T cell development. DN refers to double negative, DP to double positive and SP to single positive. D) Representative FACS plots of CD25 and SOX7::GFP expression profiles of SOX7::GFP^{high}cKIT^{+}cells cultured on OP9-DL1 cells without doxycycline. Percentages of cells are shown in the indicated quadrants. FACS plot are representative of two independent experiments

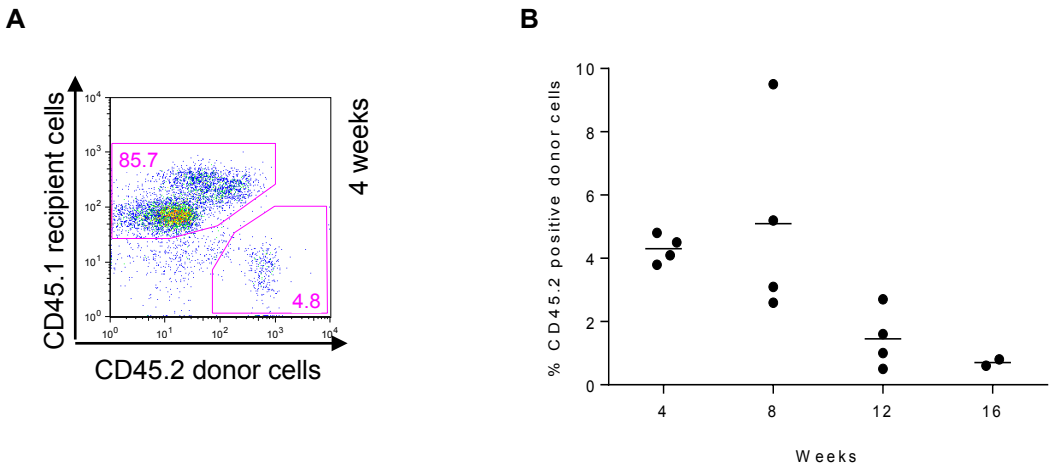


Figure 5.12. SOX7::GFP^{high}cKIT^{+} cells retain the ability to engraft *in vivo*.
A) Representative FACS plots of CD45.1 and CD45.2 expression profiles of blood samples from mice injected with SOX7::GFP^{high}cKIT^{+} cells. Percentages of cells are shown in the indicated quadrants. B) Scatter chart of the CD45.2^{+} donor cell percentage from 4 to 16 weeks after injection. Error bars indicate mean ± standard deviation. Graph is representative of two independent experiments.

In parallel, the ability to induce haematological malignancy was also evaluated by injecting cKIT^{+}SOX7::GFP^{+} cells into immuno-compromised mice fed with doxycycline. Cells successfully engrafted *in vivo* showing low short and long-term engraftment ability. The frequency of donor cell contribution was slightly lower than in mice not fed with doxycycline (Figure 5.13 A). Interestingly, the frequency of CD45.2^{+} cells dramatically increased sixteen weeks post-engraftment in one out of four mice (Figure 5.13 B), suggesting the emergence of a disease or the maintenance of self-renewing HSCs at a low frequency upon doxycycline feeding. However, to confirm this observation, further experiments will need to be performed. Analysis of the lymphoid and myeloid compartment in the mouse expressing high frequency of CD45.2 expression was performed in order to verify any haematological imbalance upon over-expression of *Sox7 in vivo*. In general, cKIT^{+}SOX7::GFP^{+} cells retained multi-lineage engraftment ability but no haematological malignancy was observed at 32 weeks post-

engraftment (Figure 5.14).

Altogether, these results suggest that cKIT^{+}SOX7::GFP^{+} cells retain both myeloid and lymphoid potential and *in vivo* repopulating ability.

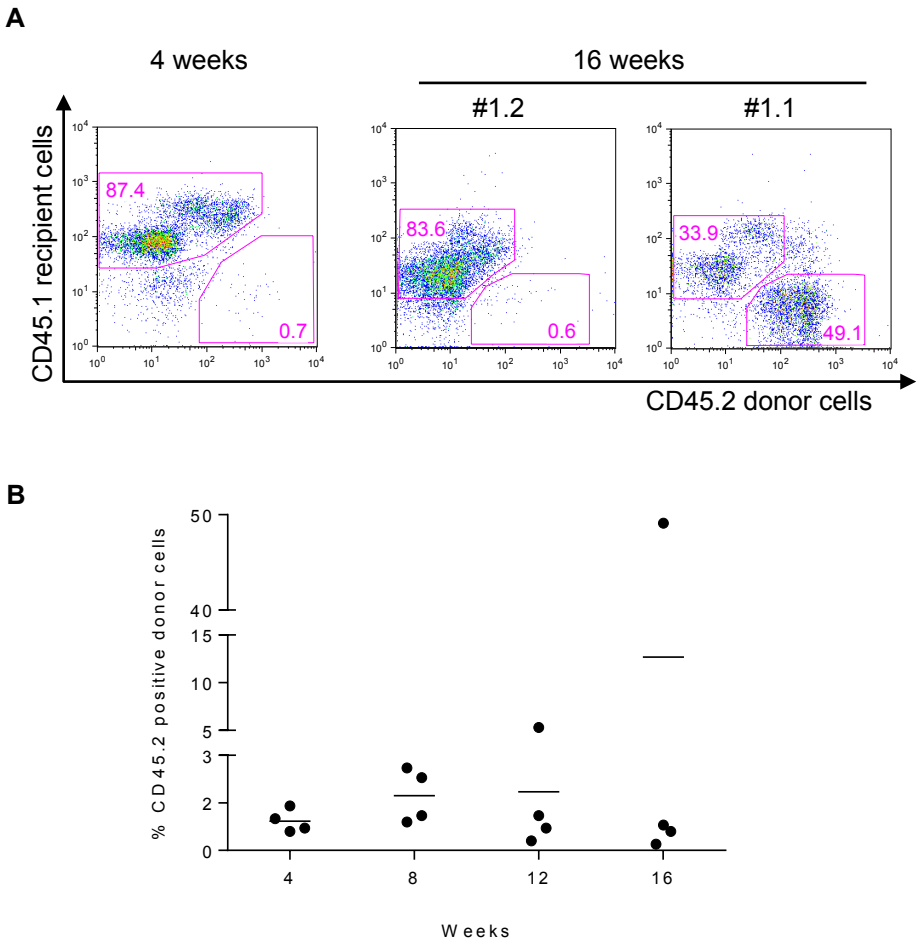


Figure 5.13. SOX7::GFP^{high}cKIT^{+} cells do not develop any type of malignancies *in vivo*.
A) Representative FACS plots of CD45.1 and CD45.2 expression profiles of blood samples from mice injected with SOX7::GFP^{high}cKIT^{+} cells and fed with doxycycline. Percentages of cells are shown in the indicated quadrants. B) Scatter chart of the CD45.2^{+} donor cell percentage from 4 to 16 weeks after injection. Graph and FACS plots are representative of two independent experiments.

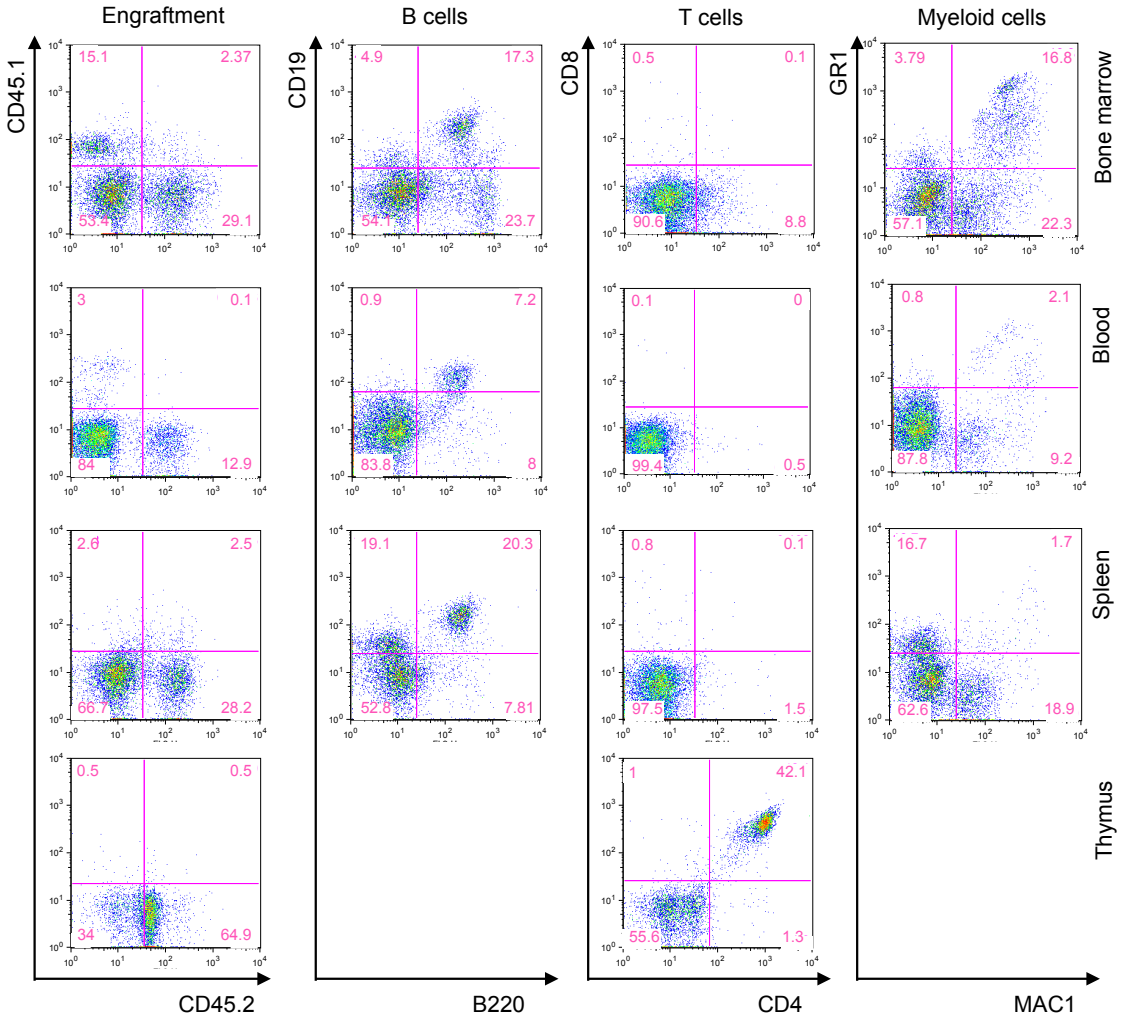


Figure 5.14. SOX7::GFP^{high}cKIT^{+} cells retain a poor multi-lineage engraftment ability *in vivo*.

Representative FACS plots of CD45.1, CD45.2, CD19, B220, CD8, CD4, GR1 and MAC1 expression profiles of bone marrow, blood, spleen and thymus samples from mice injected with SOX7::GFP^{high}cKIT^{+} cells. Analysis was performed on organs from the mouse with high frequency of CD45.2^{+} donor cells at 32 weeks post-engraftment. Percentages of cells are shown in the indicated quadrants. FACS plots are representative of two independent experiments.

5.2.3. *Sox7*^{high} population is enriched for stem cell genes

To better characterize the impact of *Sox7*-enforced expression at the molecular level, global gene expression profilings were performed. Two populations were sorted based on their level of SOX7::GFP expression, SOX7::GFP^{low} and SOX7::GFP^{high} (Figure 5.15 A) with two biological replicates used for hybridization on mouse exon arrays. Multidimensional scaling (MDS) plot was used to check the concordance between replicates and divergence between the two populations. As shown in figure 5.15 B, the two biological replicates were relatively close to each other. Genes that changed by at least 2-fold between SOX7::GFP^{low} and SOX7::GFP^{high} populations with a FDR< 5% were considered significant (Figure 5.15 C). The gene expression profile analysis showed 1164 significantly differentially expressed genes. Of these, the expression of 642 genes was increased and the expression of 522 genes was decreased in the SOX7::GFP^{high} relative to SOX7::GFP^{low} population (Figure 5.15 D). Genes significantly increased upon high expression of *Sox7* included stem cell genes such as *Gata2, Flt3* and *Cd27* (Table 5.1). In contrast, genes significantly decreased upon high expression of *Sox7* contained lymphocyte-associated genes such as *Rag1/2, Ebf1, Pax5* and *Cd19* (Table 5.2). These data were further confirmed by gene set enrichment analysis (GSEA) (Mootha et al., 2003; Subramanian et al., 2005) performed on all differentially expressed genes. GSEA clearly demonstrated a significant enrichment of haematopoietic stem cell genes (Jaatinen and Laine, 2007; Liang et al., 2007) in the SOX7::GFP^{high} population while a significant enrichment of B-lymphocyte-associated genes (Haddad et al., 2004; Mori et al., 2008) in the SOX7::GFP^{low} population (Figure 5.16).

In parallel, I performed a canonical pathway analysis using the Ingenuity Pathway Analysis IPA software. This analysis showed that most of pathways involved

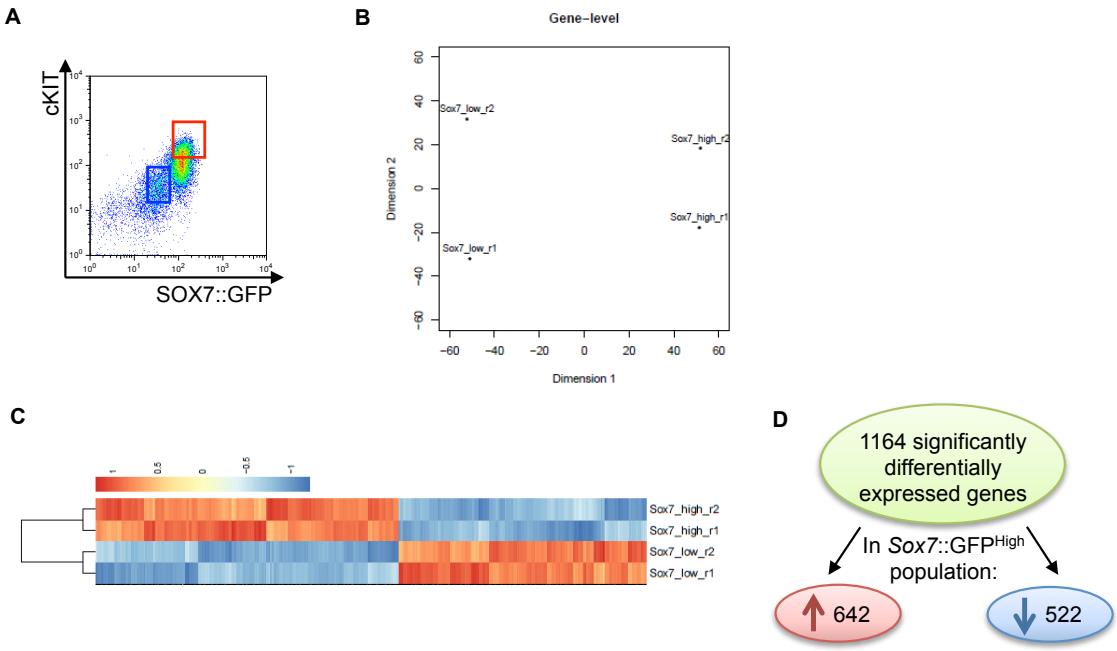


Figure 5.15. Sort and microarray analysis of *Sox7*-expressing cells.
A) Representative FACS plots of cKIT and SOX7::GFP expression profiles of *iSox7* bone marrow cells cultured with doxycycline. The indicated gates show the two sorted populations used for microarray analysis. The *Sox7*^{high} population is gated in red while the *Sox7*^{low} population is in blue. B) Multidimensional scaling (MDS) plot shows the separation between treatment groups. C) Heatmap of differentially expressed genes. Genes that changed by at least 2-fold between SOX7::GFP^{low} and SOX7::GFP^{high} populations with FDR<5% were considered significant. D) For gene-level analysis, 1164 genes were significantly differentially expressed. Of these, 642 were highly expressed and 522 were poorly expressed in SOX7::GFP^{high} relative to SOX7::GFP^{low} populations.

A

Genes highly expressed in SOX7::GFP^{high} population:

Gene	P-value	Log fold change
CDH5	0.00052	2.776087143
GATA2	7.03E-05	2.7744595
SMAD1	0.00139	1.70400375
CD27	6.33E-05	2.237015
FLT3	0.00247	2.250735714
RUNX2	1.16E-05	3.201388864

B

Genes poorly expressed in SOX7::GFP^{high} population:

Gene	P-value	Log fold change
RAG1	0.00027	-4.851588571
RAG2	0.00058	-4.238242
PAX5	5.09E-06	-3.561190556
IL7R	0.00031	-3.2314295
EBF1	0.00022	-4.178694444
CD79A	0.00058	-4.229631
CD79B	0.00058	-4.456061875
CD19	0.00058	-3.927118889
BLK	0.00058	-2.370453333
BCL2L1	0.00058	-1.598764583

Table 5.1-5.2. List of genes highly and poorly expressed in SOX7::GFP^{high} population. A) List of genes highly expressed in SOX7::GFP^{high} relative to SOX7::GFP^{low} populations. B) List of genes poorly expressed in SOX7::GFP^{high} relative to SOX7::GFP^{low} populations. P-value and log fold change are shown in the tables.

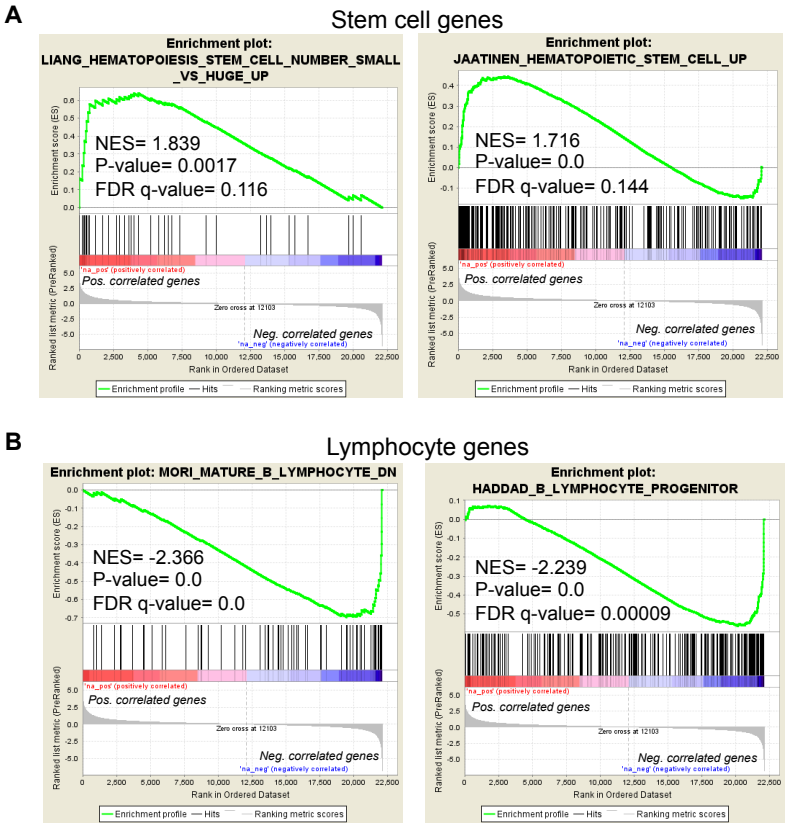


Figure 5.16. SOX7::GFP^{high} population is enriched for stem cell genes.
A) Gene set enrichment analysis demonstrates significant enrichment of hematopoietic stem cell genes in SOX7::GFP^{high} population. B) Gene set enrichment analysis demonstrates significant enrichment of B lymphocyte genes in SOX7::GFP^{low} population.

in B cell development were indeed affected by the enforced expression of *Sox7* (Figure 5.17). Overall, many genes involved in B cell development were poorly expressed in SOX7::GFP^{high} population (Figure 5.18). Moreover, the Ingenuity disease and function analysis showed that most genes highly expressed in SOX7::GFP^{high} population were in the haematopoiesis and haematological development categories. Interestingly, several genes belonged to the cellular growth or cell death and survival categories as highlighted in table 5.3 (Figure 5.19), findings correlating with the high proliferation potential of SOX7::GFP^{high} cells observed during the *in vitro* culture.

Taken together, these data further extend the results obtained *in vitro* showing that the enforced expression of *Sox7* in adult bone marrow cells has a profound impact on B cell maturation while enhancing a stem cell signature.

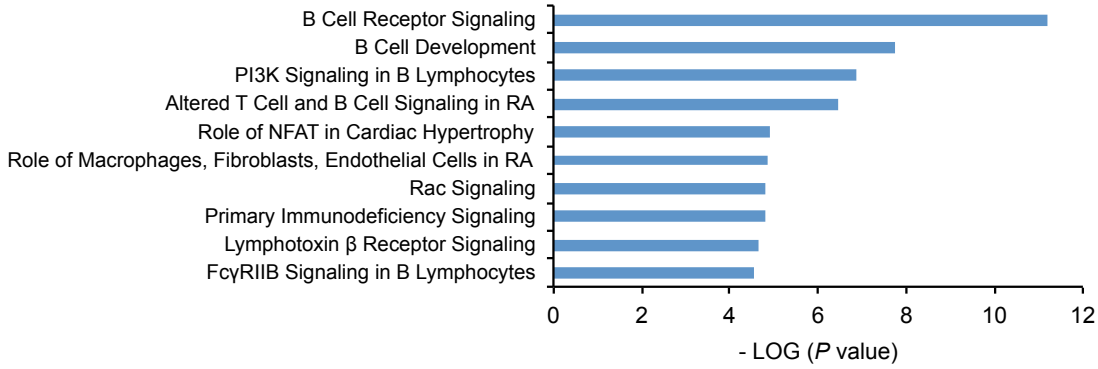


Figure 5.17. Pathways involved in B cell development are affected by the enforced expression of *Sox7*.
Ingenuity canonical pathways analysis of genes poorly expressed in SOX7::GFP^{high} population are shown in the graph. The bar chart shows the first ten pathways (RA: Rheumatoid Arthritis).

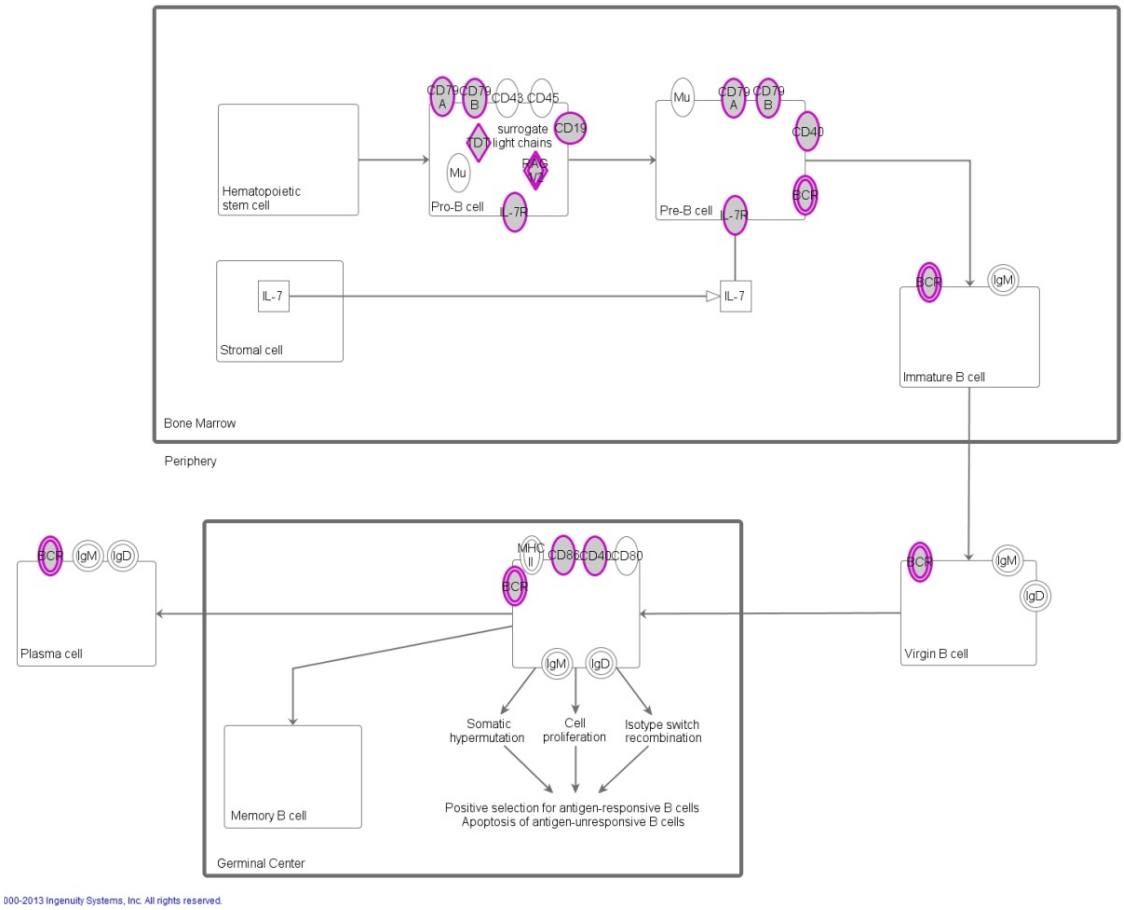


Figure 5.18. Genes involved in the B cell development are poorly expressed in SOX7::GFP^{high} population.
Ingenuity canonical pathways analysis of genes involved in the B cell development poorly expressed in SOX7::GFP^{high} population. Genes highlighted in purple are down-regulated in the SOX7::GFP^{high} population.

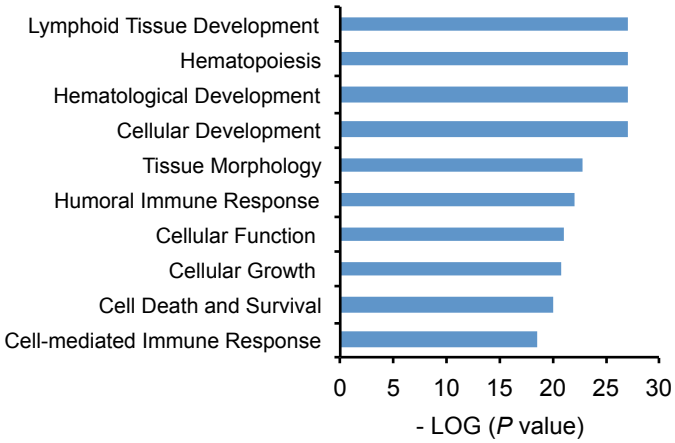


Figure 5.19. Genes highly expressed in SOX7::GFP^{high} population fall in haematopoiesis and cellular growth categories.
Ingenuity disease and function analysis of genes highly expressed in SOX7::GFP^{high} population is shown in the bar graph. The bar chart shows the first ten disease and functions.

Proliferation Genes	Log fold change	P-value
AHR	1.690985714	0.000341
CASP3	1.02398125	0.002236
CD27	2.237015	6.33E-05
CD28	1.716735	0.000613
ETS2	3.605361	1.13E-05
FLT3	2.250735714	0.002473
GATA2	2.7744595	7.03E-05
GFI1B	1.47926125	0.0008
IGF2	1.086665357	0.001961
IKZF2	2.819884667	4.97E-05
IL6	2.767303	0.001091
IRF6	1.582841111	0.000403
MYCN	1.531179286	0.000402
PECAM1	1.305463261	0.00122
TCF7	4.444998125	2.71E-05

Table 5.3. List of genes involved in cellular growth highly expressed in SOX7::GFP^{high} relative to SOX7::GFP^{low} population

5.3. Summary of findings

Since *SOX7* is mainly expressed in B-ALL, in this part of the project I investigated the consequences of *Sox7*-enforced expression on adult B cell differentiation. The data described in the present chapter reveal for the first time that the enforced expression of *Sox7* in adult bone marrow cells induces proliferation while blocking B lineage maturation. The immuno-phenotype analysis revealed that SOX7::GFP^{high} cells expressed high level of VE-cadherin and cKIT while they were negative for the expression of the B cell markers, B220 and CD19. Moreover, I observed a progressive increased in the cKIT^{+}SOX7::GFP^{high} population, suggesting that this blast-like population may maintain the proliferative potential of *iSox7* bone marrow cells.

Further analysis revealed that the SOX7::GFP^{low} cells were blocked at a pre-pro B differentiation stage or even earlier in the progenitor compartment while SOX7::GFP^{high} cells showed an LSK immuno-phenotype. These *in vitro* cellular data were complemented by a molecular profiling analysis, which revealed that the enforced expression of *Sox7* in adult bone marrow cells globally impaired B cell maturation while enhancing a stem cell signature.

Since an expansion of immature progenitor cells was observed, their ability to differentiate into lymphoid and myeloid lineages in the absence of doxycycline was investigated. cKIT^{+}SOX7::GFP^{+} cells retained erythroid, myeloid and lymphoid potential *in vitro* as well as repopulating capacity *in vivo*.

Finally, to assess *in vivo* whether *Sox7* expression may confer a leukaemic potential, cKIT^{+}SOX7::GFP^{+} cells derived from serial replating were injected into NSG mice fed with doxycycline to maintain the over-expression of *Sox7*. The cKIT^{+}SOX7::GFP^{+} cells showed repopulating ability but no haematological disease was observed.

Chapter 6

6. Analysis of *Sox7*-enforced expression *in vivo*

6.1. Introduction

The analysis of *Sox7*-enforced expression in adult bone marrow cells, performed in semi-solid culture in the presence of myeloid cytokines, showed that the mis-expression of *Sox7* in adult cells confers a proliferative advantage while blocking lineage maturation as it was previously observed in embryonic haematopoiesis (Gandillet et al., 2009).

In addition, I also reported the over-expression of *SOX7* in multiple cases of ALL, in particular in B-ALL. Therefore, I investigated how *Sox7*-enforced expression might affect adult B cell differentiation, possibly leading to haematological malignancy. The data, presented in the previous chapter, revealed that the enforced expression of *Sox7* in adult bone marrow cells induces enhanced proliferation while blocking B lineage maturation. Moreover, an increased in SOX7::GFP^{high}cKIT^{+} population was observed suggesting that this blast-like population may maintain the proliferative potential of *iSox7* bone marrow cells as observed in the semi-solid culture experiments. These cells also retained short-term repopulating capacity *in vivo*.

Given these facts, this part of my project has investigated how enforced expression of *Sox7* affects haematopoiesis *in vivo*. First, I evaluated whether *Sox7*-enforced expression impaired adult B cell differentiation. Secondly, I defined which

haematopoietic subsets were responsive to *Sox7*-enforced expression and finally, I analysed whether *Sox7* over-expression conferred a leukaemic potential *in vivo*.

6.2. Results

6.2.1. *Sox7*-enforced expression induces B cell impairment *in vivo*

In order to establish a model system in which *Sox7*-enforced expression is only inducible in the haematopoietic system, total bone marrow obtained from *iSox7* mouse was injected into sub-lethally irradiated immuno-compromised mice. After re-establishment of the haematopoietic system, engrafted mice were fed or not with doxycycline, and blood, bone marrow, spleen, liver and thymus were monitored over 12-month period (Figure 6.1). Analysis of blood samples for the expression of CD45.2 present on donor cells indicated a full reconstitution of haematopoietic system in both experimental groups, with and without doxycycline at two weeks post-engraftment (Figure 6.2).

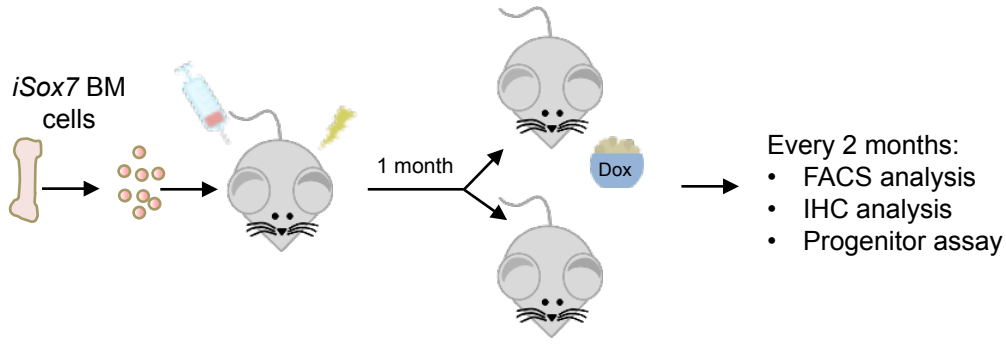


Figure 6.1. Scheme of the protocol used to test the over-expression of *Sox7 in vivo*. The bone marrow was extracted from femurs of inducible *Sox7* transgenic mice and single cell suspension was injected into sub-lethally irradiated (230 cGy) NSG mice. After a month, mice were fed with or without doxycycline. Different analyses and assays were performed every two months for a 12 months period.

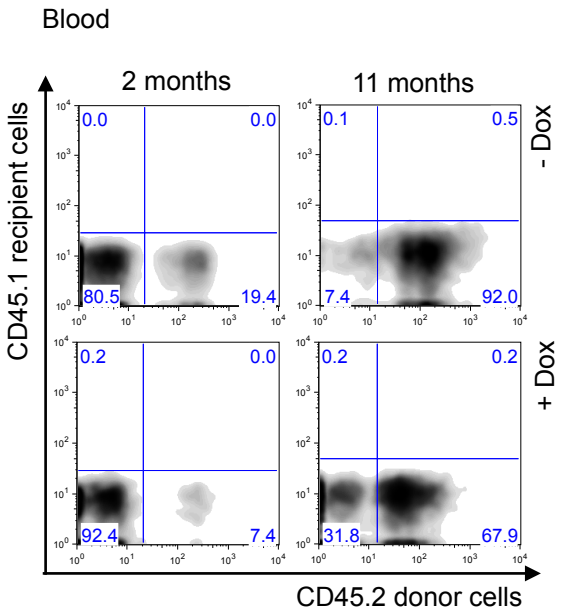


Figure 6.2. Full reconstitution of the haematopoietic system *in vivo*.
Representative FACS plots of CD45.1 and CD45.2 expression profiles of blood samples from mice injected with bone marrow extracted from femurs of inducible *Sox7* transgenic mice. Percentages of cells are shown in the indicated quadrants. FACS plots are representative of two independent experiments.

The frequency of the CD45.2$^+$ donor cells increased throughout the experiment reaching 68% and 92% in mice fed with or without doxycycline, respectively. The analyses also indicated that *Sox7*-enforced expression resulted in a progressive increase in the frequency of SOX7::GFP$^+$ cells in the analysed organs during the time course of the experiment (Figure 6.3).

Since B cell impairment was observed *in vitro* upon *Sox7*-enforced expression, B cell marker expression was analysed in the haematopoietic organs. In line with the *in vitro* data, the frequency of cells expressing B220, CD19 and IgM was significantly decreased in blood, bone marrow and spleen of mice over-expressing *Sox7*. The expression of B220, CD19 and IgM was restricted to the SOX7::GFP negative population while it was absent in the *Sox7*-expressing cells when compared to the control (Figure 6.4).

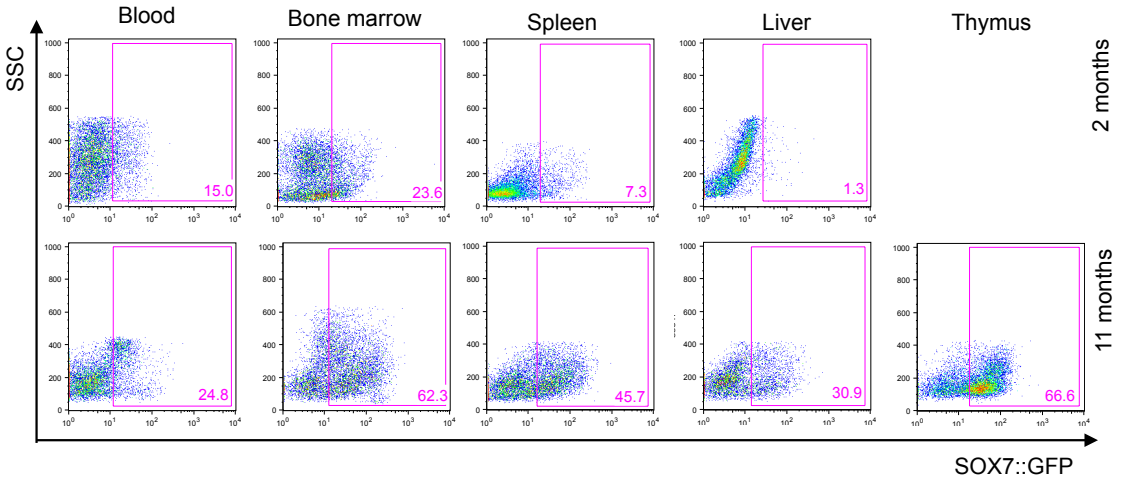


Figure 6.3. *Sox7*-enforced expression results in an increase in the frequency of SOX7::GFP-expressing cells.
Representative FACS plots of SSC and SOX7::GFP expression profiles of blood, bone marrow, spleen, liver and thymus cells extracted from mice fed with doxycycline. Percentages of cells are shown in the indicated gates. FACS plots are representative of two independent experiments.

In general, mature B cells expressing B220 and CD19 or IgM and IgD were mainly detectable in the control group while a significant decrease in the frequency of cells expressing these markers was observed in the *Sox7* over-expressing samples (Figure 6.5 A-B). The B cell development was further analysed based on the expression of B220 and IgM; expression of these markers distinguishes three different sub-populations from early B cells to mature B cells. In particular, pre B cells express only low level of B220 while low and high level of both B220 and IgM markers are expressed by immature B cells and mature B cells, respectively. The *Sox7*-expressing samples showed an absence of $IgM^{+}B220^{+}$ mature and $IgM^{low}B220^{low}$ immature populations compared to the control showing normal B cell development (Figure 6.5 C).

Altogether, these results reveal that the over-expression of *Sox7* impairs B cell development *in vivo*, extending further the data obtained *in vitro*.

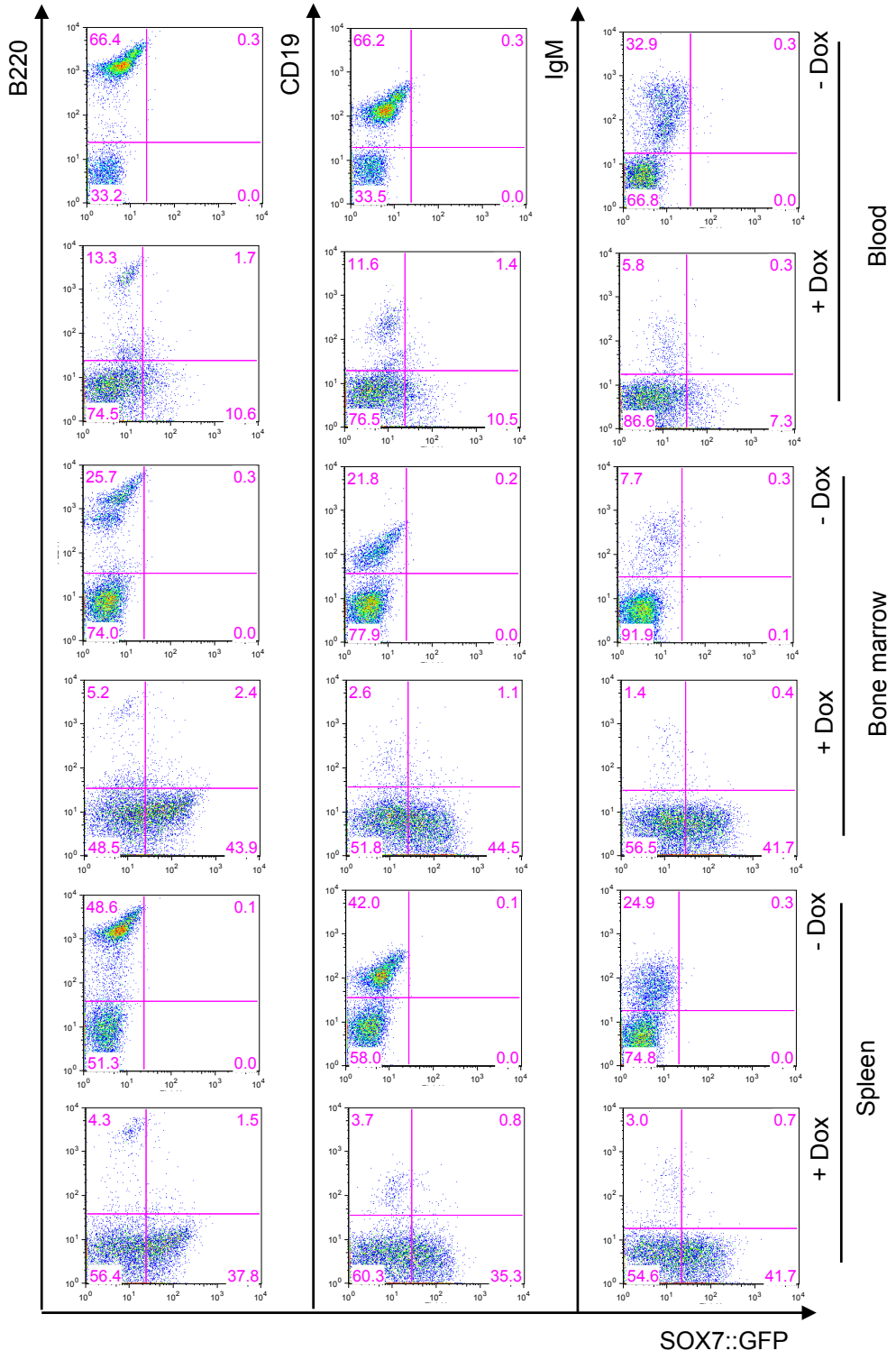


Figure 6.4. *Sox7*-enforced expression induces the loss of B cell markers *in vivo*. Representative FACS plots of B220, CD19, IgM and SOX7::GFP expression profiles of blood, bone marrow, spleen cells extracted from mice fed with or without doxycycline (11 months). Percentages of cells are shown in the indicated gates. FACS plot are representative of two independent experiments.

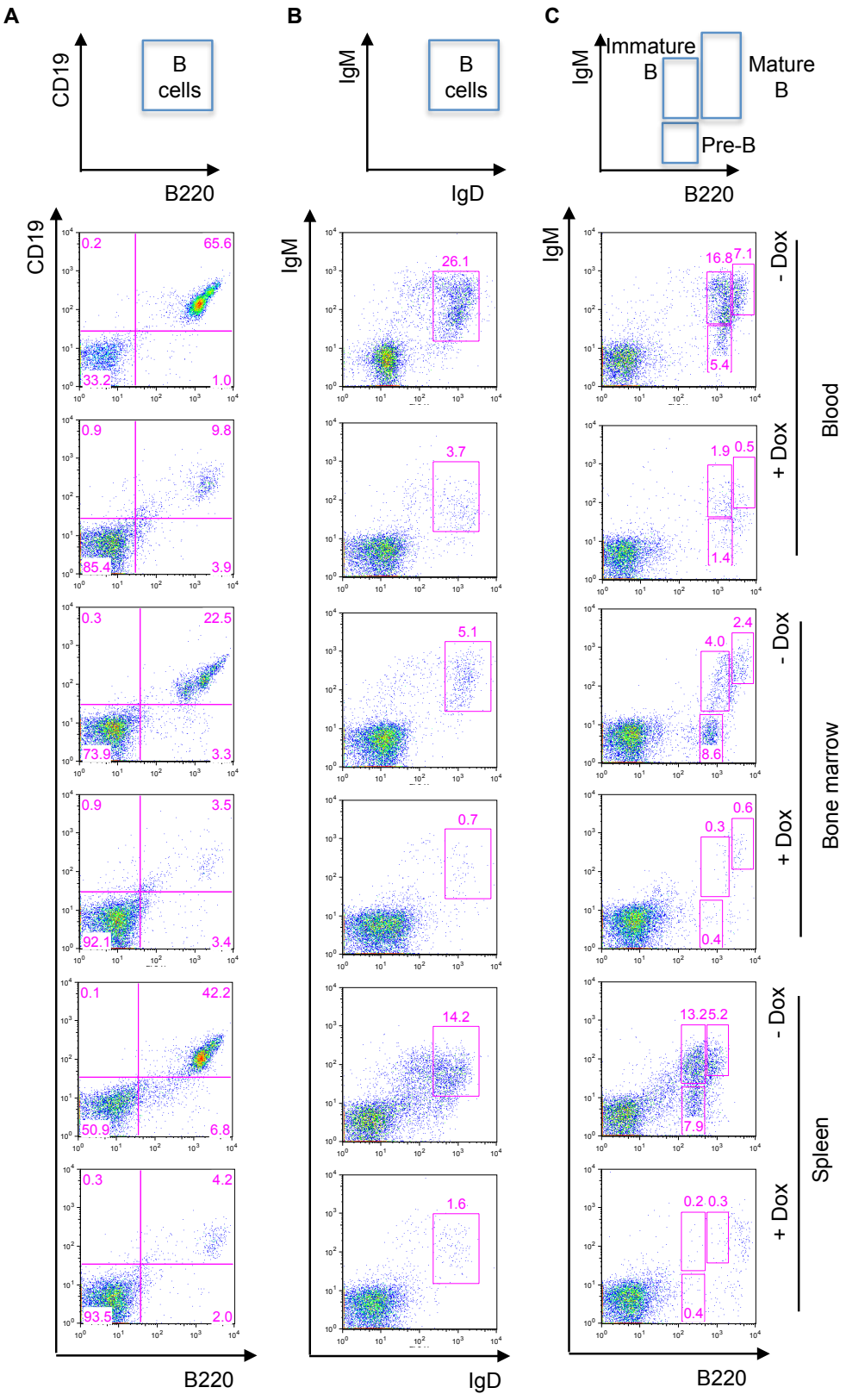


Figure 6.5. *Sox7*-enforced expression induces B cell impairment *in vivo*.
Representative FACS plots of B220, CD19 (A), IgD, IgM (B) and IgM, B220 (C) expression profiles of blood, bone marrow, spleen cells extracted from mice fed with or without doxycycline (11 months). Percentages of cells are shown in the indicated gates. FACS plots are representative of two independent experiments.

Since the over-expression of *Sox7* is induced by doxycycline in all haematopoietic cells, the analysis of all blood lineages was further undertaken. As described above, the percentage of $B220^{+}IgM^{+}$ cells was significantly decreased in doxycycline treated samples (Figure 6.6 A). In general, the myeloid and T lymphoid compartments were not affected by *Sox7*-enforced expression as shown by the percentage of $MAC1^{+}$ and $CD8^{+}$ cells in the bone marrow (figure 6.6 B-C). Of interest, the percentage of $TER119^{+}$ erythroid cells was increased by the over-expression of *Sox7* (Figure 6.6 D). This increase in erythroid cells seems to be due to an accumulation of reticulocytes but this aspect will be discussed in more detail in the next section.

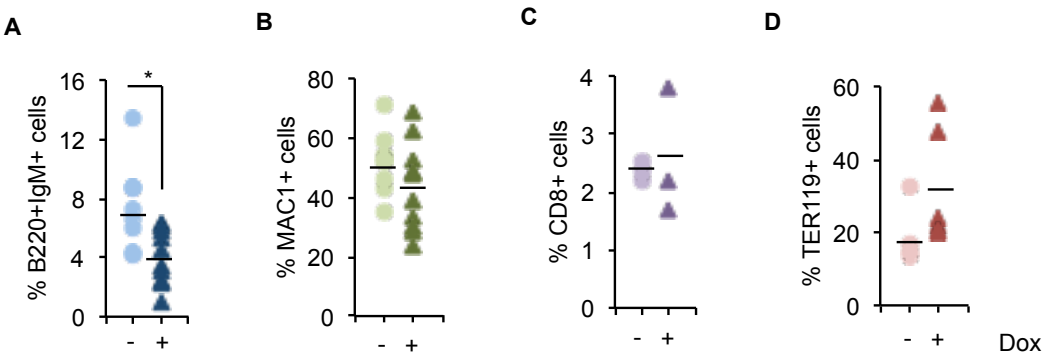


Figure 6.6. Analysis of the consequences of *Sox7* over-expression on blood lineages. Scatter chart of the $B220^{+}IgM^{+}$ (A), $MAC1^{+}$ (B), $CD8^{+}$ (C) and $TER119^{+}$ (D) cell percentage in the bone marrow of mice fed or not with doxycycline. Error bars indicate mean (n=8 for $B220^{+}IgM^{+}$, $MAC1^{+}$ and $TER119^{+}$, n=3 for $CD8^{+}$; * $P< 0.05$). Graphs are representative of two independent experiments.

The expression of the haematopoietic marker, CD45, was also analysed. Interestingly, its expression is reduced of about 10% in both bone marrow and spleen samples indicating defects in the haematopoiesis (Figure 6.7). This reduction is might be linked to the absence of the isoform CD45R, also called B220, in the *Sox7*-expressing cells or it might be due to the increase of erythrocytes that are CD45 negative.

Overall, these results suggest that *Sox7*-enforced expression impairs B cell development and maintenance *in vivo* while not affecting myeloid and T cell homeostasis.

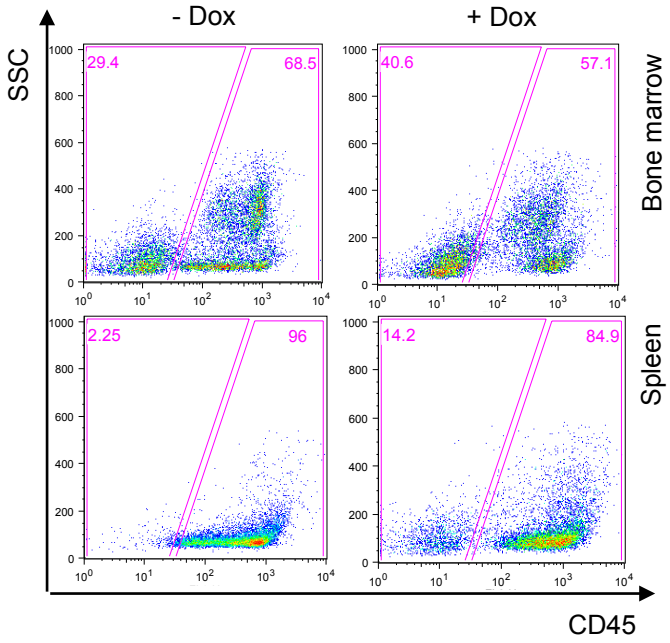


Figure 6.7. *Sox7*-enforced expression induces a decrease in CD45 expression *in vivo*.
Representative FACS plots of SSC and CD45 expression profiles in bone marrow and spleen cells extracted from mice fed with or without doxycycline (11 months). Percentages of cells are shown in the indicated gates. FACS plots are representative of two independent experiments.

6.2.2. Analysis of haematological parameters using Sysmex fluorescent flow cytometry

To complement the FACS analysis and to have a clearer overview at a clinical level, a detail analysis of haematological parameters was performed on blood samples using the Sysmex technology, a diagnostic tool routinely used in haematology. Sysmex fluorescent flow cytometry provides high sensitivity for measuring and differentiating cell types in blood samples. This instrument can classify normal white blood cells, red blood cells and platelets (Mathers et al., 2008).

To confirm the observation made upon the immuno-phenotypic analysis, the first parameter taken into account was lymphocytes number. As shown in figure 6.8 lymphocytes number decreased when *Sox7* is over-expressed.

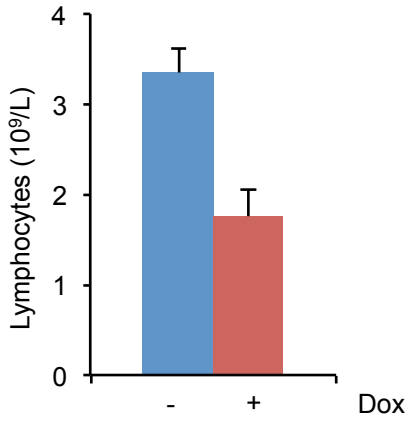


Figure 6.8. *Sox7*-enforced expression induces a decrease in lymphocytes *in vivo*. The bar chart shows lymphocyte count in the blood of mice fed with or without doxycycline. Error bars indicate mean ± standard deviation (n=6). Graph is representative of two independent experiments.

In parallel, several parameters related to red blood cells were analysed. Interestingly, an increase in the mean red blood cell volume (MCV) was reported. In detail, both standard deviation (RDW-SD) and coefficient variation of the red cell

distribution width (RDW-CV) was increased in mice fed with doxycycline suggesting a high variability in erythrocyte size (Figure 6.9 A). The standard deviation of the red cell distribution width measured the width of the erythrocyte distribution curve while the coefficient variation of the red cell distribution width depends on the mean size of erythrocytes. These results were confirmed by monitoring the morphology of peripheral blood cells on blood smears. The May-Grünwald staining clearly showed an increase in size of erythrocytes in mice fed with doxycycline (Figure 6.9 B). The MCV, in combination with the RDW, is used in the classification of anaemia. An increase in both MCV and RDW is associated with pernicious anaemia but further analyses will have to be done to confirm this clinical aspect (Thomas et al., 2005a).

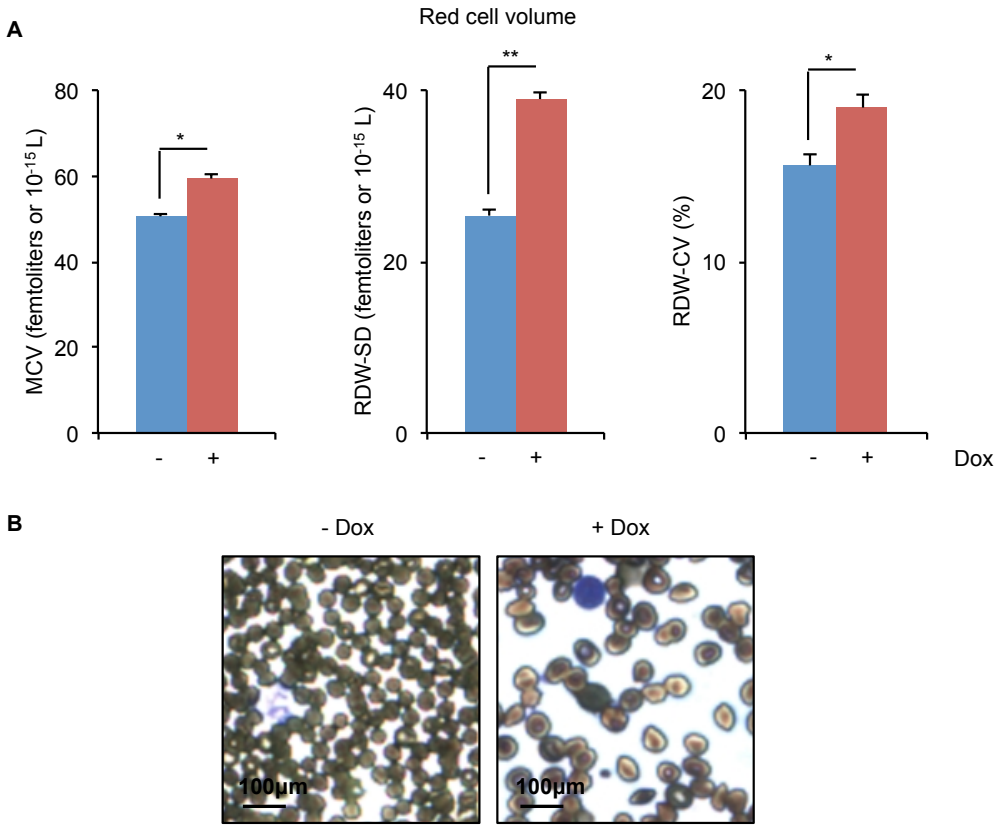


Figure 6.9. *Sox7*-enforced expression induces an increase in erythroid size *in vivo*. The bar chart shows three red blood cell volume parameters in mice fed with or without doxycycline. Error bars indicate mean ± standard deviation (n=6). Graphs are representative of two independent experiments. MCV: mean cell volume, RDW-SD: red cell distribution width-standard deviation, RDW-CV: red cell distribution width-coefficient variation.

Since the erythrocyte compartment was affected by *Sox7*-enforced expression, a more accurate analysis of erythrocyte development was carried out. The erythrocytes are generated by the expulsion of nucleus from the reticulocytes (Chiabrando et al., 2014). Therefore, a defect in the erythroid development might be due to an accumulation of reticulocytes. To verify this hypothesis, a deep analysis of reticulocyte sub-populations was performed. The reticulocyte development is characterized by several maturation stages. By removing of the endoplasmatic reticulum, the reticulocytes develop into mature red blood cells (Figure 6.10 A). To measure the reticulocytes count, blood samples are incubated with a fluorescent dye for the RNA content. Then, cells are acquired and counted by flow cytometry using the Sysmex machine. Therefore, the classification of reticulocyte maturation stages is defined by their RNA content (Corberand, 1996; Lee et al., 2014). Three main sub-populations are subdivided using this method. The high florescence reticulocytes (HFR) are characterized by high level of RNA content and represent the immature reticulocyte population. In contrast, the low fluorescence reticulocytes (LFR) contains little amount of RNA and indicates mature reticulocytes. Between these two sub-populations, it is possible to identify the medium fluorescence reticulocyte fraction (MFR) that contains medium level of RNA and indicates semi-mature reticulocytes (Figure 6.10 B).

In general, an increase in the percentage of reticulocytes was observed by fluorescent flow cytometry (Figure 6.11 A). Moreover, an increase in immature and semi-mature reticulocytes was detected resulting, as expected, in an increase in the immature reticulocyte fraction. Consequently, the mature reticulocyte sub-population was the only fraction to be reduced in the performed analysis (Figure 6.11 B). To confirm these observations, FACS analysis was performed on cells extracted from blood, bone marrow and spleen based on the expression of CD71 and TER119 markers.

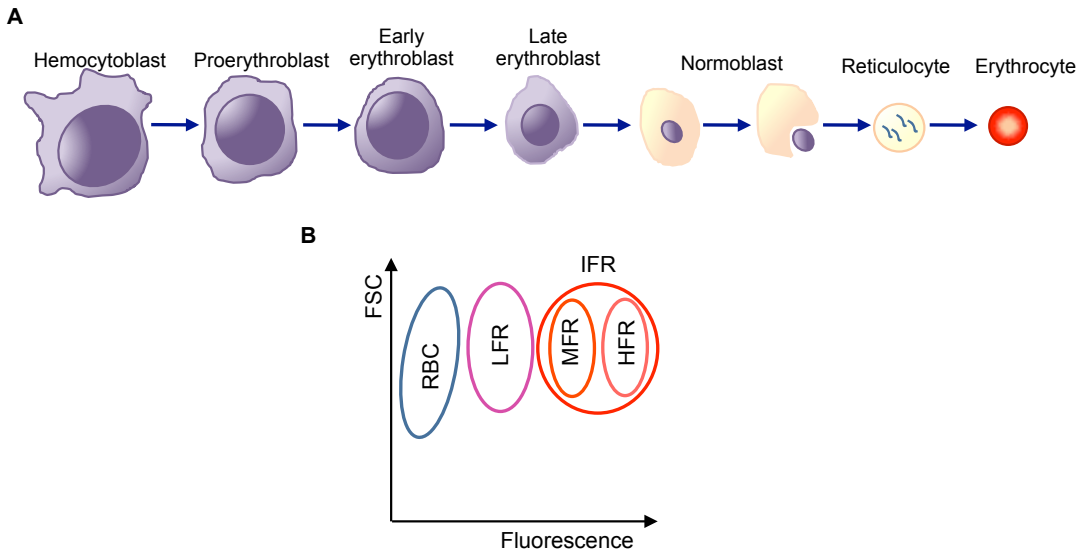


Figure 6.10. Scheme of reticulocyte development.
A) Schematic diagram of reticulocyte development. B) Scheme of FACS plot for the reticulocyte fractions. RBC: red blood cell, IRF: immature reticulocyte fraction (MFR+HFR), LFR: low-florescence reticulocyte (mature reticulocyte), MFR: middle-florescence reticulocyte (semi-mature reticulocyte), HFR: high-florescence reticulocyte (immature reticulocyte).

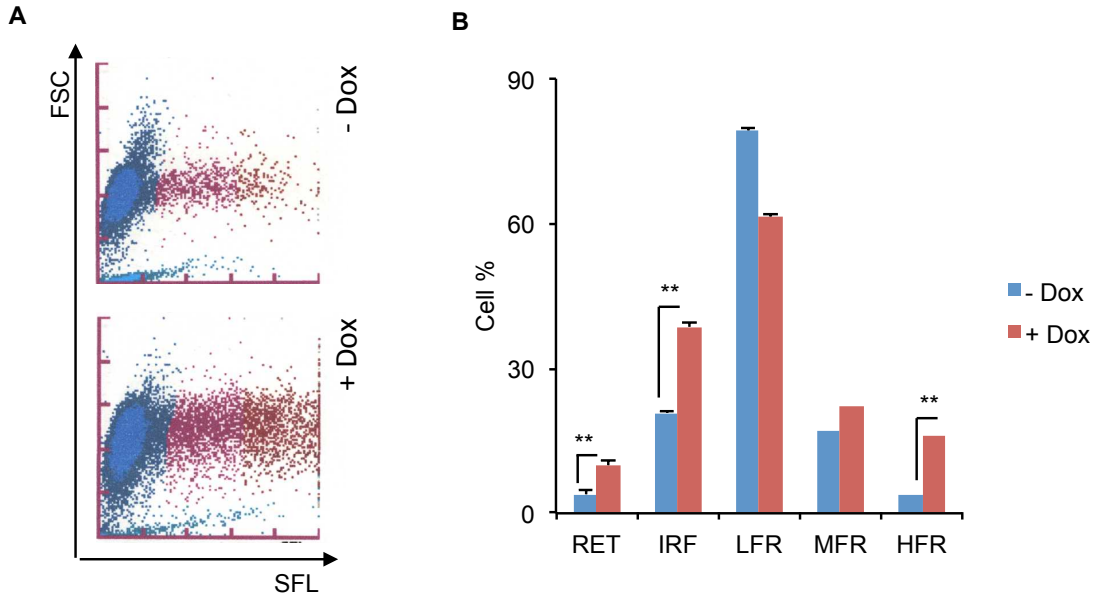


Figure 6.11. *Sox7*-enforced expression induces an increase in reticulocytes *in vivo*.
A) Representative FACS plots of FSC and SFL profiles in blood samples from mice fed with or without doxycycline (11 months). B) The bar chart shows different reticulocyte fractions in mice fed with or without doxycycline. Error bars indicate mean ± standard deviation (n=6). Graph is representative of two independent experiments. RET: reticulocytes, IRF: immature reticulocyte fraction (MFR+HFR), LFR: low-florescence reticulocyte (mature reticulocyte), MFR: middle-florescence reticulocyte (semi-mature reticulocyte), HFR: high-florescence reticulocyte (immature reticulocyte).

In all analysed samples an increase in $CD71^{+}TER119^{+}$ fraction was observed in *Sox7* over-expressing samples (Figure 6.12 A). Moreover, the FACS analysis showed an accumulation in late basophilic erythroblasts (EryB) in the $TER119^{+}$ population as reported in figure 6.12 B. Overall, these analyses suggest a defect in the erythroid maturation.

Sysmex analysis also revealed a defect in platelets formation as shown by the dramatic and significant drop in platelets count upon *Sox7*-enforced expression (Figure 6.13 A). Analysis of CD41 and GFP expression revealed that SOX7::GFP-expressing cells expressed very low level of CD41 (Figure 6.13 B). The megakaryocyte and erythrocyte lineages arise from the megakaryocyte-erythrocyte progenitors (MEP) (Murphy et al., 2013; Pang et al., 2005). By analysing the expression of CD41 and CD71, it is possible to discriminate the megakaryocytes from the erythrocyte cells. Interestingly, an accumulation of $CD41^{-}CD71^{+}$ cells was observed in the blood of mice fed with doxycycline suggesting an augment in erythroid progenitors. In contrast, a decrease in megakaryocyte progenitors was observed relative to the control sample suggesting a defective megakaryopoiesis upon *Sox7* enforced expression (Figure 6.13 C). The low frequency of erythroid progenitors in the control sample correlated with the presence of $TER119^{+}$ mature red blood cells in the peripheral blood as shown in figure 6.12.

Altogether, these results show that the over-expression of *Sox7* affects the development of several blood lineages.

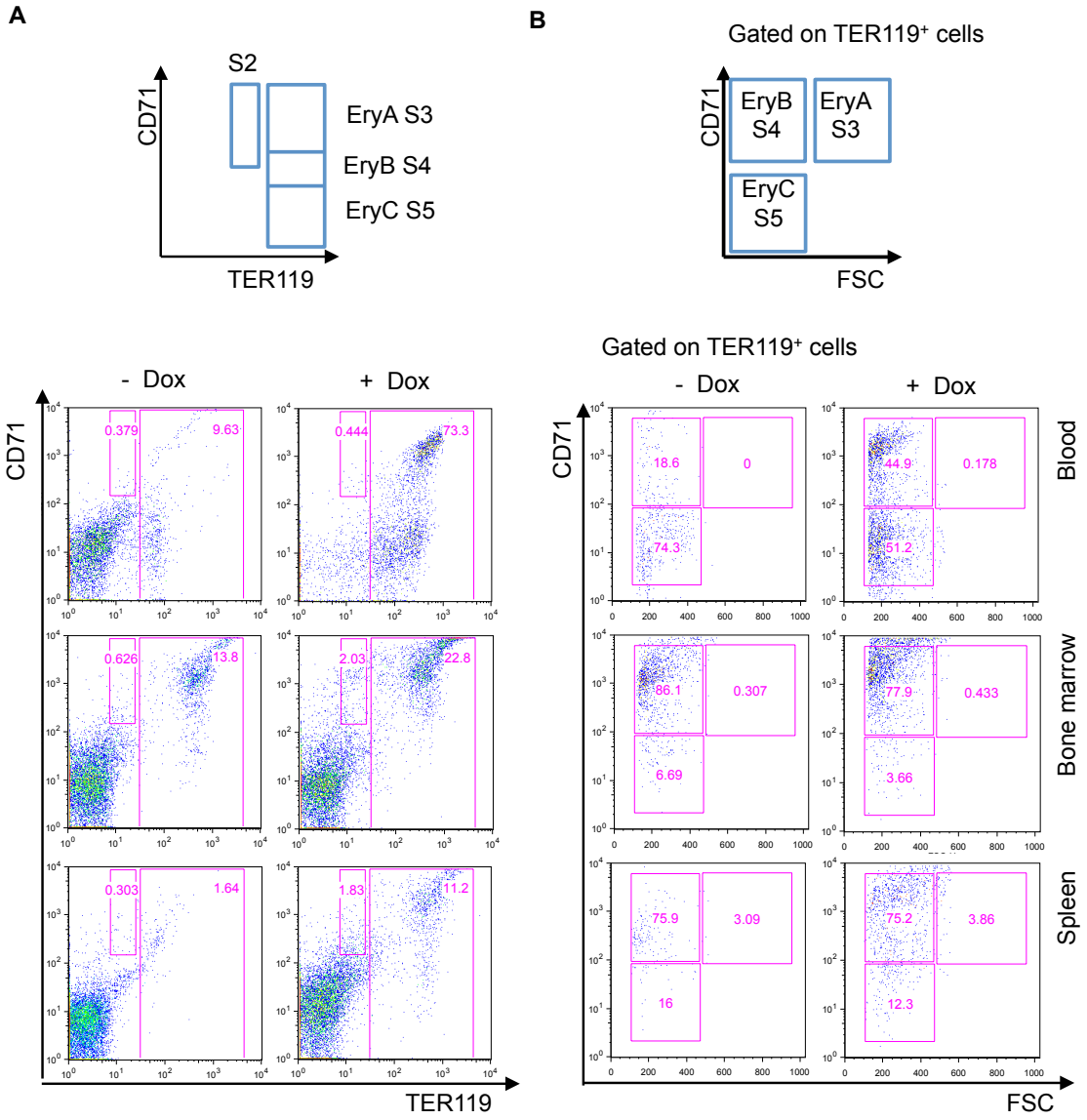


Figure 6.12. *Sox7*-enforced expression induces an increase in immature erythrocytes.
A) Representative FACS plots of CD71 and TER119 expression profiles of blood, bone marrow, spleen cells extracted from mice fed with or without doxycycline (11 months). B) Representative FACS plots of CD71 and FSC expression profiles of blood, bone marrow, spleen cells extracted from mice fed with or without doxycycline (11 months). Percentages of cells are shown in the indicated gates. FACS plots are representative of two independent experiments. EryA: early basophilic erythroblasts, EryB: late basophilic erythroblasts, EryC: poly/orthochromatic erythroblasts.

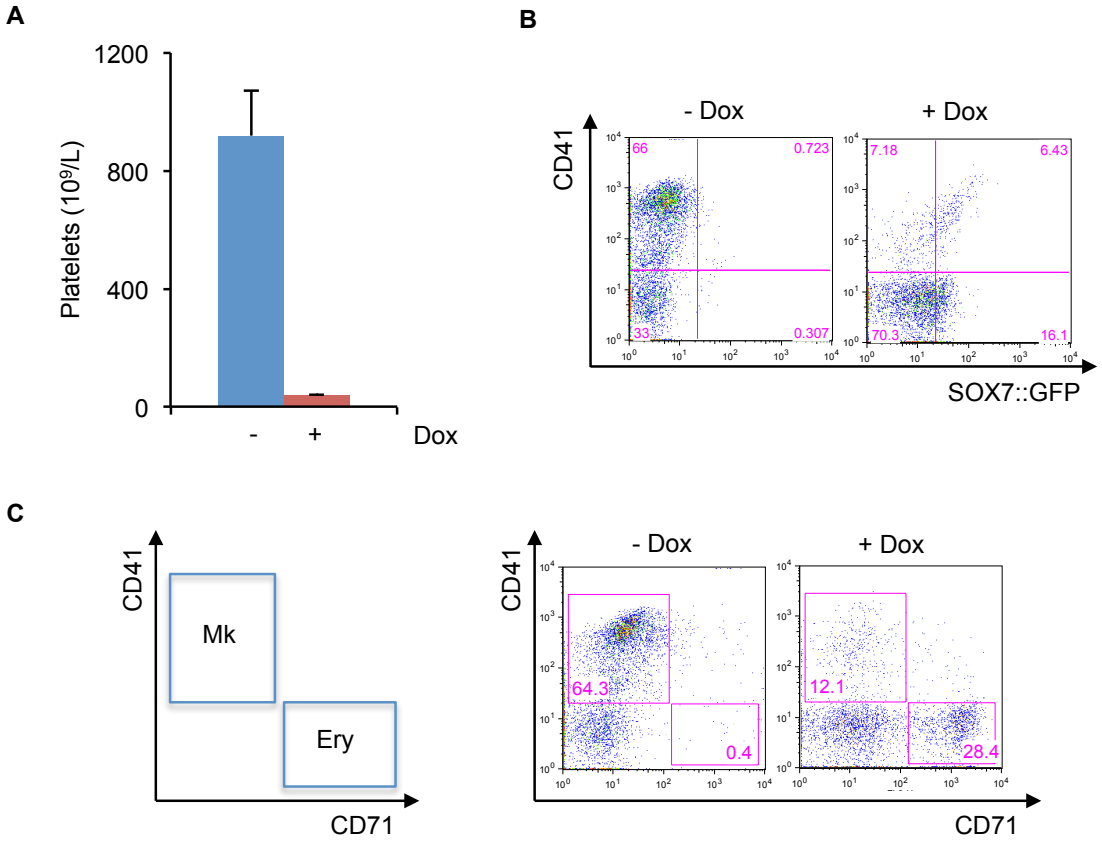


Figure 6.13. *Sox7*-enforced expression induces a decrease in platelets.
A) The bar chart shows platelets count in the blood of mice fed with or without doxycycline. Error bars indicate mean ± standard deviation (n=6). B) Representative FACS plots of CD41 and SOX7::GFP expression profiles of blood cells extracted from mice fed with or without doxycycline (11 months). Percentages of cells are shown in the indicated gates. C) Representative FACS plots of CD41 and CD71 expression profiles of blood cells extracted from mice fed with or without doxycycline (11 months). Percentages of cells are shown in the indicated gates. Mk: megakaryocyte, Ery: erythrocyte. Graph and FACS plots are representative of two independent experiments.

6.2.3. *Sox7*-enforced expression induces bone marrow failure and splenomegaly

To investigate whether *Sox7*-enforced expression induces signs of malignancy, the status and morphology of extracted organs was carefully analysed. Organs were first weighted then; they were used for flow cytometry analysis, morphologic analysis and immuno-histochemistry staining.

Sox7-enforced expression induced clearly splenomegaly demonstrated by a significant increase in spleen size. Moreover, a partial hepatomegaly was also observed in *Sox7*-over-expressing samples (Figure 6.14 A-B). Furthermore, the spleen architecture was completely altered. The spleen is characterized by two main compartments: white pulp and red pulp. The red pulp is involved in filtering the blood and recycling iron from old erythrocytes while the white pulp represents the lymphoid regions within the spleen (Mebius and Kraal, 2005). In particular, the lymphoid follicles were disrupted in *Sox7*-enforced expressing samples compared to the control samples where a clear distinction between the follicles and red pulp is visible (Figure 6.14 C).

Together, these results might explain the disruption of the spleen architecture due to the loss of B cells in the *Sox7*-expressing samples.

In order to confirm the association between the altered structure of follicles and the B cell impairment as shown by the immuno-phenotypic analysis, immuno-histochemistry analysis was performed on femur and spleen sections for the B cell marker, B220. In line with all previous data, this analysis revealed that *Sox7*-enforced expression resulted in a dramatic decrease in B220 expression both in the bone marrow and in the spleen compared to the controls (Figure 6.14 D).

Haematopoietic imbalance and sign of splenomegaly suggested bone marrow failure. One of the diagnostic criteria for bone marrow abnormality is the presence of fibrotic

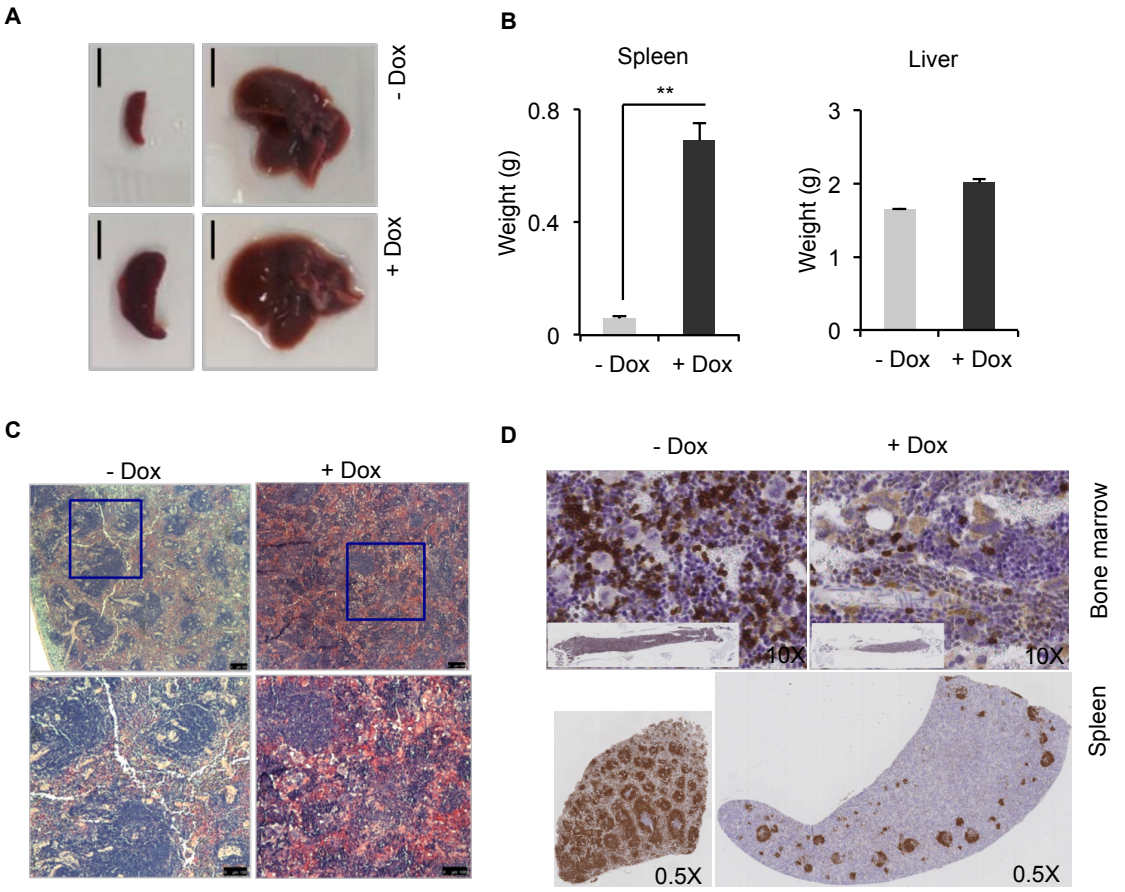


Figure 6.14. *Sox7*-enforced expression induces splenomegaly, spleen architecture disruption and decrease in B220 expression *in vivo*.
A) Representative pictures of spleen and liver of immuno-compromised mice feed with and without doxycycline (Leica and scan software, scale bar: 0.75 cm). B) The bar graph shows spleen and liver weight in grams. Error bars indicate mean ± standard deviation (n=8; **P≤0.01). C) Haematoxylin and eosin stained spleen sections of immuno-compromised mice engrafted with *iSox7* bone marrow cells fed with and without doxycycline (Leica and scan software, scale bar: 100 µm, higher magnification in the bottom panels). D) Immunohistochemistry of femur sections in the upper panels (Leica and scan software, magnification: 10X) and spleen sections in the bottom panels (Leica and scan software, magnification: 0.5X) stained with anti-B220. Pictures and graphs are representative of two independent experiments.

fibres (Kuter et al., 2007). To this end, reticulin staining was performed on femur sections. As suspected, fibrotic structures were identified in the bone marrow of *Sox7*-expressing samples (Figure 6.15 A). The staining showed a diffuse fibre network including thick and reduplicated fibres similar to Grade 3 of the modified Bauermeister scale (Kuter et al., 2007; Thiele and Kvasnicka, 2005). These data suggest the presence of a non-favourable structure and environment for haematopoiesis within the bone marrow.

The first organ able to replace the function of the bone marrow for blood production is the spleen. Indeed, in order to restore the levels of blood formation, spleen has to increase its size to react to the high demand of blood cells from the organism. The increase in size correlated with the presence of fibrosis in the spleen of *Sox7*-expressing samples (Figure 6.15 B). Splenic fibrosis is often induced by injury or inflammation and occurs as reparative process (Suttie, 2006). In particular, a dense network of thick reticulin fibres throughout the spleen was observed in the *Sox7* over-expressing samples compared to the few fibres restricted to the vessel walls in the control. These results might indicate that the fibrosis reported in the spleen has a more structural function than a pathological one like in the bone marrow. This might be explained by the fact that the bone marrow has a limited space delimited by femoral bone while the spleen has less restriction in space.

Altogether, these results confirmed a decrease in B cell differentiation, and documented bone marrow failure and splenomegaly.

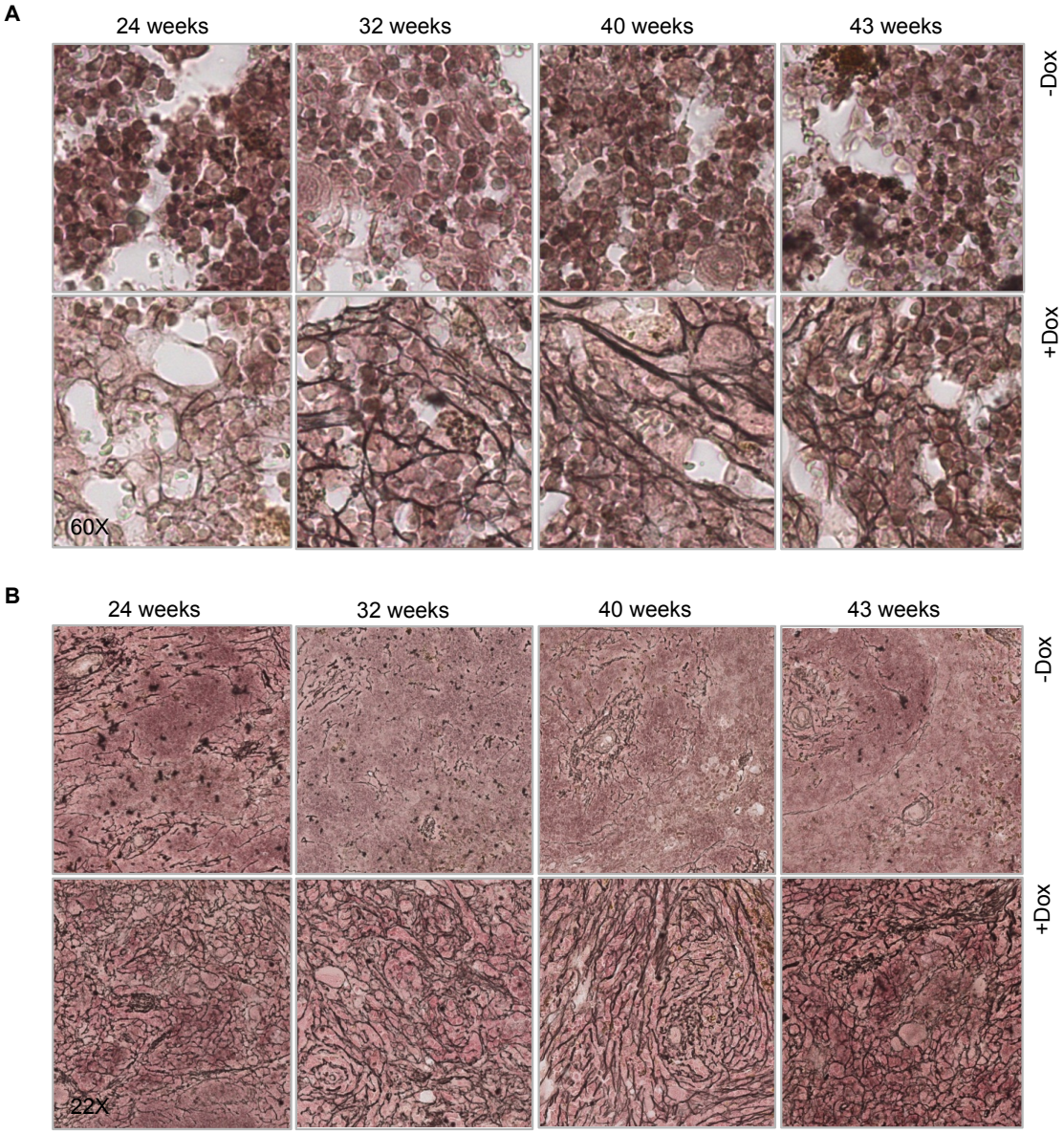


Figure 6.15. *Sox7*-enforced expression induces fibrosis in bone marrow and spleen. A) Reticulin stained femur sections of immuno-compromised mice engrafted with *iSox7* bone marrow cells fed with and without doxycycline (Leica and scan software, magnification: 60X). B) Reticulin stained spleen sections of immuno-compromised mice engrafted with *iSox7* bone marrow cells fed with and without doxycycline (Leica and scan software, magnification: 22X). Pictures are representative of two independent experiments.

6.2.4. *Sox7*-enforced expression induces extra-medullary haematopoiesis

Since sign of extra-medullary haematopoiesis was observed, the presence of progenitors in the spleen and liver was next investigated using clonogenic-replating assay. Cells were extracted from the organs and single cell suspension was plated in semi-solid culture without doxycycline. As shown in the figure 6.16, cells, extracted from the spleen and liver of mice fed with doxycycline, were enriched for progenitors as demonstrated by the significant increase in colony number.

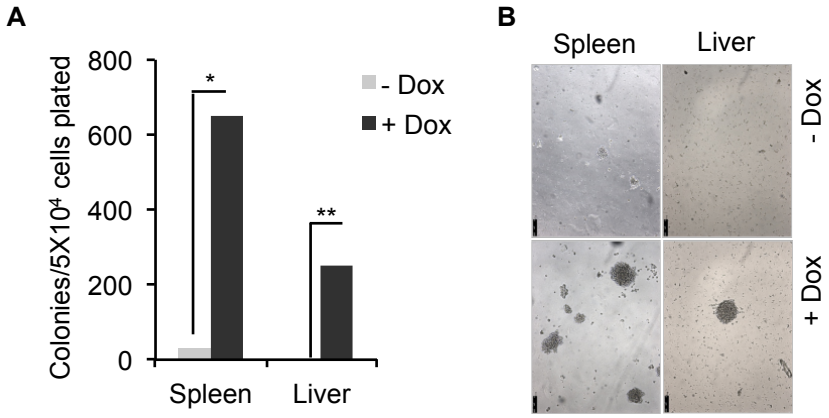


Figure 6.16. *Sox7*-enforced expression induces extra-medullary haematopoiesis. A) Spleen and liver cells were extracted from engrafted immuno-compromised mice fed with or without doxycycline and cultured in semi-solid culture without doxycycline for progenitor assay (n=3, ** P<0.01; * P<0.05). B) Representative picture of colonies obtained in the progenitor assay (Leica scan and software, scale bar: 100 μm). Pictures and graphs are representative of two independent experiments.

To assay the functionality of these extra-medullary generated progenitors, spleen cells, extracted from mice fed with doxycycline, were injected into sub-lethally irradiated recipient mice (Figure 6.17 A). FACS analysis revealed that spleen cells retained the ability to engraft recipient mice as reported by the expression of $CD45.2^+$ donor cells in the blood circulation (Figure 6.17 B). Moreover, cells retained both short and long-term engraftment capacity as shown in figure 6.17 C by the persistence of $CD45.2^+$ cells at 16 weeks post engraftment.

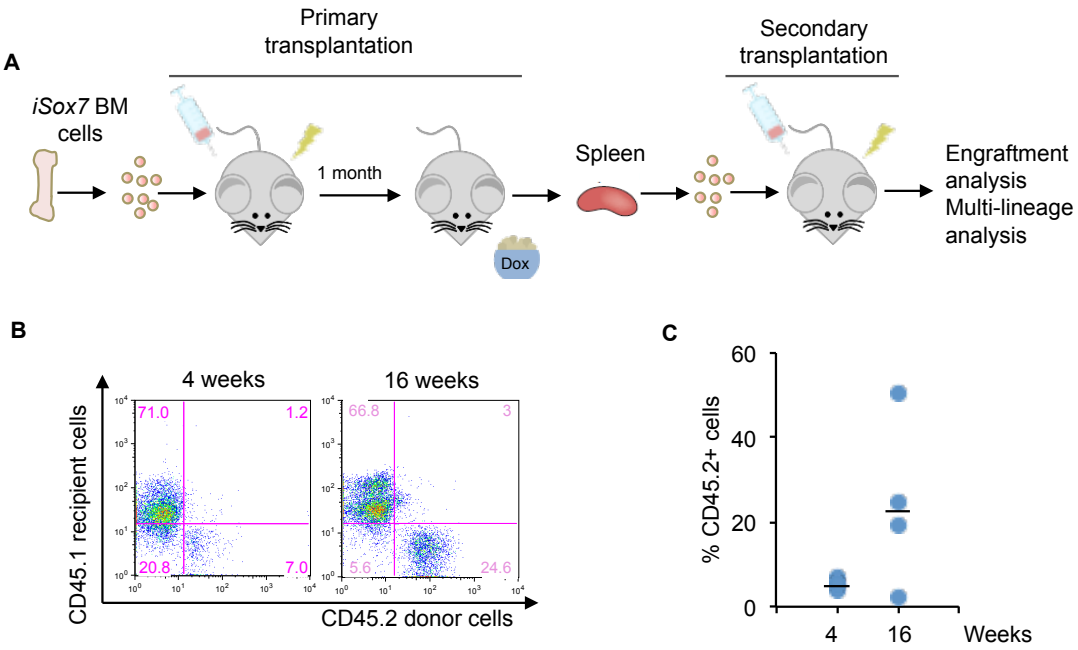


Figure 6.17. Progenitor cells within the spleen retain both short and long-term engraftment capacity.
A) Scheme of the protocol used to test the presence of progenitor cells in spleen. Spleen cells were extracted from mice engrafted with bone marrow cells from inducible *Sox7* transgenic mice fed with doxycycline. Single cell suspension was injected into irradiated secondary recipients. Engraftment analysis and multi-lineage analysis were performed. B) Representative FACS plots of CD45.1 and CD45.2 expression profiles of blood samples from secondary recipients. Percentages of cells are shown in the indicated quadrants. C) Scatter chart of the CD45.2 positive cell percentage in mice fed without doxycycline. Error bars indicate mean (n=4). Graph and FACS plots are representative of two independent experiments.

To investigate the multi-lineage engraftment capability of these cells, cells from the haematopoietic organs were extracted at 20 weeks post engraftment and analysed for lineage contribution. In particular, the expression of CD71 and TER119 markers were analysed for presence of erythroid cells, MAC1 and GR1 for myeloid cells, IgM and B220 or CD4 and CD8 for B and T lymphoid cells, respectively. As shown in figure 6.18, $CD45.2^+$ spleen cells retained the ability to contribute to all lineages in all organs analysed.

Given that the over-expression of *Sox7* induces an accumulation of progenitor cells *in vitro* and causes haematological imbalance *in vivo*, spleen cells were tested for their ability to induce haematological malignancy upon secondary transplantation in mice fed with doxycycline (Figure 6.19 A). No defects in the ability to engraft were observed by flow cytometry analysis for the expression of $CD45.2^+$ cells (Figure 6.19 B-C). In addition, analysis of the multi-lineage engraftment was assessed in the haematopoietic organs upon secondary transplantation. Only, and as expected, a defect was observed in the production of B lymphocyte cells in the bone marrow compartment but no malignancy was identified (Figure 6.20).

Overall, these results demonstrate the presence of extra-medullary haematopoiesis in spleen and liver and that the infiltrated progenitors retain long and multi-lineage engraftment *in vivo*.

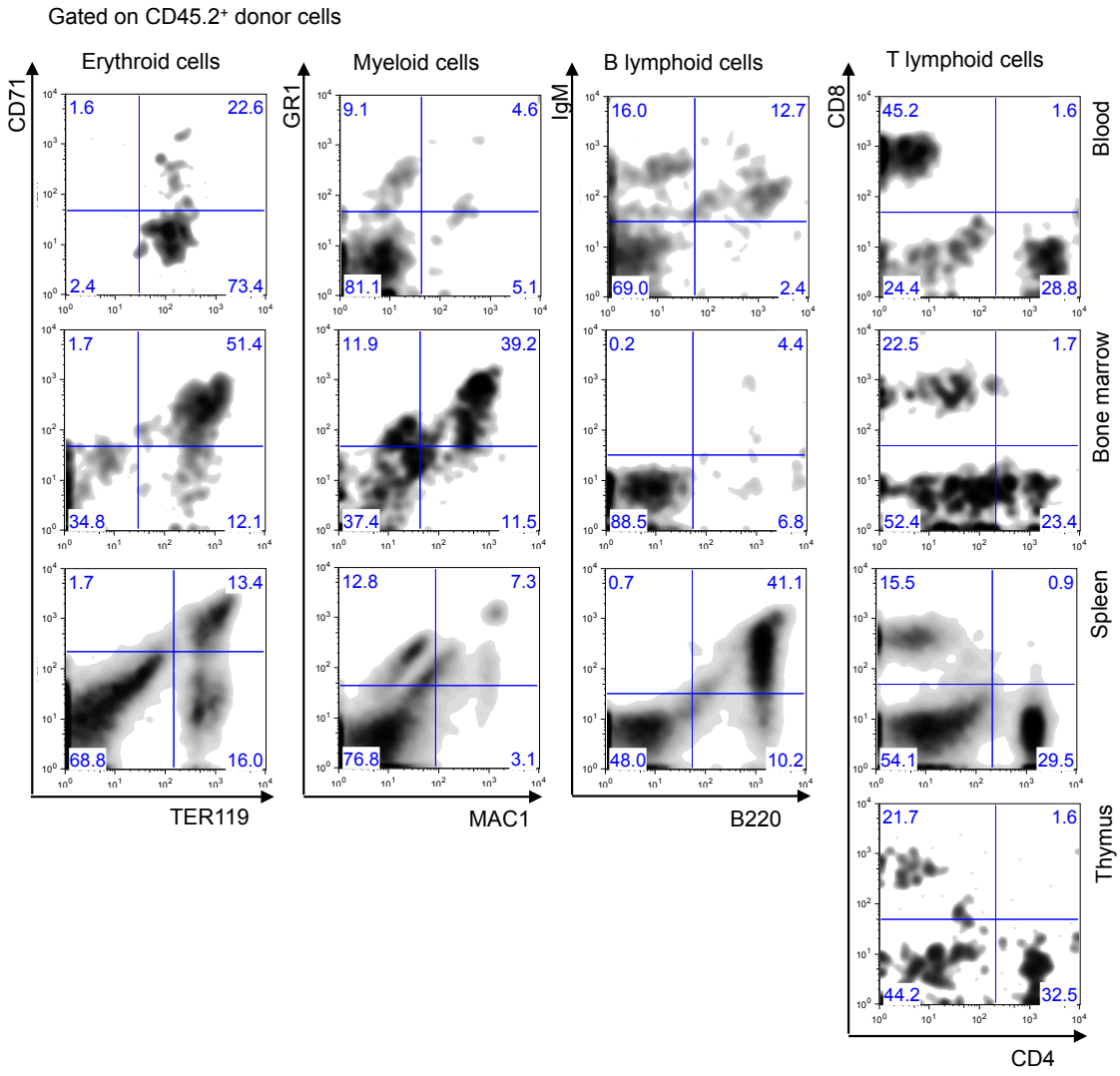


Figure 6.18. Progenitor cells within the spleen retain multi-lineage engraftment ability.
Representative FACS plots of TER119, CD71, GR1, MAC1 CD19, B220, CD8 and CD4, expression profiles of blood, bone marrow, spleen and thymus samples from secondary recipients. Analysis was performed at 20 weeks post engraftment. FACS plots are gated on CD45.2^{+} donor cells. Percentages of cells are shown in the indicated quadrants. FACS plots are representative of two independent experiments.

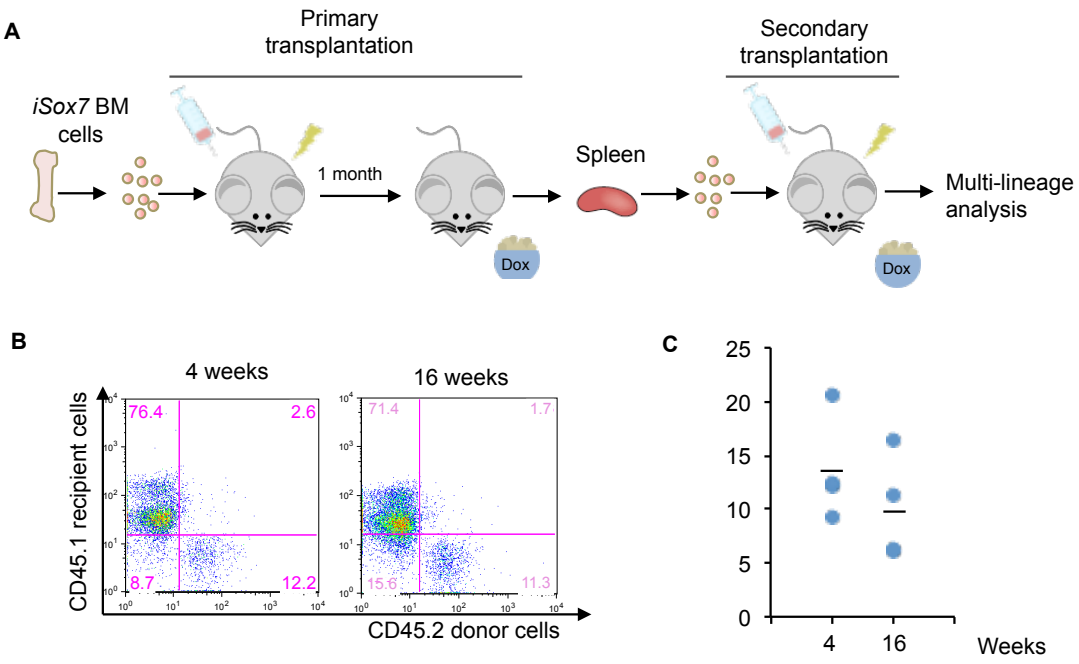


Figure 6.19. Progenitor cells within the spleen retain both short and long-term engraftment capacity upon doxycycline treatment.
A) Scheme of the protocol used to test the ability to induce haematological malignancy of progenitor cells in spleen. Spleen cells were extracted from immuno-compromised mice engrafted with bone marrow cells from inducible *Sox7* transgenic mice fed with doxycycline. Single cell suspension was injected into irradiated secondary recipients fed with doxycycline. Engraftment analysis and multi-lineage analysis were performed. B) Representative FACS plots of CD45.1 and CD45.2 expression profiles of blood samples from secondary recipients. Percentages of cells are shown in the indicated quadrants. C) Scatter chart of the CD45.2 positive cell percentage in mice fed with doxycycline. Error bars indicate mean (n=4). Graph and FACS plots are representative of two independent experiments.

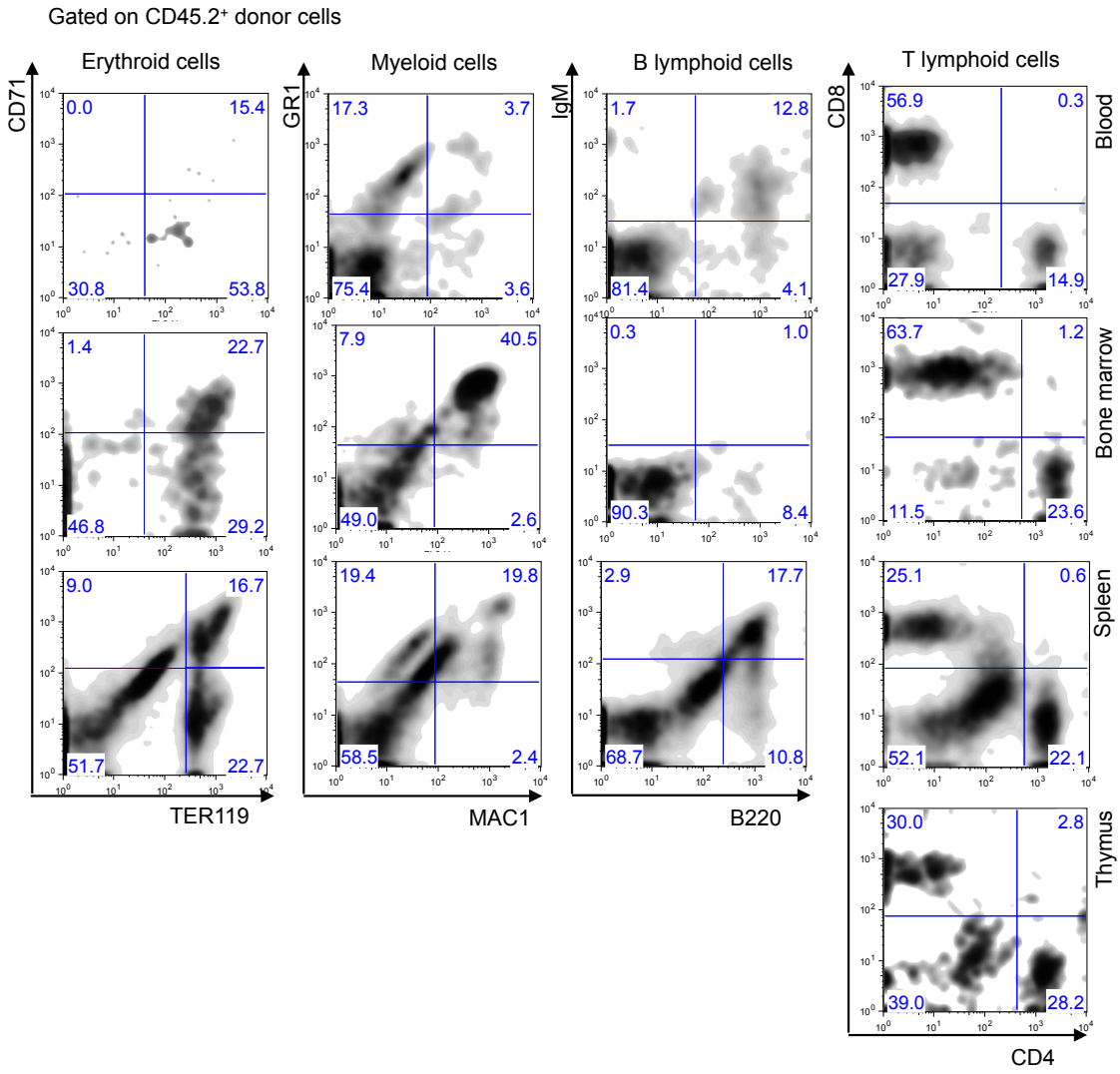


Figure 6.20. Progenitor cells within the spleen do not induce haematological malignancy upon doxycycline treatment.
Representative FACS plots of TER119, CD71, GR1, MAC1 CD19, B220, CD8 and CD4, expression profiles of blood, bone marrow, spleen and thymus samples from secondary recipients fed with doxycycline. Analysis was performed at 20 weeks post-engraftment. FACS plots are gated on CD45.2^{+} donor cells. Percentages of cells are shown in the indicated quadrants. FACS plots are representative of two independent experiments.

6.3. Summary of findings

In order to investigate the contribution of *Sox7*-enforced expression to haematopoietic malignancy, sub-lethally irradiated immuno-compromised mice were injected with total bone marrow derived from *iSox7* mouse. In this *in vivo* model, the over-expression of *Sox7* is only inducible in haematopoietic cells. The analyses first indicated that *Sox7*-enforced expression resulted in a progressive increase in the frequency of SOX7::GFP^{+} cells in the analysed organs during the time course of the experiment. Moreover, the expression of the B cell markers was significantly and severely impaired in blood, bone marrow and spleen samples. In contrast, the myeloid and T lymphoid compartments were not affected by *Sox7*-enforced expression. Of interest, the percentage of TER119^{+} erythroid cells was increased by the over-expression of *Sox7*.

A detailed analysis of haematological parameters was performed using the Sysmex apparatus during the time course of the experiment. Briefly, the analysis confirmed a drop in lymphocyte number when *Sox7* is over-expressed. An accumulation of immature erythrocytes was also revealed in *Sox7*-expressing samples. In parallel, an increase in the frequency of reticulocytes was observed revealing defects in the erythroid maturation. In addition, a defect in platelets formation was also observed by the dramatic and significant drop in platelets count in *Sox7*-enforced blood samples.

Next, to have a full picture of the consequences of *Sox7* over-expression *in vivo*, the morphology of the organs were analysed. Interestingly, I observed a significant increase in spleen size while a modest one in the liver. The spleen architecture was completely altered; in particular the lymphoid follicles were disrupted. In parallel to this imbalance in the lymphoid organization, I also observed that *Sox7*-enforced expression results in a dramatic decrease in B220 expression both in the bone marrow and in the

spleen compared to the control. To complement these analyses and to confirm bone marrow failure, the presence of reticulin fibres were investigated. Both, bone marrow and spleen contained fibrotic fibres upon *Sox7*-enforced expression. Diseases with increased bone marrow fibrosis are often associated with abnormalities in megakaryocyte and platelets production. Therefore, the presence of fibres in the bone marrow might be due to the impairment in platelet formation but further investigations have to be done.

Finally, since the bone marrow was fibrotic and the development of several haematopoietic lineages was compromised, the presence of extra-medullary haematopoiesis in secondary haematopoietic organs was analysed. As hypothesized, the main production of blood cells was displaced to the spleen and liver. These organs were enriched in progenitor cells that retained long-term and multi-lineage engraftment capacity.

Given these haematological imbalances upon *Sox7*-enforced expression *in vivo*, I extended this series of experiments analysing to determine whether the continued over-expression of *Sox7* upon secondary transplantation might lead to more pronounced haematological malignancy. The progenitor cells derived from the spleen did show long-term and multi-lineage repopulating ability but no significant haematopoietic imbalance or haematological diseases were observed.

Chapter 7

7. SOX7 in leukaemogenesis

7.1. Introduction

Acute lymphoblastic leukaemia is one of the most frequent cancers in children. Treatment of acute lymphoblastic leukaemia has improved substantially in last years reaching 80 per cent of 5-year survival rate in children compared to 40 per cent in adult. However, there are still poor prognosis cases associated with relapse or resistant ALL (Pui et al., 2004; Pui et al., 2008). Therefore, the identification of new molecular targets or understanding of the molecular mechanisms of action of primary genetic lesions involved in leukaemia is definitely necessary.

In general, leukaemia is induced by the mis-expression of proto-oncogenes, mutated transcription factors or chromosomal translocation. The most common alterations in children are hyperdiploidy (25%), TEL-AML1 (22%) and MLL rearrangements (8%) while BCR-ABL and MLL rearrangements in adults with a frequency of 25% and 10%, respectively (Pui et al., 2004).

SOX7 expression has been considered a tumour related gene. Its expression is down-regulated in many solid tumours such as prostate, colon and breast cancers (Guo et al., 2008; Chan et al., 2012). A study reported by Gandillet and colleagues has suggested a possible role for SOX7 in blood malignancy. Using a murine model, they have reported that *Sox7* knock-down *in vitro* significantly inhibits haematopoietic differentiation and that the enforced expression of *Sox7* in the earliest committed

haematopoietic precursors promotes proliferation while blocking their differentiation (Gandillet et al., 2009).

Since the expression of *SOX7* is specifically found in multiple cases of B cell acute lymphoblastic leukaemia and its enforced expression in adult bone marrow cells induced B cell impairment, the next step in my project was to define the possible relationship between *SOX7* expression and leukaemogenesis in human cell lines by knocking-down *SOX7* using shRNA. First, I analysed the requirement of *SOX7* expression for cell growth and colony formation *in vitro*. Then, since these cell lines are able to induce leukaemia in mice, I studied the possible role for *SOX7* in leukaemogenesis testing whether *SOX7* expression was required for the engraftment and formation of leukaemia *in vivo*.

7.2. Results

7.2.1. Down-regulation of *SOX7* induces a significant decrease in proliferation *in vitro*

To investigate the role for *SOX7* in leukaemogenesis, the expression of *SOX7* was knocked-down in human leukaemia cell lines using two different shRNA constructs. Both the shRNAs and Scramble control constructs carry sFFV as a promoter and GFP-2A-FireflyLuciferase to monitor the transduced cells both *in vitro* and *in vivo*. The efficiency of eight independent shRNAs to knock-down *SOX7* expression was tested in HUVEC cells by analysing *SOX7* expression levels by qRT-PCR and SOX7 protein levels by western blotting as reported in figure 7.1.

SOX7 expression level was down-regulated by all the shRNA constructs. All shRNAs worked extremely but so the shRNA_2 and _3 were chosen to carry on further analysis

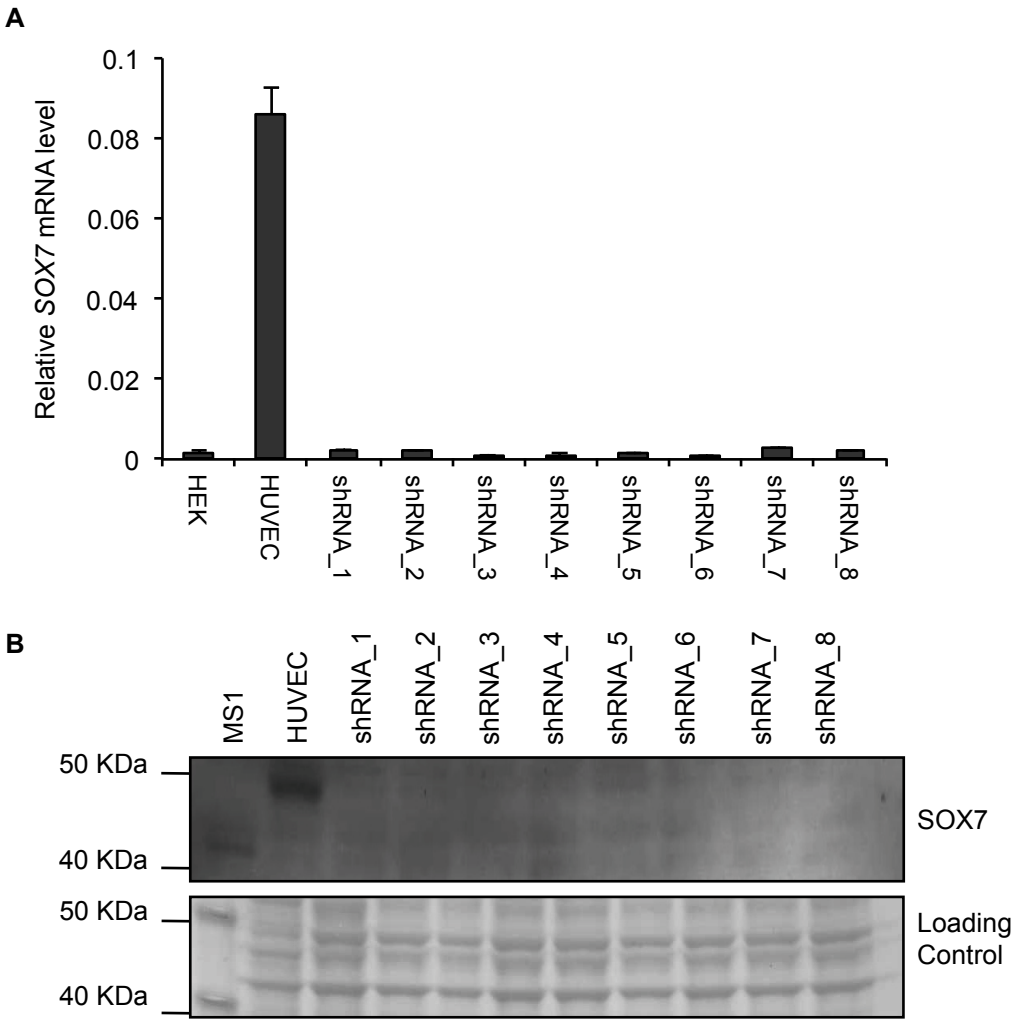


Figure 7.1. Efficiency of shRNA against *SOX7* in HUVEC cells.
A) Quantitative RT-PCR analysis of *SOX7* transcript level relative to β2microglobulin in HUVEC cells transduced with different shRNA against *SOX7*. HEK cells were used as a negative control. Error bars indicate mean ± standard deviation (n=3). B) Western blotting analysis for SOX7 of whole protein extract from HUVEC cells transduced with different shRNA against *SOX7*. Protein extract from MS1 cell line was used as a mouse positive control. Ponceau was used as loading control. SOX7 molecular weight is 42-49 KDa. Graph and western blotting are representative of one experiment.

due to their high efficiency and their different binding to the *SOX7* locus as reported by the blast alignment analysis in figure 7.2.

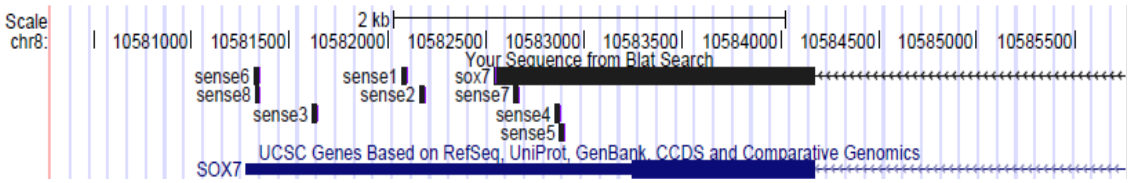


Figure 7.2. Binding of shRNAs to *SOX7* locus.
Representation of the binding of the different shRNA to the human *SOX7* locus. shRNAs are indicated as sense together with a number. Only shRNA 4, 5 and 7 are binding to *SOX7* coding sequence, all the other shRNAs are binding to the UTR region. The alignment was performed using BLAST browser.

Four different B-ALL leukaemia cell lines were transduced with lentiviruses carrying the shRNA against *SOX7* or the control construct. The molecular genetic of REH cell line is characterized by the expression of the fusion gene resulting from the TEL-AML1 translocation while RS4:11 cell line contains a MLL-AF4 chromosomal translocation (e10-e4). The NALM6 cell line contains a PDGFRβ-ETV6 translocation and finally, the 697 cell line is characterized by TCF3-PBX1 fusion gene.

Leukaemia cell lines were efficiently transduced with the lentiviruses as reported by the frequency of more than 90 per cent of GFP expression after puromycin selection (Figure7.3). The efficiency of the knock-down was tested by qRT-PCR. As observed in figure 7.4, *SOX7* expression was decreased of about 50% by the two shRNAs against *SOX7* compared to the control one. In addition, the knock-down efficiency with sh*SOX7*_3 worked slightly better than the one performed with sh*SOX7*_2.

To test whether the down-regulation of *SOX7* might affect the proliferation of these leukaemia cell lines, cells were cultured in their growth medium for more than 12

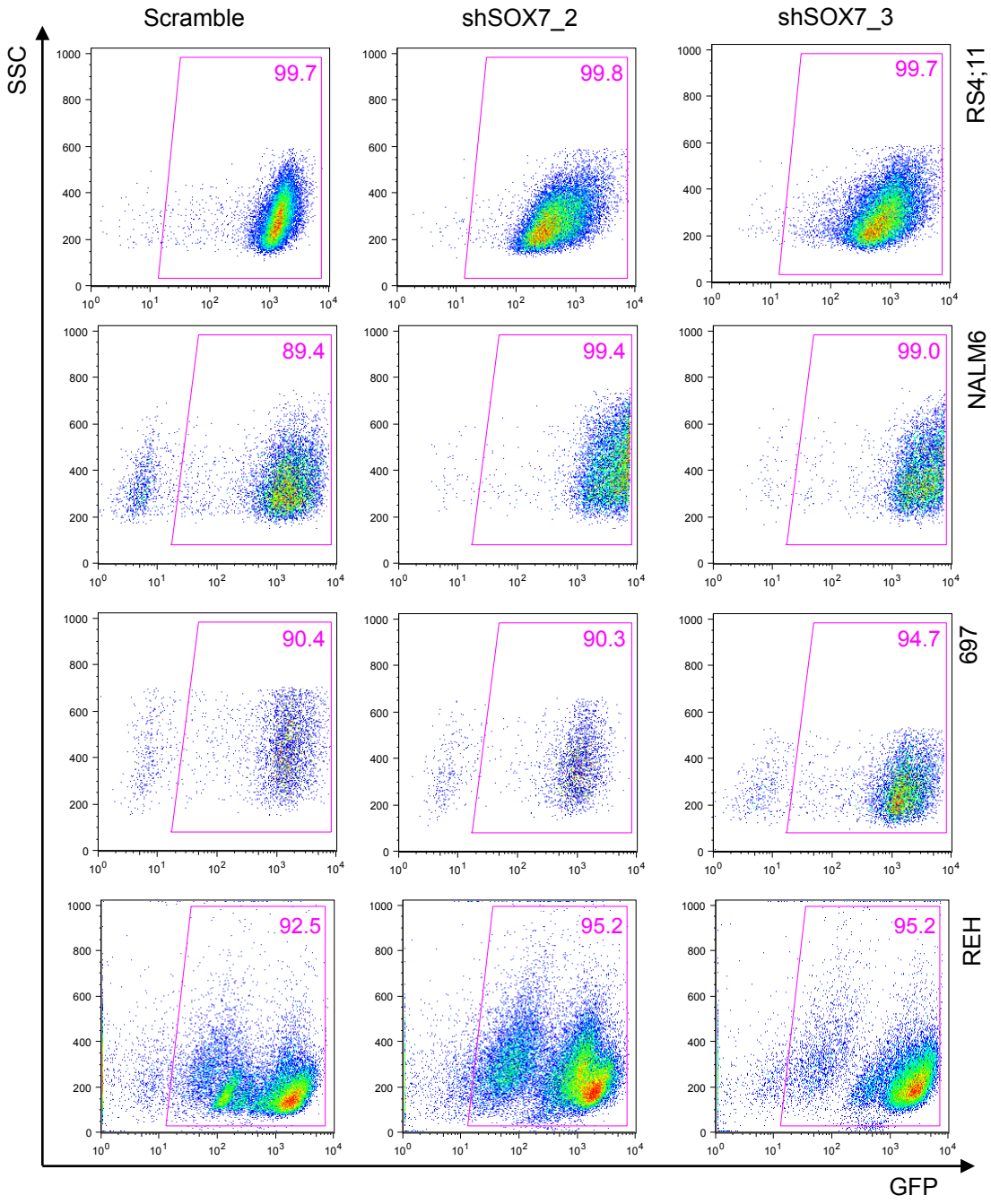


Figure 7.3. GFP expression in transduced leukaemia cell lines.
Representative FACS plots of SSC and GFP expression profiles of RS4;11, NALM6, 697 and REH human leukaemia cells transduced with two different shRNAs against *SOX7* (sh*SOX7*_2, sh*SOX7*_3) and or control construct (Scramble). Percentages of cells are shown in the indicated quadrants. FACS plots are representative of two independent experiments.

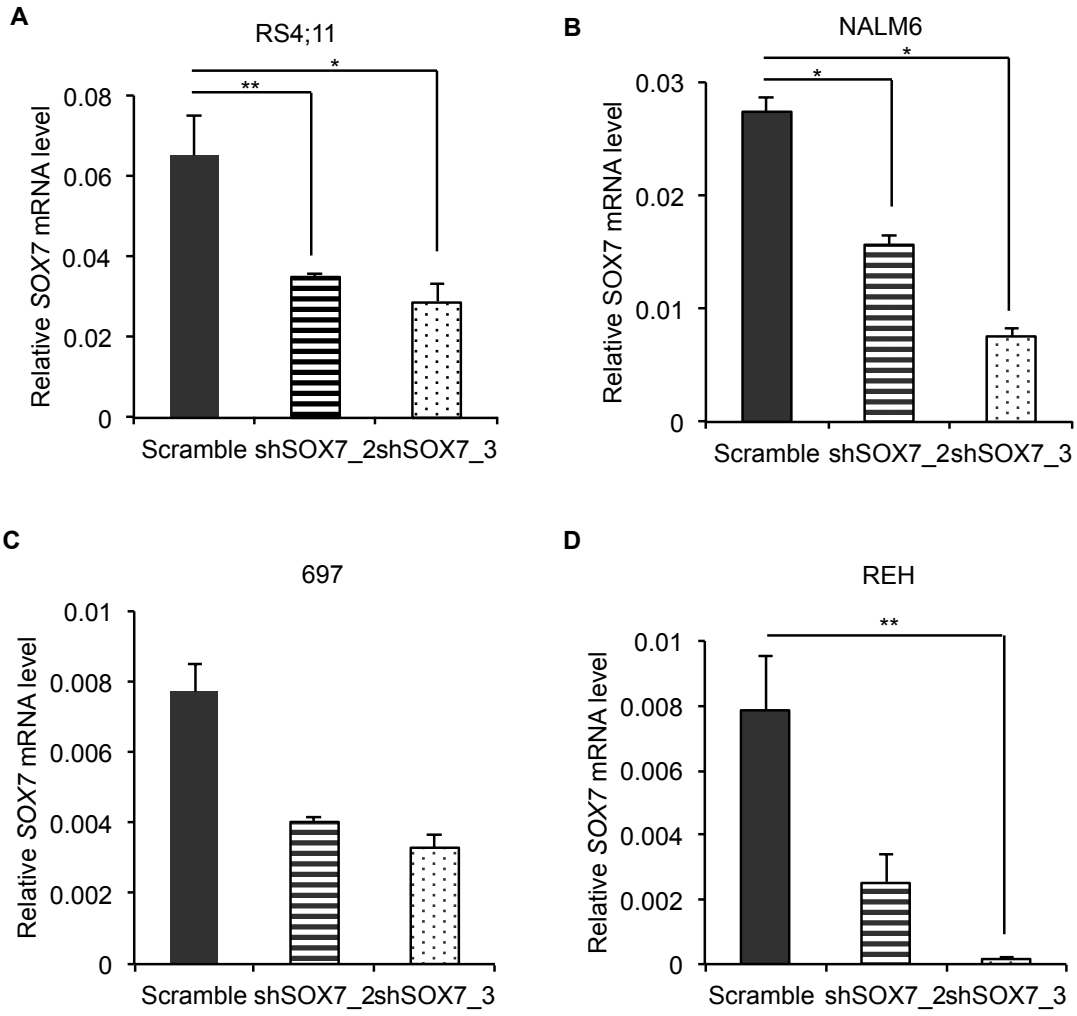


Figure 7.4. The two shRNAs against *SOX7* induce a decrease in *SOX7* expression in human leukaemia cell lines.
Quantitative RT-PCR analysis of *SOX7* transcript level relative to β2microglobulin in RS4;11 (A), NALM6 (B), 697 (C) and REH (D) human leukaemia cells transduced with two different shRNAs against SOX7 (sh*SOX7*_2, sh*SOX7*_3). Leukaemia cells transduced with a control construct (Scramble) were used as a positive control. Error bars indicate mean ± standard deviation (n=3, * P<0.05, ** P<0.01). Bar graphs are representative of two independent experiments.

To test whether the down-regulation of *SOX7* might affect the proliferation of these leukaemia cell lines, cells were cultured in their growth medium for more than 12 days. Interestingly, the down-regulation of *SOX7* in RS4;11 and NALM6 with both shRNAs induced a significant decrease in cell growth compared to the control shRNA *in vitro* (Figure 7.5 A-B). A less pronounced decrease in cell growth was also observed for 697 cells while no differences were seen for the REH cell line (Figure 7.5 C-D). To further demonstrate the effect of SOX7 knock-down on proliferation, analysis of the cell cycle status was performed via EdU incorporation This analysis confirmed the results obtained in the proliferation assay, revealing a reduction of about 5% and 10% in the frequency of cells in the S-phase of the cell cycle for RS4;11 cells transduced with sh*SOX7*_2 and sh*SOX7*_3, respectively. This diminution in the frequency of cells in the phase of DNA replication of the cell cycle correlated with an increase in the frequency of cells in the G1 phase. The down-regulation of *SOX7* induced a decrease of 6% in the frequency of cells in the S-phase of the cell cycle in NALM6 cells. As expected, the decrease in cell frequency in the S-phase of the cell cycle was less in 697 cells while no significant differences were observed in REH cell line (Figure7.6).

Since leukaemia cells were proliferating less when *SOX7* was down-regulated, the next step was to determine whether the reduction in proliferation resulted from cells dying by apoptosis. To this end, cells were analysed by flow cytometry for Annexin-V expression to the phosphatidylserine on the extracellular membrane of apoptotic cells. Overall, cells were viable and were not going under apoptosis. An increased in the frequency of Annexin-$V^{+}7ADD^{+}$ cells was only observed in REH cells transduced with sh*SOX7*_2 (Figure 7.7) but did not correlate with a decrease in proliferation. Altogether, these results suggested that the down-regulation of *SOX7* in leukaemia cell lines, except in REH, induces a significant decrease in proliferation *in vitro*.

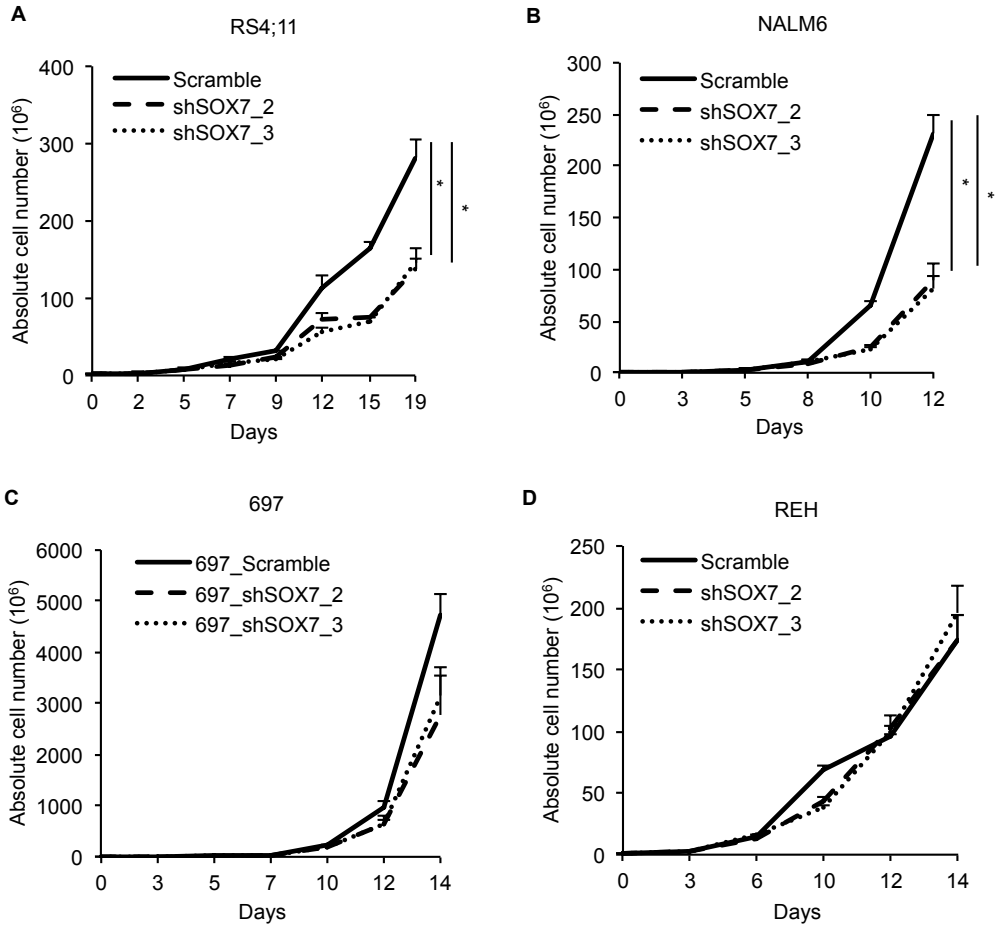


Figure 7.5. *SOX7* down-regulation induces a decrease in proliferation in human leukaemia cell lines.
Proliferation assay performed with RS4;11 (A), NALM6 (B), 697 (C) and REH (D) human leukaemia cells transduced with two different shRNAs against *SOX7* (sh*SOX7*_2, sh*SOX7*_3) or control construct (Scramble). A total of 0.25-1 million cells were plated in the presence of their growth medium. Cells were counted every 2-3 days and re-seeded in their growth medium for about 12 days or more. Data are shown as the mean of absolute cell number from three wells (n=3, * $P<0.05$). Graphs are representative of three independent experiments.

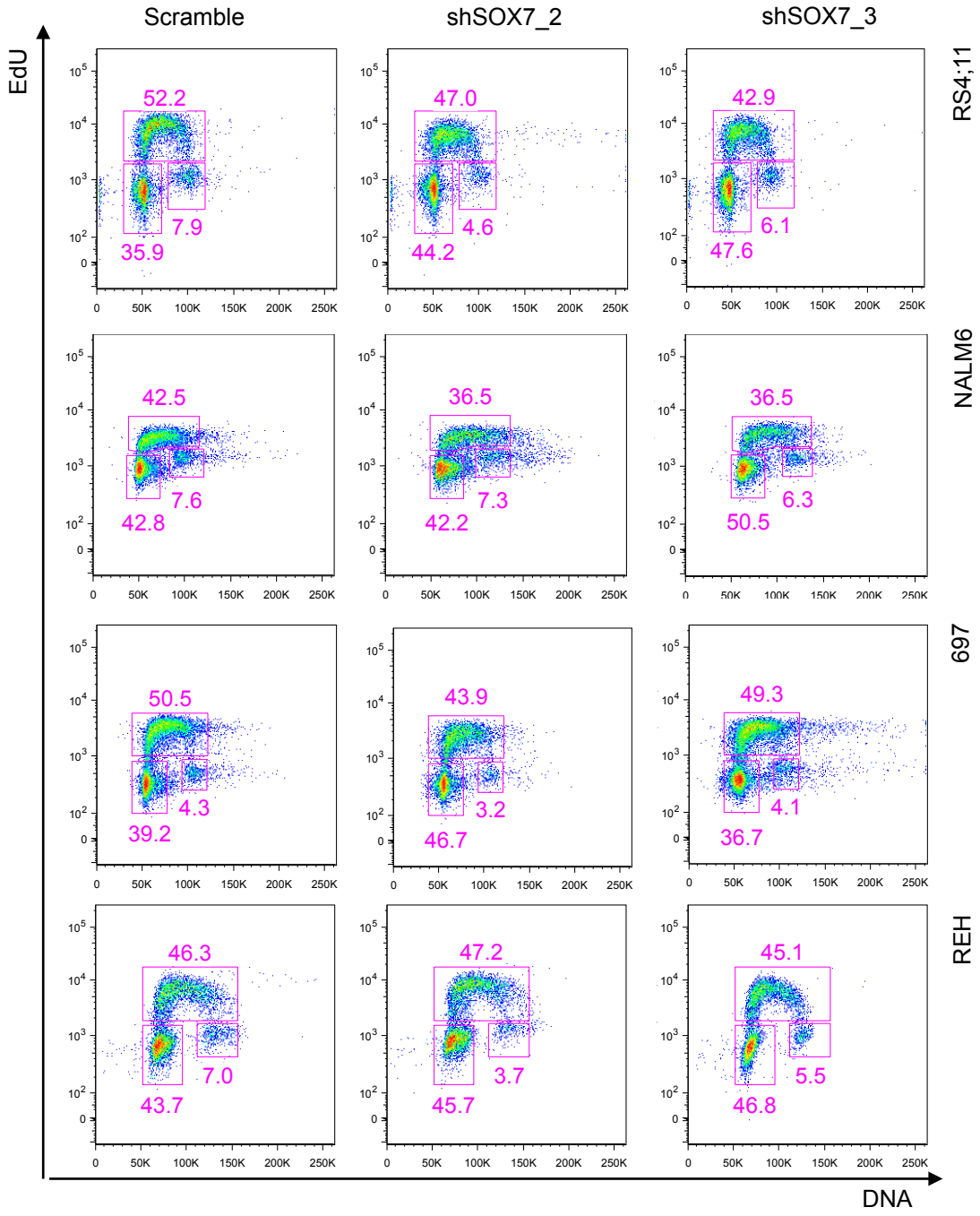


Figure 7.6. *SOX7* down-regulation induces a decrease in the frequency of cells in the S phase of the cell cycle in human leukaemia cell lines.
Cell cycle status of RS4;11, NALM6, 697 and REH human leukaemia cells transduced with two different shRNAs against *SOX7* (sh*SOX7*_2, sh*SOX7*_3) or control construct (Scramble) was assessed after 20 minutes to 2 hours of EdU incubation around 20 days after transduction. Percentages of cells in the G0/G1, S and G2 phases are shown in the indicated gates. FACS plots are representative of two independent experiments.

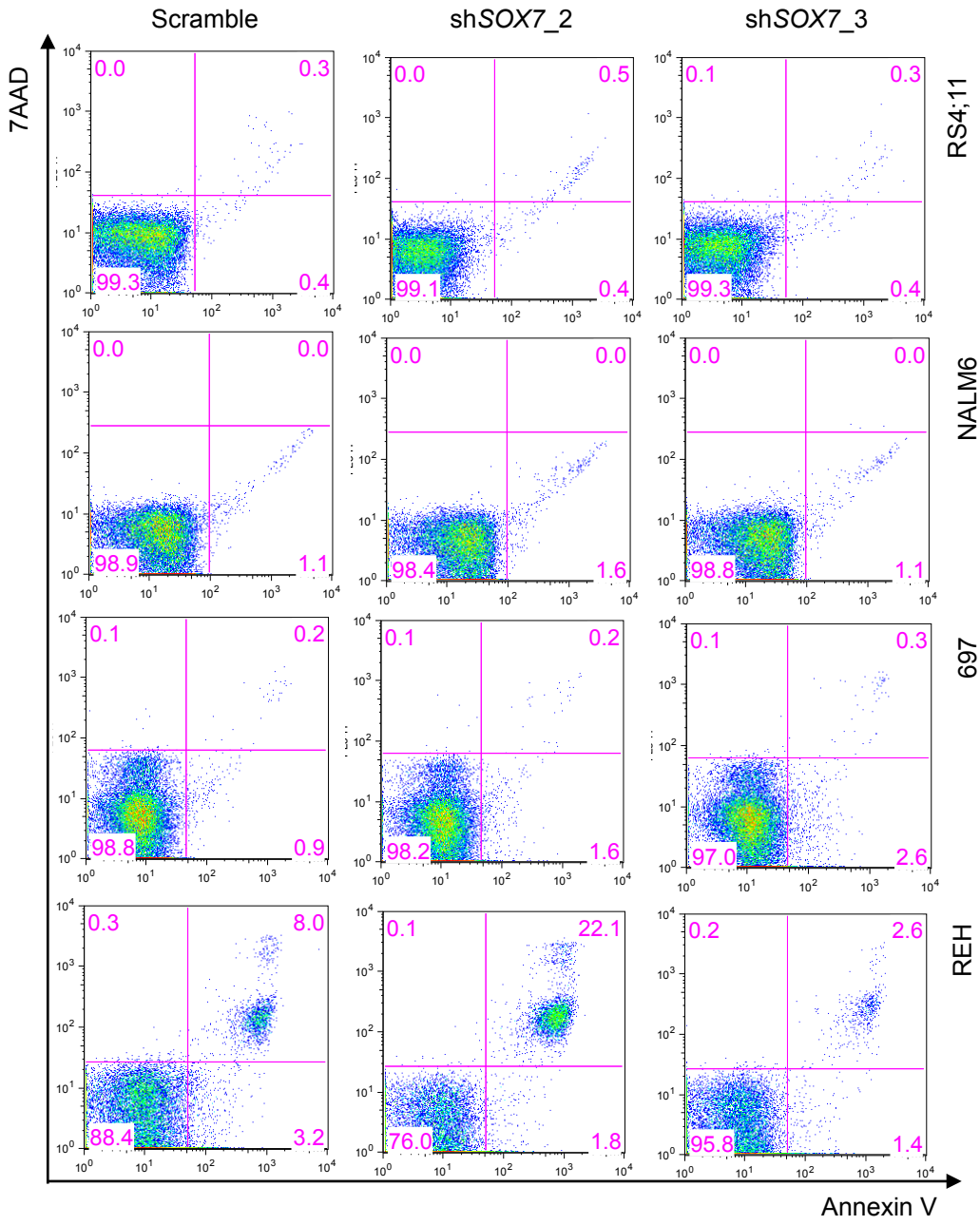


Figure 7.7. *SOX7* down-regulation does not induce apoptosis in human leukaemia cell lines.
Representative FACS plots of 7AAD and Annexin-V expression profiles on RS4;11, NALM6, 697 and REH human leukaemia cells transduced with two different shRNAs against *SOX7* (sh*SOX7*_2, sh*SOX7*_3) and or control construct (Scramble). The analysis was performed around 20 days after transduction. Percentages of cells are shown in the indicated quadrants. FACS plots are representative of two independent experiments.

7.2.2. Down-regulation of *SOX7* induces a significant decrease in colony formation

Next, to test whether *SOX7* down-regulation may induce the cells to differentiate or to lose their clonogenic capacity, transduced cell lines were cultured in semi-solid colony assay. After ten days in culture, the down-regulation of *SOX7* induced a significant decrease in colony formation in RS4;11, NALM6 and 967 cell lines, but not in REH (Figure 7.8 A-D). Moreover, the colonies formed by cells transduced with shRNA against *SOX7* were extremely reduced not only in number but also in size as shown in figure 7.9 A-D.

Altogether, these data revealed that the down-regulation of *SOX7* in leukaemia cell lines induces a significant decrease in colony formation.

7.2.3. Down-regulation of *SOX7* induces a significant increase in survival and a slight decrease in organs infiltration *in vivo*

Next, to investigate on the role of *SOX7* in leukaemogenesis *in vivo*, leukaemia cell lines transduced with shRNAs against *SOX7* and control were injected into non-irradiated immuno-compromised mice. Upon engraftment in recipient mice, the knock-down of *SOX7* expression in RS4;11 cell line induced a slight but significant increase in survival *in vivo* (Figure 7.10) in line with the reduced proliferation observed *in vitro*. In particular, an increase of 11 and 14 days was reached for the groups with the two different shRNAs, respectively.

In order to improve the down-regulation of *SOX7* expression, RS4;11 cells were transduced with both shRNAs against *SOX7*. The presence of the two shRNA together did not further improve *SOX7* knock-down or a decrease in proliferation (Figure 7.11). As these cells presented a similar level of GFP-luciferase expression as the scramble

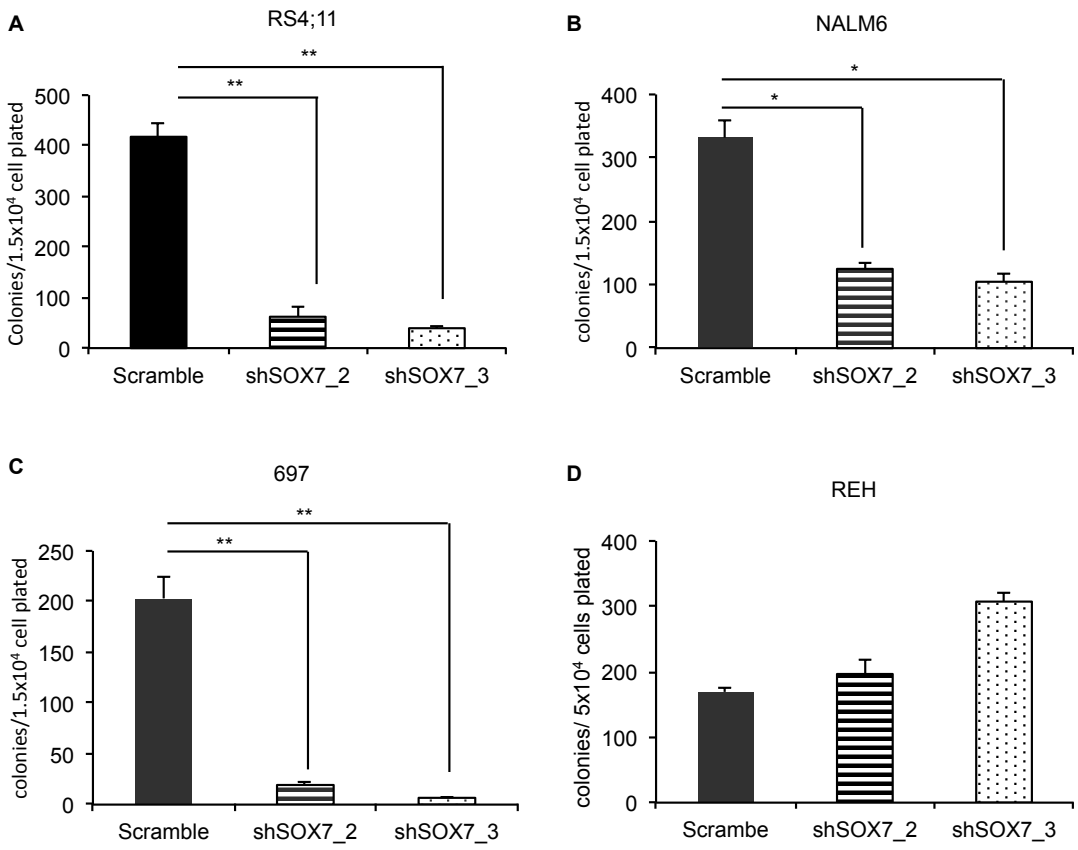


Figure 7.8. *SOX7* down-regulation induces a decrease in colony formation in human leukaemia cell lines.
Human leukaemia cell lines were plated in semi-solid culture and colonies were counted after 10 days in culture. The bar chart shows the colony number in RS4;11 (A), NALM6 (B), 697 (C) and REH (D). Error bars indicated mean ± standard deviation (n=3, * $P<0.05$, ** $P<0.01$). The analysis was performed around 20 days after transduction. Graphs are representative of three independent experiments.

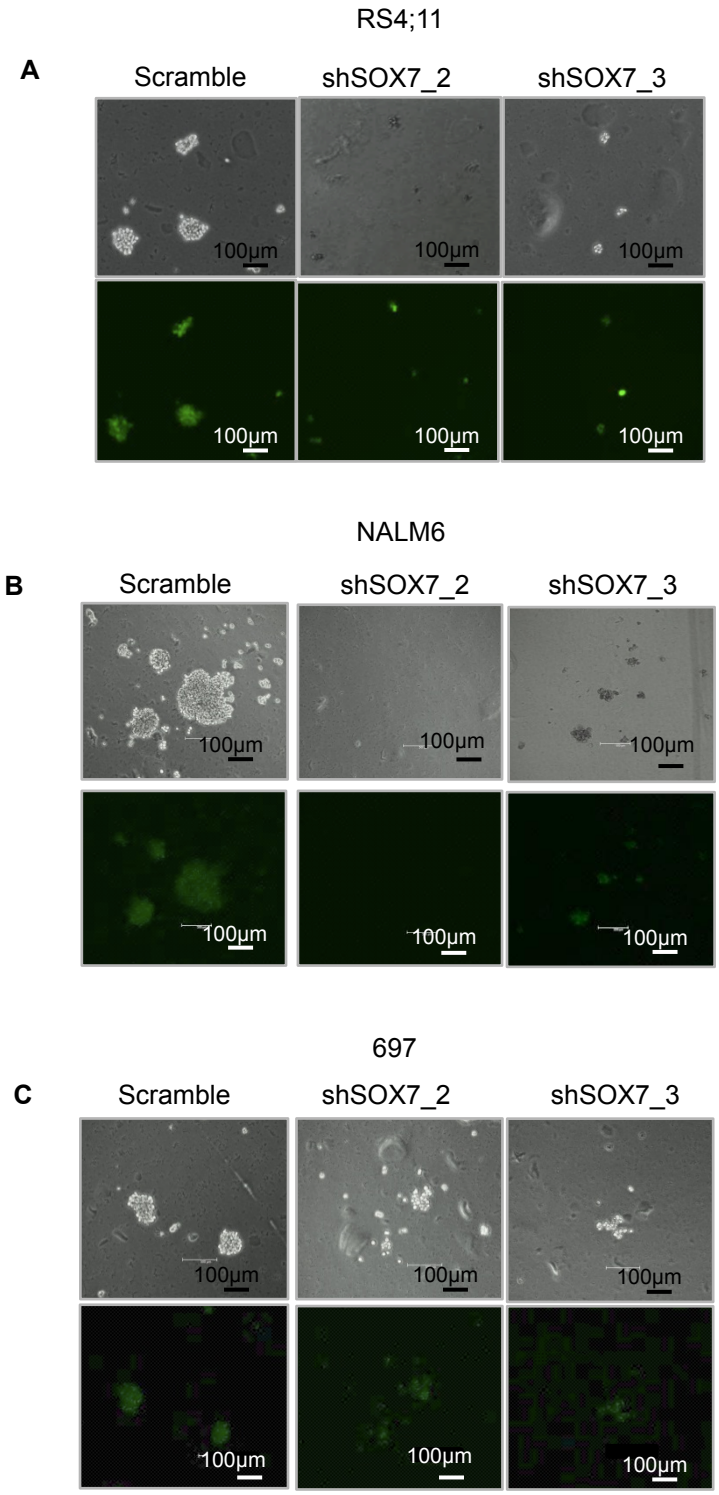


Figure 7.9. *SOX7* down-regulation induces a decrease in colony size in human leukaemia cell lines.
Representative picture of colonies obtained in semi-solid culture assay (Leica scan and software, scale bar: 100 µm). Pictures are representative of three independent experiments.

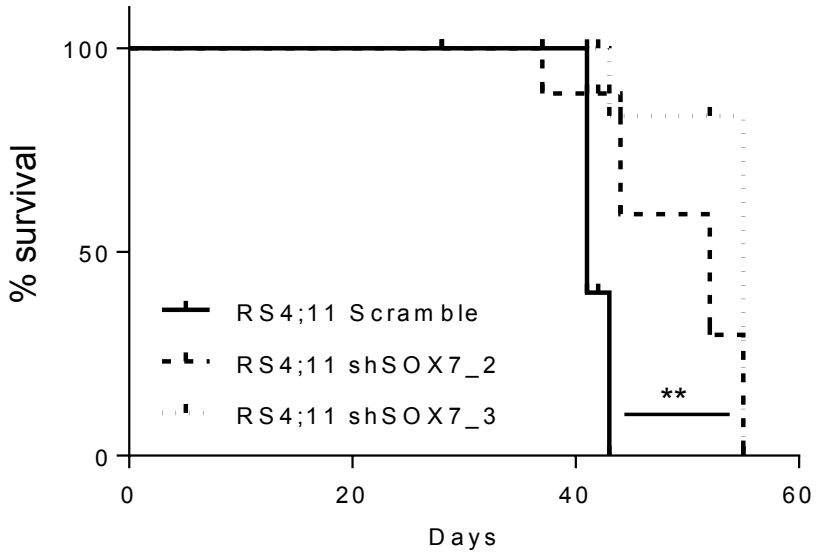


Figure 7.10. *SOX7* down-regulation in RS4;11 human leukaemia cell line induces an increase in survival *in vivo*.
RS4;11 leukaemia cells were transplanted intravenously into non-irradiated immuno-compromised mice and survival percentage was calculated using the Kaplan–Meier survival curve (n=4 mice per group). Log-rank (Mentel-Cox) test was used to analyze the P value (** $P<0.01$). Graph is representative of three independent experiments.

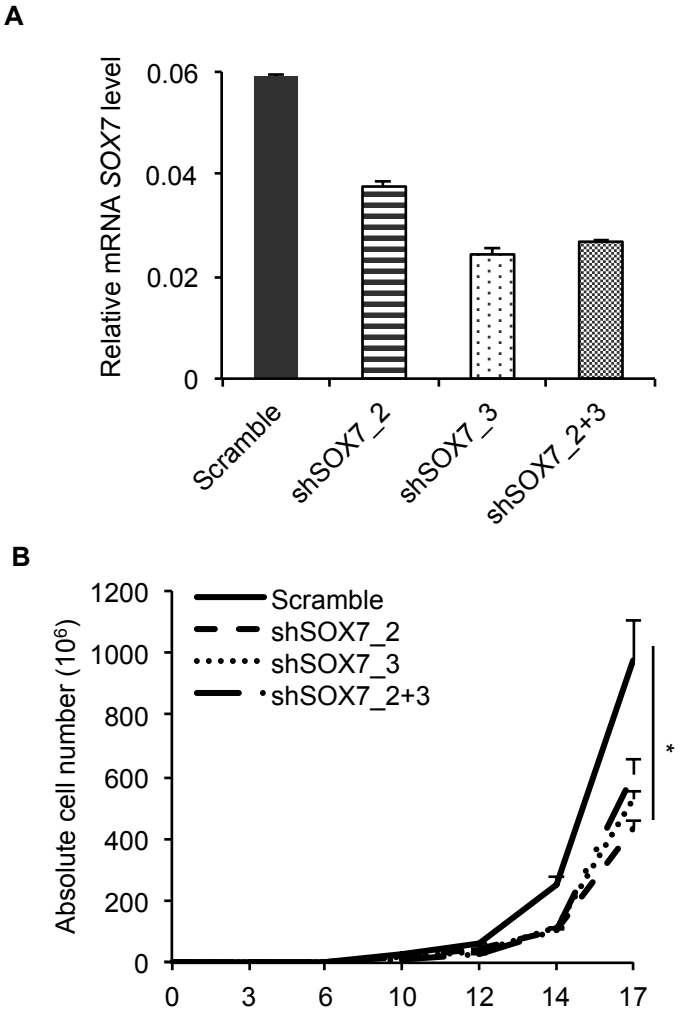


Figure 7.11. Transduction with both shRNAs does not improve the knock-down of *SOX7*.

A) Quantitative RT-PCR analysis of *SOX7* transcript level relative to β2microglobulin in RS4;11 human leukaemia cells transduced with one or two shRNAs against *SOX7* (sh*SOX7*_2, sh*SOX7*_3, sh*SOX7*_2+3). Leukaemia cells transduced with a control construct (Scramble) were used as a positive control. Error bars indicate mean ± standard deviation (n=3). B) Proliferation assay performed with RS4;11 human leukaemia cells transduced with one or two shRNAs against *SOX7* or control construct. A total of 0.25-1 million cells were plated in the presence of their growth medium. Data are shown as the mean of absolute cell number from three wells (n=3, * $P<0.05$). Graphs are representative of one independent experiment.

shRNA control cells (Figure 7.12), they were used for *in vivo* assays to monitor the early stages of leukaemia progression using an *in vivo* imaging system (IVIS) to monitor Luciferase activity. The chemical substance, Luciferin, penetrates the membrane of cell and when oxidized under the catalytic effects of Luciferase and ATP, a bluish-green light is produced (Hastings, 1996). Interestingly, the increase in survival corresponded with a decrease in proliferation and in organs infiltration *in vivo* (Figure 7.13). In particular, mice injected with RS4;11 cells transduced with both shRNAs showed signs of luciferase activity starting between 2 and 3 weeks after injection compared to the control shRNA that instead showed the first signs of luciferase activity already between one and two weeks after injection. Moreover, in the control group it was possible to observe the evolution of the leukaemia. First, leukaemic cells homed to the bone marrow, then they started to circulate throughout the body and finally they infiltrated the spleen. The analysis showed a massive bone infiltration, in particular along the spine and the skull from 3 weeks after injection. These IVIS results are extremely interesting because they might suggest an involvement of *SOX7* in migration but more analysis have to be done to confirm this hypothesis. On the contrary, mice injected with RS4;11 cells transduced with both shRNAs showed bone infiltration along the spine and the skull starting from week 4 only in two mice out of four. Importantly, at 6 weeks after the injection leukaemic cells were still spread throughout the body compared to the control mice where leukaemic cells completely infiltrated the spleen.

To further confirm the lower level of infiltration of leukaemic cells in the spleen upon *SOX7* knock-down, leukaemia cells were counted on spleen section from engrafted mice. The May-Grünwald staining showed leukaemia cells in dark purple due to their high nucleus:cytoplasm ratio (Figure 7.14 A). This analysis revealed a significant decrease in leukaemia cells infiltration in spleen from the groups engrafted

with the two different shRNAs compared to the control group (Figure 7.14 B).

Altogether, these data suggested that the down-regulation of *SOX7* induces a significant increase in survival and a delay in organs infiltration *in vivo*.

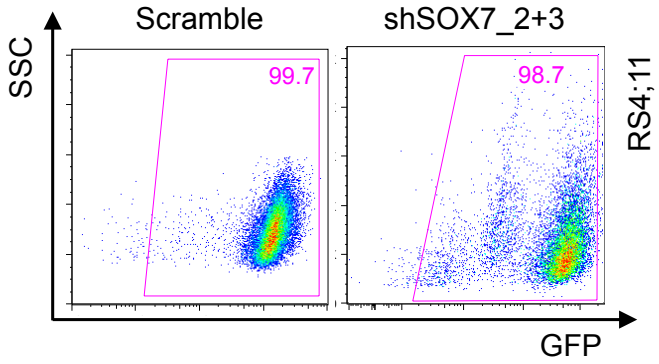


Figure 7.12. GFP expression in transduced leukaemia cell lines.
Representative FACS plots of SSC and GFP expression profiles of RS4;11 human leukaemia cells transduced with two different shRNAs against *SOX7* (sh*SOX7*_2+3) or control construct (Scramble). Percentages of cells are shown in the indicated quadrants. FACS plots are representative of two independent experiments.

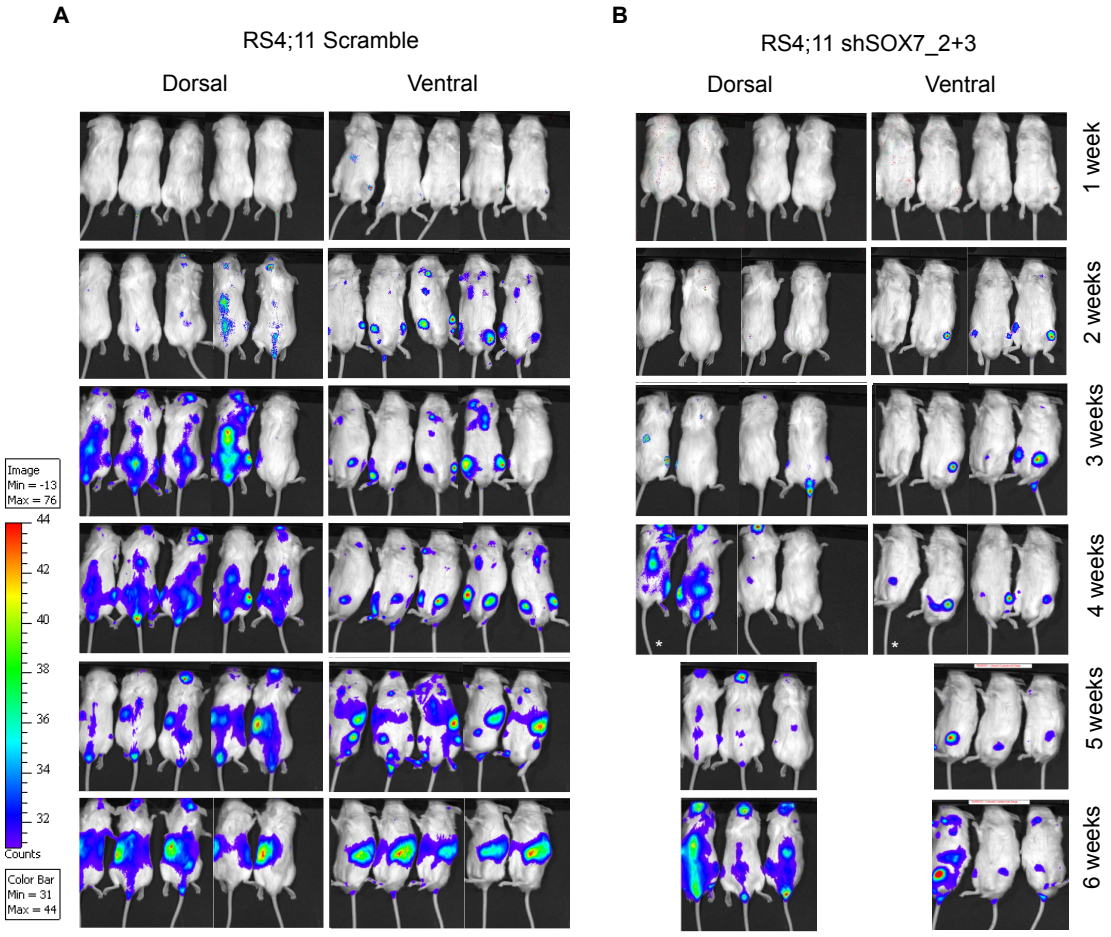


Figure 7.13. *SOX7* down-regulation in RS4;11 human leukaemia cell line induces a delay in organ infiltration *in vivo*.
Mice were injected intraperitoneally with a Luciferin solution (dose of 75 mg/kg; PerkinElmer). After mice were fully anesthetized, they were transferred into the imaging chamber (IVIS Lumina, Caliper Life Sciences). The imaging time is 30 seconds-1 minute per side (dorsal/ventral). A represents the control group (Scramble) while B the group with *SOX7* knock-down. One mouse (*) from the group with *SOX7* knock-down died between 4 and 5 weeks.

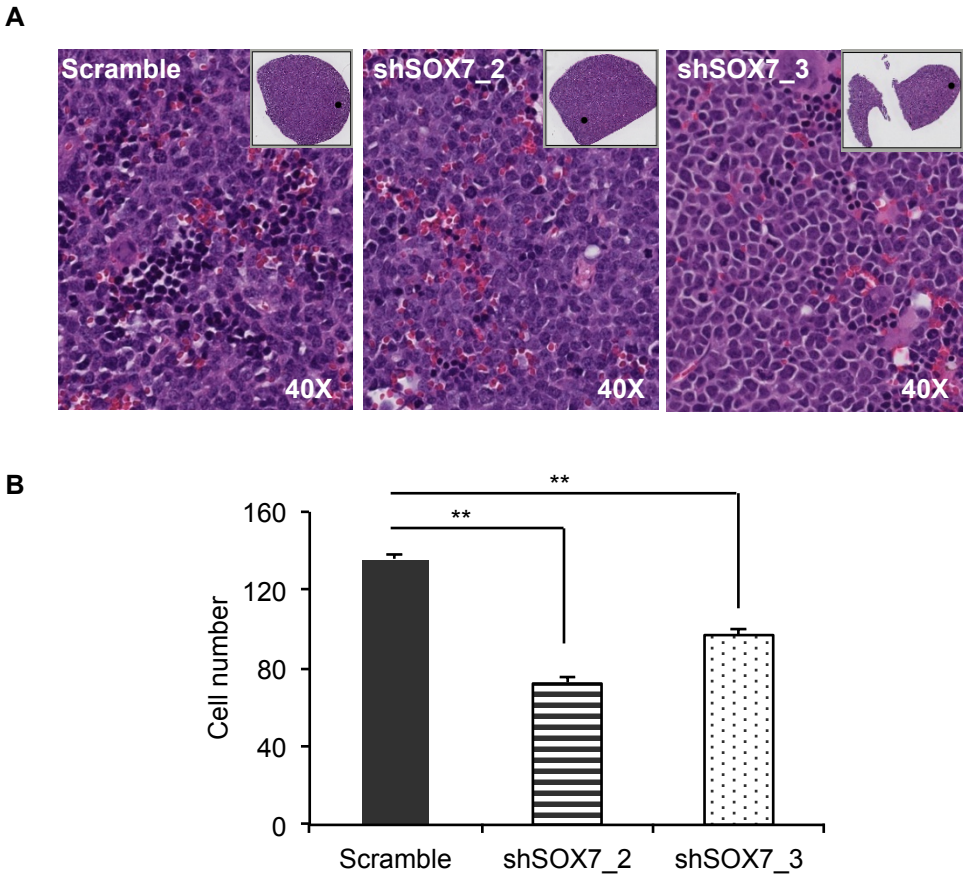


Figure 7.14. *SOX7* down-regulation in RS4;11 human leukaemia cell line induces a decrease in spleen infiltration *in vivo*.
A) May-Grünwald-Giemsa staining of spleen sections from mice engrafted with RS4;11 human leukaemia cells transduced with two different shRNAs against *SOX7* (sh*SOX7*_2 and sh*SOX7*_3) or control construct (Scramble). B) Bar chart shows the number of leukaemia cells per each section. Data are shown as the mean of cell number counted from four different areas in each section. Error bars indicate mean ± standard deviation (n=4, Leica scan and software, magnification: 40X). Graph is representative of three independent experiments.

7.3. Summary of finding

Since *SOX7* expression correlated with B-ALL and its over-expression in adult bone marrow cells induced B cell impairment, this part of the project has investigated the role for *SOX7* in leukaemogenesis. To this end, I have knocked-down the expression of *SOX7* in four different human leukaemia cell lines using two different shRNAs. Interestingly, the down-regulation of *SOX7* induced a significant decrease in proliferation *in vitro* in RS4;11, NALM6 and 697 cell lines. In contrast, this effect was not observed in the REH cell line. I reported that *SOX7* over-expression correlates with B-ALL but does not correlate with a specific type of translocation or mutation involved in B-ALL. Therefore, it is interestingly to notice no significant effects by *SOX7*-knock-down in REH cells. Further analyses will need to be done in order to verify whether this is linked to this specific cell line or to insufficient down-regulation of *SOX7*.

Upon engraftment in immuno-compromised mice, I observed that the knock-down of *SOX7* expression in RS4;11 cell line induced a slight but significant increase in survival and a delay in spleen infiltration *in vivo*.

To test whether the down-regulation of *SOX7* may induce cells to lose their clonogenic ability, I performed a colony assay. After ten days in culture, the down-regulation of *SOX7* induced a significant decrease in colony formation in RS4;11, NALM6 and 697 leukaemia cell lines.

To further characterise the knock-down of *SOX7* in human leukaemia cell lines I have performed a gene expression analysis. The analysis suggested a possible induction in B cell maturation upon *SOX7* knock-down in leukaemia cells explaining the drop off in colony formation in semi-solid culture. The microarray analysis also suggested a possible involvement of *SOX7* in the migration and cell movement confirming the delay in organ infiltration observed by IVIS imaging *in vivo*.

Together, these results suggest that the down-regulation of *SOX7* affected the proliferation of leukaemia cell lines both *in vitro* and *in vivo*.

Chapter 8

8. Discussion

8.1. SOX7 and the maintenance of immature progenitor cells

The haematopoietic program is driven by several transcription factors in response to specific signals. During the last decades, many transcriptional regulators, growth factors and signalling cascades have been shown to play a role in this process but deep understanding of the molecular mechanism underling the blood formation is required and other factors important for this process still remain to be identified. These findings will increase our ability to generate novel, more efficient therapies for blood malignancies such as leukaemia.

The Sex-determining region Y-box (SOX) family of transcription factors was first defined in mammals based on sequence conservation within the High Mobility Group (HMG) box initially identified in the testis-determining factor SRY (sex determining region Y) (Gubbay et al., 1990 Sinclair et al., 1990). Based on sequence and structural motif, the SOX family is divided into eight groups, from A to H. All SOX proteins share more than 50% amino acid similarity within the HMG box while at least 70% within members of the same group (Bowles et al., 2000). The SOX factors bind, through their HMG DNA-binding domain, specifically to the sequence motif 5'-(A/T)(A/T)CAA(A/T)G-3' (Bowles et al., 2000; Laudet et al., 1993; Stros et al., 2007).

The F group subfamily comprises 3 members: SOX7 together with its close

homologues SOX17 and SOX18. The murine *SoxF* genes are located on chromosome 14D, 1A1 and 2H4 for *Sox7*, *Sox17* and *Sox18*, respectively. Both human *SOX7* and *SOX17* are found on chromosome 8 while *SOX18* on chromosome 20 (Bowles et al., 2000).

Sox7 is expressed at the onset of blood specification in particular it marks haemogenic endothelium. Its enforced expression in $FLK1^{+}$ cells impairs the generation of mature blood cells sustaining the expression of endothelial markers (Costa et al., 2012). Previous work performed in our laboratory established that *Sox7*-sustained expression in haematopoietic precursors blocked differentiation while promoting self-renewal. Moreover, *Sox7* knock-down greatly impaired the generation of endothelial and haematopoietic precursors. Altogether, this work demonstrated the importance of SOX7, along with its close homologue SOX18, in controlling the balance between proliferation and differentiation at the onset of haematopoietic specification (Gandillet et al., 2009; Serrano et al., 2010).

Given the striking impact of *Sox7*-expression on embryonic haematopoiesis, I hypothesized that the mis-expression of *Sox7* in adult cells may confer a proliferative or survival advantage while blocking lineage maturation. In the present thesis, I reported the outcome of *Sox7* ectopic expression in cells derived from adult mice harbouring an inducible *Sox7* transgene. Using a serial clonogenic-replating assay, I demonstrated that the enforced *Sox7* expression in bone marrow cells induces an increase in colony formation together with increased proliferation. As expected, *Sox7*-expressing cells were positive for VE-cadherin but also for MAC1 and CD48. Interestingly, the immune-phenotype analysis of *Sox7*-expressing cells also revealed the presence of a sub-population of cells with different level of cKIT and GFP expression showing that the SOX7::$GFP^{high}cKIT^{+}$ sub-population contains most of the clonogenic potential.

Overall, these findings revealed that *Sox7*-enforced expression impairs haematopoietic differentiation as previously observed in embryonic haematopoiesis (Gandillet et al., 2009). Similarly, over-expression of *Sox17* in adult haematopoietic cells induced an expansion of haematopoietic progenitors *in vivo*. Moreover, *Sox17*-expressing $CD48^{+}$LSK cells expressed MAC1, VE-cadherin and AA4.1 foetal HSC markers (He et al., 2011).

Interestingly, the expression of VE-cadherin has been reported in $Philadelphia^{+}$ ALL cell lines and primary samples. Its down-regulation with siRNA or VE-cadherin antagonist induced an increase sensitive to chemotherapy agents while its over-expression in $Philadelphia^{-}$ ALL cell lines resulted in an increase in survival upon chemotherapy treatment (O'Leary et al., 2010). Since VE-cadherin is a transcriptional target of SOX7, at least in mouse, and since *SOX7* is widely expressed in B-ALL, it is tempting to speculate that SOX7, inducing the expression of VE-cadherin, might be implicated in poor prognosis B-ALL.

Since I discovered that *SOX7* is mainly expressed in B-ALL, I aimed to investigate the consequences of *Sox7*-enforced expression on adult B cell differentiation. The data described in this thesis revealed for the first time that the enforced expression of *Sox7* in adult bone marrow cells induces proliferation while blocking B lineage maturation. Similarly to the results obtained during the serial replating performed on semi-solid culture supplemented with myeloid cytokines, an increase in *Sox7*-expressing cells highly positive for VE-cadherin and cKIT was observed. The blockage during the B cell differentiation was at the pre-pro B stage or even earlier in the progenitor compartment. Moreover, a molecular profile analysis confirmed that the enforced expression of *Sox7* in adult bone marrow cells globally impaired B cell maturation while enhancing a stem cell-like identity.

In contrast to cells serially expanded in semi-solid culture, SOX7::GFP$^+$cKIT$^+$ cells cultured on OP9 were able to retain repopulating ability *in vivo*. This might be due to the culture conditions used and more probably due to the possible role for *Sox7* in B cell leukaemogenesis. In addition, SOX7::GFP$^+$cKIT$^+$ cells showed an higher frequency of engrafted cells when mice were fed with doxycycline. This might be explained by the maintenance of immature progenitor cells by doxycycline treatment.

Overall, these results are very interesting and clearly suggest a role for *Sox7* in the maintenance of an immature progenitor population.

8.2. SOX7 and B-ALL

Since the sustained expression of *Sox7* in early mouse haematopoietic precursors was sufficient to completely alter the balance between proliferation and differentiation at the onset of haematopoiesis, it was tempting to speculate that *SOX7* might be implicated in the emergence and/or maintenance of leukaemia.

Analysis of *SOX7* expression in multiple cases of human leukaemia revealed for the first time a tight correlation between *SOX7* expression and B-ALL. *SOX7* expression was heterogeneous across B-ALL samples and did not correlate with a specific type of translocation or mutation. Moreover, in order to verify whether *SOX7* expression correlated with specific sub-populations within B-ALL, primary samples were sorted based on the expression of CD19 and then on CD38 and CD34. *SOX7* was not specifically expressed at higher level in any of the sorted sub-populations suggesting the homogeneity of its expression through B-ALL sub-populations (data not shown).

Recently, it has been shown that *SOX7* is expressed in human splenic marginal zone B cells. Its expression is about eight times more than in naïve cells. Interestingly, microarray analysis also showed *SOX7* expression already in marginal zone progenitors

at low level but such level increased in marginal zone B cells. Moreover, murine MZB cells did not express *Sox7* (Descatoire et al., 2014). This study indicates the expression of *SOX7* in a specific subset of B cell cells in human and its absence in the corresponding subset in mouse, highlighting possible key differences in the regulation of SOX7 between mouse and human. It remains to be explored whether there is a link between SOX7 expression in MZB cells and the expression of this factor in B-ALL.

Differences between mice and human in B cell development have been already reported. In particular, IL7R is essential for B cell generation in mice but it is not for human where instead is important for T cell development (Giliani et al., 2005; Puel et al., 1998). Importantly, Descatoire's study reported again differences in gene expression in mouse and human in B cell sub-populations. These differences between mice and human in B cell development have to be considered when interpreting experiments.

To investigate the role for *SOX7* in leukaemogenesis, I have knocked-down the expression of *SOX7* in four different human leukaemia cell lines using two different shRNAs. Interestingly, the down-regulation of *SOX7* induced a significant decrease in proliferation *in vitro* in leukaemia cell lines. The incomplete decrease in proliferation by *SOX7* knock-down in these cell lines let me speculate that the leukaemia cells may require *SOX7* expression to growth and also that the down-regulation of *SOX7* was not stable or sufficient. Indeed, it seems that cells, grown in culture, adapted and most probably bypassed, through unknown mechanisms, the down-regulation of *SOX7*. Therefore, it will be important to find an alternative method to improve the knock-down of *SOX7*. The use of inducible shRNA construct might be a solution to homogenously and acutely silence *SOX7*.

Upon engraftment in immuno-compromised mice, I observed that the decrease in proliferation *in vitro* resulting from knocking-down *SOX7* expression in RS4;11 cell

line correlated with a significant increase in survival and a delay in spleen infiltration *in vivo*. Interestingly, the IVIS analysis *in vivo* suggested a possible involvement of *SOX7* in the migration and cell movement. This is an aspect of the project that will need to be further investigated.

Similar to *SOX7*, *SOX4* is expressed at high levels in ALL primary samples and cell lines while at low levels in AML (Ma et al., 2014). Analogous to the knock-down of *SOX7* in leukaemia cell line, the down-regulation of *Sox4* results in a comparable phenotype. Recently, two studies analysed the role for *Sox4* in acute leukaemia progression. Ma and colleagues deleted *Sox4* gene in pro B cells carrying the p190 BCR-ABL fusion protein. They observed a decrease in proliferation and colony formation *in vitro* and increase in survival *in vivo* (Ma et al., 2014). In addition, the deletion of *Sox4* in AML *C/ebpα* deficient LSK cells reduced significantly colony formation and induced myeloid maturation both *in vitro* and *in vivo* (Zhang et al., 2013).

Together, these studies and the results reported in this thesis suggest a role for *SOX* genes in acute leukaemia.

8.3. SOX7 and haematopoietic imbalance

In order to investigate the contribution of *Sox7*-enforced expression to haematopoietic malignancy, sub-lethally irradiated immuno-compromised mice were injected with total bone marrow derived from *iSox7* mouse. As observed *in vitro* culture, the B cell development was also impaired *in vivo*. Of interest, the over-expression of *Sox7* induced an increase in immature erythrocytes that reflected an increase in reticulocytes. Furthermore, the defects in the erythroid maturation have been observed in parallel to defect in platelets formation. These impairments in myeloid lineages maturation resemble to defects observed in *Runx1* deficient mice. *Runx1*$^{-/-}$ mice

died at E12.5 due to necrosis and haemorrhaging in the central nervous system. A complete block in definitive erythropoiesis, myelopoiesis and lymphopoiesis was reported as well suggesting a defect in an early haematopoietic progenitor cells (Okuda et al., 1996; Wang et al., 1996). Furthermore, in the absence of *Runx1*, haemogenic endothelium cells accumulate and the few haematopoietic cells produced are restricted to a primitive erythroid fate (Lacaud et al., 2002). A key role for RUNX1 in megakaryocyte development has been described through the conditional deletion of this factor in adult mice using *Mx1-Cre* mice. These experiments revealed the presence of immature megakaryocytes with low level of polyploidy in the bone marrow (Ichikawa et al., 2004). Microarray analysis performed on *Runx1*$^{-/-}$*Pf4-Cre,* where the deletion of *Runx1* occurred only in megakaryocytes, revealed the down-regulation of many genes (*Itga1/2*, *Mkl* and *Myl9*) involved in the megakaryopoiesis (Pencovich et al., 2013).

Importantly, in the set of *in vivo* experiments, reported in this thesis, the over-expression of *Sox7* occurred in all haematopoietic cells at all different maturation stages. The observed phenotype is the cumulative result of a series of events and therefore, these results have to be interpreted carefully. Overall, these data suggest a possible direct or more probably indirect role for *Sox7* in both erythroid and megakaryocytic lineages. However, further analyses will be necessary to demonstrate this hypothesis.

Next, an alteration of the organs morphology was observed upon *Sox7*-enforced expression. The spleen was enlarged showing an altered architecture. Enlargement of the spleen is a sign of leukaemia. Indeed, patients with acute leukaemia often show increased in spleen size (Pui, 2012b). Similarly to *Sox7*-enforced expression, mice reconstituted with *Sox17*-expressing cells exhibited signs of leukaemia showing increased spleen size together with extra-medullary haematopoiesis. In contrast, the

splenic architecture was normal in *Sox17*-enforced expressing samples (He et al., 2011).

Bone marrow failure upon *Sox7*-enforced expression was confirmed by the presence of reticulin fibres. Interestingly, fibrotic fibres were reported in the spleen as well. Diseases with increased bone marrow fibrosis are often associated with abnormalities in megakaryocyte and platelets production. The generation of collagen and reticulin fibres is mediated by the presence of cytokines such as transforming growth factor β (TGFβ), fibroblast growth factor (FGF), vascular endothelial growth factor (VEGF) and platelet-derived growth factor (PDGF) produced by megakaryocytes. Increased expression of FGF in megakaryocytes and increased expression of TGFβ receptor in endothelial cells have been reported in patients with myelofibrosis (Chou et al., 2003). Moreover, it has been shown that the two types of fibres have different clinical implication. Reticulin fibres show less strong correlation with disease severity compared to collagen fibres (Thiele and Kvasnicka, 2005). An increase in bone marrow fibres has also been described in patients with ALL or AML (Hann et al., 1978; Manoharan et al., 1979). Moreover, it has been shown that *SOX7* is expressed in patients with myelodisplastic syndrome but its role in fibrosis formation as not been investigated yet (Fan et al., 2012). Therefore, the presence of fibres in the bone marrow in mice engrafted with *Sox7*-expressing cells might be due to the impairment in megakaryocyte and platelet formation but further investigations will have to be performed.

Extra-medullary haematopoiesis was observed in secondary haematopoietic organs, spleen and liver. These organs were enriched in progenitor cells that retained long-term and multi-lineage engraftment capacity. These set of experiment further confirm the results obtained on semi-solid culture and on OP9 suggesting that enforced expression of *Sox7* maintains the proliferation of immature progenitor cells.

The continued over-expression of *Sox7* upon secondary transplantation did not lead to blood disease. No haematological malignancies were observed probably because mice were sacrificed before the appearance of the disease. Similarly and as mentioned previously, constitutive expression of *Sox17* in bone marrow cells was not sufficient to induce leukaemia; however, mice over-expressing *Sox17* died with features of leukaemia within 374 days after transplantation (He et al., 2011). In addition, the expression of *Sox7* was induced in all haematopoietic cells and created, as mentioned above, a mis-regulation of different blood lineages. Probably, in order to induce leukaemia sequential and critical events have to occur on a specific sub-population prone to promote such disease. For instance, since the in utero origin for some ALL has been reported (Greaves, 2003), infecting embryonic haematopoietic cells with retroviral construct expressing *Sox7* might be closer to recapitulate the events occurring during human leukaemia development. Also, the level of *Sox7* over-expression might be an important factor to consider for the development of a blood disease. Not all the engrafted cells were positive for the expression of GFP. This might suggest a possible silencing of the HPRT locus that, therefore, cannot be activated by doxycycline.

8.4. SOX7 and second hit to induce leukaemia

Since no haematological malignancies were observed upon *Sox7*-enforced expression, the requirement of a second hit seems necessary. I first analysed whether the deletion of *p53* might cooperate with *Sox7* over-expression to induce blood malignancy. P53 is a tumour suppressor protein that controls the apoptotic cell death response upon stress. Inhibition of *P53* results essential for the progression of many cancers (Vousden and Lu, 2002). In some cases of AML with *AML1-ETO* translocation, loss of *P53* is associated with tumour progression (Krejci et al., 2008) and in some cases of AML with

MLL fusion genes drug resistant is observed due to diminished P53 activity (Zuber et al., 2009). In ALL patients, *P53* mutations are observed in 8.2 per cent of cases and are detected mainly in patients negative for fusion genes (Chiaretti et al., 2013). In addition, monoallelic *P53* deletions have been reported in 11 per cent (5/46) ALL patients (Studniak et al., 2013).

The experiments carried out with loss of *p53* further confirmed that *Sox7*-enforced expression impairs haematopoietic differentiation and that the $cKIT^{+}GFP^{high}$ sub-population contains most clonogenic potential. The deletion of *p53* only increases the colony formation but when injected *in vivo* did not show sign of engraftment and therefore did not appear to confer leukaemia. Therefore, these experiments suggest that the deletion of *p53* does not cooperate with *Sox7* over-expression to induce blood malignancy.

8.5. SOX7 and B cell differentiation

To define a possible role for *Sox7* in murine adult lymphopoiesis, the function of *Sox7* was explored during B cell development by using conditional deletion of *Sox7* gene mediated by *CD19-Cre.* In this experimental set-up, *Cd19* promoter controls the expression of the Cre recombinase that deletes *Sox7* sequence, flanked by *LoxP* sites (Appendix 1 A). Therefore, the deletion of *Sox7* occurs only in CD19-expressing B cells.

After confirming the deletion of *Sox7* gene by genotyping, organs were analysed for any defects in lymphoid development (Appendix 1 B). No significant differences in the frequencies of B cells were reported between *CD19-Cre* mice and control in the analysed organs. Only a slight low level in the intensity of B cell markers was observed in *CD19-Cre* mice (Appendix 1 C). Overall, these preliminary results suggest that the

deletion of *Sox7* in B cells do not confer any impairment in B cell development. The absence of any defects might be explained by the fact that the expression of *Sox7* might be important in the early stages of B cell maturation therefore deleting *Sox7* in $CD19^+$ cells might be too late in the development. For this reason, a possible role for *Sox7* in early haematopoietic progenitor was analysed upon conditional deletion using the *Mx1-Cre* transgenic mice. In these mice, the deletion of *Sox7* occurs in all haematopoietic cells. The introduction of poly-inosinic-poly-cytidylic acid (plpC) by intra-peritoneal injection activates the *Mx1* promoter that drives the Cre recombinase expression (Appendix 2 A). Seven doses of plpC were injected in mice and peripheral blood was analysed 2 days and 30 days after the final dose to verify a complete deletion of *Sox7* gene. Deleted band was observed in both homozygous $Sox7^{fl/fl}/Mx1\text{-}Cre^+$ and heterozygous $Sox7^{wt/fl}/Mx1\text{-}Cre^+$. No band was observed for the Cre recombinase in the analysed samples suggesting an incomplete deletion of *Sox7* or defects in the PCR analysis. Then, 44 days after the final dose of plpC, organs were analysed for any defects in haematopoietic development and homeostasis. No significant differences in the frequencies of B cells were reported, both in the bone marrow and spleen, in mice: plpC treated heterozygous $Sox7^{wt/fl}/Mx1\text{-}Cre^-$ and homozygous $Sox7^{fl/fl}/Mx1\text{-}Cre^+$ and plpC untreated heterozygous $Sox7^{wt/fl}/Mx1\text{-}Cre^+$ mouse. A slight decrease in mature IgM^+B220^+ B cells was observed mainly in the bone marrow only in heterozygous $Sox7^{wt/fl}/Mx1\text{-}Cre^+$ mouse treated with plpC compared to the controls mice. Once again, the intensity of CD19 marker was slightly lower in the heterozygous $Sox7^{wt/fl}/Mx1\text{-}Cre^+$ mouse treated with plpC compared to the controls (Appendix 2 B-C). Interestingly, in this experimental set-up, a decrease in the intensity of B220 marker was also observed in the heterozygous $Sox7^{wt/fl}/Mx1\text{-}Cre^+$ mouse treated with plpC.

Together, these preliminary results showed that the deletion of *Sox7* in adult B

or HSC cells do not significantly impair mouse B cell development and homeostasis. Maybe, *Sox7* is important for B cell development in a different contest, in particular it might have a role during the early haematopoiesis in the foetal liver rather than in the adult lymphopoiesis. Also, *Sox7* might become important upon activation of B cells or upon stress. It might be interesting generate immune-responses in *Sox7*$^{fl/fl}$/*Mx1-Cre*$^{+}$ mice and analyse the B cell memory formation.

Therefore, changes in when and where to analyse the impact of *Sox7* deletion might reveal a function for *Sox7* in B cell development or activity.

8.6. How SOX7 blocks B cell differentiation?

The results presented in this thesis clearly suggest a role for *Sox7* in the maintenance of an immature progenitor population. However, the underlying mechanisms are still not clear. It has been reported in the literature that RUNX1 has a role in B cell development, in particular at the pre B cell stage. *Runx1* deficient mice showed a decrease in early B cell progenitors that is rescued by the expression of *Bcl2* (Niebuhr et al., 2013). Niebuhr and colleagues also reported the down-regulation of genes involved in the B cell development such as *Ebf1*, *Pax5*, *Src-kinases*, *Irf4/8*, *SpiB* and *Ikzf3* in *Runx1*-deficient cells (Niebuhr et al., 2013). Interestingly, these genes were also down-regulated in *Sox7*-expressing cells cultured on OP9. Of note, only *Blk* kinase gene was down-regulated upon *Sox7*-enforced expression while the expression of other genes, within the *Lyn* subfamily of SRC-family kinases, such as *Hck*, *Lck* and *Lyn* were decreased in *Runx1* null cells. The down-regulation of the SRC kinase, *Blk*, might be directly related to B cell impairment in *Sox7*-expressing cells. SRC kinases are important for the activation of the BCR signalling. Recently, it has been shown that BLK contributes to the BANK1-PLCγ2 interaction upon BCR stimulation. Its silencing

with siRNA in human B-cell line reduced the association between BANK1 and PLCγ2 (Bernal-Quiros et al., 2013). Therefore, a possible link between an impaired BCR signaling and the blockage of *Sox7*-expressing cells at the pre-pro B stage might be underlying the observed phenotype. However, further analyses of BCR signaling in bone marrow cells cultured on OP9 will need to be performed to test this hypothesis.

Another study on the role of RUNX1 in B cell lymphopoiesis showed in *Runx1*-deficient pro B cells an accumulation of the repressive mark, H3K27 trimethylation, at the *Ebf1* proximal promoter. In addition, RUNX1-CBFβ binding was observed on two RUNX recognition sites (RRS) on *Ebf1* proximal promoter (Seo et al., 2012). Moreover, unpublished data from our group suggest that RUNX1 and SOX7 can physically interact with each other (data not shown). Therefore, I hypothesized a possible mechanism of action where SOX7 binds RUNX1 preventing the activation of *EBF1* by RUNX1. To this end, I cloned the human *EBF1* proximal promoter sequence into a FireFly-Luciferase vector (Appendix 3 A) and performed Luciferase assay with human *SOX7*, human *RUNX1* and *Cbfβ* plasmids alone or in combination in HEK293 cells (Appendix 3 B). As reported in appendix 3, SOX7 can activate *EBF1* proximal promoter but less than RUNX1 and CBFβ. In particular, one-fold increase in Luciferase activity was observed by RUNX1 while more than 2-fold were reported by CBFβ compared to SOX7. All the combinations between the three factors activated *EBF1* proximal promoter whit no significant differences compared to CBFβ alone. Interestingly, the presence of SOX7 in combination with RUNX1 increased the Luciferase activity compared to RUNX1 alone. Unfortunately, these data do not support the formulated hypothesis but this preliminary result might suggest that SOX7 binding to *EBF1* proximal promoter. In addition, it is worth noting that the proposed mechanism might not work in HEK293 cells and therefore, similar experiments have to be

performed in specific context such as B cells.

8.7. Conclusive remarks and future works

The results described in present thesis revealed for the first time the correlation between *SOX7* expression and B-ALL. Moreover, the over-expression of *Sox7* in adult bone marrow cells impaired B cell maturation and induced an accumulation of immature progenitor cells both *in vitro* and *in vivo*. In addition, *Sox7*-enforced expression also impaired erythroid and megakaryocyte/platelets lineages. *Sox7*-enforced expression induced splenomegaly, bone marrow failure with fibrosis and extra-medullary haematopoiesis *in vivo*. Importantly, the down-regulation of *SOX7* in human leukaemia cell line induced a decrease in proliferation *in vitro* that correlates with an increase in survival *in vivo*.

To extend and complement the data presented in this thesis, further experiments will have to be performed to address the role of *SOX7* in the proliferative and leukaemogenic potential observed in B-ALL cell lines. Preliminary *in vivo* studies with NALM6 and 697 cell lines have been initiated. Results were not as successful as with the RS4;11 cell line (data not shown). This might be linked to the different mutations carried by these cells that confer a very high proliferative potential in these specific leukaemia cell lines. Therefore, since the expression of *SOX7* seems essential for the proliferation of leukaemic cells, it is important to down-regulate *SOX7* enough to affect proliferation. To better control the knock-down of *SOX7*, these B-ALL lines will be transduced with scramble or *SOX7* shRNA using an inducible system. Subsequently, leukaemic cells will be analyzed *in vitro* for proliferation and clonogenicity as well as *in vivo* for leukaemogenic potential upon the acute deletion of *SOX7* expression. Similar experiments will also be performed in primary B-ALL samples.

These complementary experiments should possibly help us to establish conclusively the critical role of *SOX7* in promoting proliferation and leukaemogenesis in other B-ALL cell lines and primary samples. Moreover, since no haematological malignancies have been observed upon *Sox7*-enforced expression in mouse models, the requirement for other mutational events might be necessary. One candidate might be the fusion gene *MLL-AF4* since it expression is not sufficient to recapitulate the human disease when expressed on its own. To this end, I cloned the human *MLL-AF4* translocation in a retroviral vector that contains the *Tomato* reporter gene under the control of the *Ef1* promoter (Appendix 4). Preliminary experiments were performed to test the transduction efficiency in $cKIT^+$ mouse bone marrow cells. A new set of experiments will be carrying out on progenitor bone marrow cells extracted from inducible *Sox7* mice. Then, transduced cells will be engrafted in recipient mice and *Sox7* over-expression will be induced by doxycycline. These *in vivo* experiments should define a possible co-operation between SOX7 and MLL-AF4 for leukaemia induction.

To identify the downstream effectors of *SOX7* function in B-ALL, global transcriptome changes occurring upon *SOX7* knock-down in B-ALL cell lines using the Affymetrix All-Exon array platform or RNA-seq will be performed. Comparative analysis will allow the identification of genes specifically modulated upon *SOX7* down-regulation and common between different B-ALL cell lines. To complement this analysis and determine SOX7 direct targets, SOX7 chromatin immuno-precipitation followed by massive parallel DNA sequencing (ChIP-seq) will be performed in the B-ALL cell lines. The combined analysis of transcriptome and ChIP-seq datasets should define a list of target genes regulated by SOX7 in B-ALL cell lines.

Altogether, these experiments will help to establish a list of candidate genes likely involved in the molecular program controlling the balance between differentiation

and proliferation of haematopoietic precursors. Finally, identifying the molecular mechanism underlying SOX7 activity will help to define novel potentially druggable targets.

References

Alexander, J., and Stainier, D. Y. (1999). A molecular pathway leading to endoderm formation in zebrafish. Current biology : CB *9*, 1147-1157.

Allman, D., Lindsley, R. C., DeMuth, W., Rudd, K., Shinton, S. A., and Hardy, R. R. (2001). Resolution of three nonproliferative immature splenic B cell subsets reveals multiple selection points during peripheral B cell maturation. Journal of immunology *167*, 6834-6840.

An, X., Tiwari, A. K., Sun, Y., Ding, P. R., Ashby, C. R., Jr., and Chen, Z. S. (2010). BCR-ABL tyrosine kinase inhibitors in the treatment of Philadelphia chromosome positive chronic myeloid leukemia: a review. Leukemia research *34*, 1255-1268.

Andreasson, P., Schwaller, J., Anastasiadou, E., Aster, J., and Gilliland, D. G. (2001). The expression of ETV6/CBFA2 (TEL/AML1) is not sufficient for the transformation of hematopoietic cell lines in vitro or the induction of hematologic disease in vivo. Cancer genetics and cytogenetics *130*, 93-104.

Aoki, T. O., David, N. B., Minchiotti, G., Saint-Etienne, L., Dickmeis, T., Persico, G. M., Strahle, U., Mourrain, P., and Rosa, F. M. (2002). Molecular integration of casanova in the Nodal signalling pathway controlling endoderm formation. Development (Cambridge, England) *129*, 275-286.

Aue, G., Du, Y., Cleveland, S. M., Smith, S. B., Dave, U. P., Liu, D., Weniger, M. A., Metais, J. Y., Jenkins, N. A., Copeland, N. G., and Dunbar, C. E. (2011). Sox4 cooperates with PU.1 haploinsufficiency in murine myeloid leukemia. Blood *118*, 4674-4681.

Azhikina, T., Kozlova, A., Skvortsov, T., and Sverdlov, E. (2011). Heterogeneity and degree of TIMP4, GATA4, SOX18, and EGFL7 gene promoter methylation in non-small cell lung cancer and surrounding tissues. Cancer genetics *204*, 492-500.

Bain, G., Maandag, E. C., Izon, D. J., Amsen, D., Kruisbeek, A. M., Weintraub, B. C., Krop, I., Schlissel, M. S., Feeney, A. J., van Roon, M., and et al. (1994). E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. Cell *79*, 885-892.

Bao, Y., Liu, X., Han, C., Xu, S., Xie, B., Zhang, Q., Gu, Y., Hou, J., Qian, L., Qian, C.*, et al.* (2014). Identification of IFN-gamma-producing innate B cells. Cell research *24*, 161-176.

Bass, A. J., Watanabe, H., Mermel, C. H., Yu, S., Perner, S., Verhaak, R. G., Kim, S. Y., Wardwell, L., Tamayo, P., Gat-Viks, I.*, et al.* (2009). SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nature genetics *41*, 1238-1242.

Beck, K., Peak, M. M., Ota, T., Nemazee, D., and Murre, C. (2009). Distinct roles for E12 and E47 in B cell specification and the sequential rearrangement of immunoglobulin light chain loci. The Journal of experimental medicine *206*, 2271-2284.

Bemark, M., Holmqvist, J., Abrahamsson, J., and Mellgren, K. (2012). Translational Mini-Review Series on B cell subsets in disease. Reconstitution after haematopoietic stem cell transplantation - revelation of B cell developmental pathways and lineage phenotypes. Clinical and experimental immunology *167*, 15-25.

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

Bernal-Quiros, M., Wu, Y. Y., Alarcon-Riquelme, M. E., and Castillejo-Lopez, C. (2013). BANK1 and BLK act through phospholipase C gamma 2 in B-cell signaling. PloS one *8*, e59842.

Bernardin, F., Yang, Y., Cleaves, R., Zahurak, M., Cheng, L., Civin, C. I., and Friedman, A. D. (2002). TEL-AML1, expressed from t(12;21) in human acute lymphocytic leukemia, induces acute leukemia in mice. Cancer research *62*, 3904-3908.

Bertrand, J. Y., Kim, A. D., Violette, E. P., Stachura, D. L., Cisson, J. L., and Traver, D. (2007). Definitive hematopoiesis initiates through a committed erythromyeloid progenitor in the zebrafish embryo. Development (Cambridge, England) *134*, 4147-4156.

Betz, B. L., and Hess, J. L. (2010). Acute myeloid leukemia diagnosis in the 21st century. Archives of pathology & laboratory medicine *134*, 1427-1433.

Bhayat, F., Das-Gupta, E., Smith, C., McKeever, T., and Hubbard, R. (2009). The incidence of and mortality from leukaemias in the UK: a general population-based study. BMC cancer *9*, 252.

Bijl, J., Sauvageau, M., Thompson, A., and Sauvageau, G. (2005). High incidence of proviral integrations in the Hoxa locus in a new model of E2a-PBX1-induced B-cell leukemia. Genes & development *19*, 224-233.

Boiers, C., Carrelha, J., Lutteropp, M., Luc, S., Green, J. C., Azzoni, E., Woll, P. S., Mead, A. J., Hultquist, A., Swiers, G*., et al.* (2013). Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells. Cell stem cell *13*, 535-548.

Boisset, J. C., van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E., and Robin, C. (2010). In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. Nature *464*, 116-120.

Bomken, S., Fiser, K., Heidenreich, O., and Vormoor, J. (2010). Understanding the cancer stem cell. British journal of cancer *103*, 439-445.

Bowles, J., Schepers, G., and Koopman, P. (2000). Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Developmental biology *227*, 239-255.

Brotherton, T. W., Chui, D. H., Gauldie, J., and Patterson, M. (1979). Hemoglobin ontogeny during normal mouse fetal development. Proceedings of the National Academy of Sciences of the United States of America *76*, 2853-2857.

Bueno, C., Montes, R., Melen, G. J., Ramos-Mejia, V., Real, P. J., Ayllon, V., Sanchez, L., Ligero, G., Gutierrez-Aranda, I., Fernandez, A. F.*, et al.* (2012). A human ESC model for MLL-AF4 leukemic fusion gene reveals an impaired early hematopoietic-endothelial specification. Cell research *22*, 986-1002.

Butler, T., and Gribben, J. G. (2010). Biologic and clinical significance of molecular profiling in Chronic Lymphocytic Leukemia. Blood reviews *24*, 135-141.

Carvalho, T. L., Mota-Santos, T., Cumano, A., Demengeot, J., and Vieira, P. (2001). Arrested B lymphopoiesis and persistence of activated B cells in adult interleukin 7(-/)-mice. The Journal of experimental medicine *194*, 1141-1150.

Castellanos, A., Pintado, B., Weruaga, E., Arevalo, R., Lopez, A., Orfao, A., and Sanchez-Garcia, I. (1997). A BCR-ABL(p190) fusion gene made by homologous recombination causes B-cell acute lymphoblastic leukemias in chimeric mice with independence of the endogenous bcr product. Blood *90*, 2168-2174.

Castor, A., Nilsson, L., Astrand-Grundstrom, I., Buitenhuis, M., Ramirez, C., Anderson, K., Strombeck, B., Garwicz, S., Bekassy, A. N., Schmiegelow, K., et al. (2005). Distinct patterns of hematopoietic stem cell involvement in acute lymphoblastic leukemia. Nature medicine *11*,630-637.

Cermenati, S., Moleri, S., Cimbro, S., Corti, P., Del Giacco, L., Amodeo, R., Dejana, E., Koopman, P., Cotelli, F., and Beltrame, M. (2008). Sox18 and Sox7 play redundant roles in vascular development. Blood *111*, 2657-2666.

Chan, D. W., Mak, C. S., Leung, T. H., Chan, K. K., and Ngan, H. Y. (2012). Down-regulation of Sox7 is associated with aberrant activation of Wnt/b-catenin signaling in endometrial cancer. Oncotarget *3*, 1546-1556.

Chen, K., Huang, Y. H., and Chen, J. L. (2013). Understanding and targeting cancer stem cells: therapeutic implications and challenges. Acta pharmacologica Sinica *34*, 732-740.

Chen, W., Li, Q., Hudson, W. A., Kumar, A., Kirchhof, N., and Kersey, J. H. (2006). A murine Mll-AF4 knock-in model results in lymphoid and myeloid deregulation and hematologic malignancy. Blood *108*, 669-677.

Chiabrando, D., Mercurio, S., and Tolosano, E. (2014). Heme and erythropoieis: more than a structural role. Haematologica *99*, 973-983.

Chiaretti, S., Brugnoletti, F., Tavolaro, S., Bonina, S., Paoloni, F., Marinelli, M., Patten, N., Bonifacio, M., Kropp, M. G., Sica, S.*, et al.* (2013). TP53 mutations are frequent in adult acute lymphoblastic leukemia cases negative for recurrent fusion genes and correlate with poor response to induction therapy. Haematologica *98*, e59-61.

Chimonidou, M., Strati, A., Malamos, N., Georgoulias, V., and Lianidou, E. S. (2013). SOX17 promoter methylation in circulating tumor cells and matched cell-free DNA isolated from plasma of patients with breast cancer. Clinical chemistry *59*, 270-279.

Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J. C., and Keller, G. (1998). A common precursor for hematopoietic and endothelial cells. Development (Cambridge, England) *125*, 725-732.

Chou, J. M., Li, C. Y., and Tefferi, A. (2003). Bone marrow immunohistochemical studies of angiogenic cytokines and their receptors in myelofibrosis with myeloid metaplasia. Leukemia research *27*, 499-504.

Chun, J. J., Schatz, D. G., Oettinger, M. A., Jaenisch, R., and Baltimore, D. (1991). The recombination activating gene-1 (RAG-1) transcript is present in the murine central nervous system. Cell *64*, 189-200.

Ciau-Uitz, A., Monteiro, R., Kirmizitas, A., and Patient, R. (2014). Developmental hematopoiesis: Ontogeny, genetic programming and conservation. Experimental hematology *42*, 669-683.

Clark, M. R., Mandal, M., Ochiai, K., and Singh, H. (2014). Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling. Nature reviews Immunology *14*, 69-80.

Clarke, R. L., Yzaguirre, A. D., Yashiro-Ohtani, Y., Bondue, A., Blanpain, C., Pear, W. S., Speck, N. A., and Keller, G. (2013). The expression of Sox17 identifies and regulates haemogenic endothelium. Nature cell biology *15*, 502-510.

Claus, R., Almstedt, M., and Lubbert, M. (2005). Epigenetic treatment of hematopoietic malignancies: in vivo targets of demethylating agents. Seminars in oncology *32*, 511-520.

Clayton, E., Bardi, G., Bell, S. E., Chantry, D., Downes, C. P., Gray, A., Humphries, L. A., Rawlings, D., Reynolds, H., Vigorito, E., and Turner, M. (2002). A crucial role for the p110delta subunit of phosphatidylinositol 3-kinase in B cell development and activation. The Journal of experimental medicine *196*, 753-763.

Collins-Underwood, J. R., and Mullighan, C. G. (2010). Genomic profiling of high-risk acute lymphoblastic leukemia. Leukemia *24*, 1676-1685.

Copp, A. J. (1995). Death before birth: clues from gene knockouts and mutations. Trends in genetics : TIG *11*, 87-93.

Corberand, J. X. (1996). Reticulocyte analysis using flow cytometry. Hematology and cell therapy *38*, 487-494.

Corfe, S. A., and Paige, C. J. (2012). The many roles of IL-7 in B cell development; mediator of survival, proliferation and differentiation. Seminars in immunology *24*, 198-208.

Costa, G., Kouskoff, V., and Lacaud, G. (2012a). Origin of blood cells and HSC production in the embryo. Trends in immunology *33*, 215-223.

Costa, G., Mazan, A., Gandillet, A., Pearson, S., Lacaud, G., and Kouskoff, V. (2012b). SOX7 regulates the expression of VE-cadherin in the haemogenic endothelium at the onset of haematopoietic development. Development (Cambridge, England) *139*, 1587-1598.

Cumano, A., Dorshkind, K., Gillis, S., and Paige, C. J. (1990). The influence of S17 stromal cells and interleukin 7 on B cell development. European journal of immunology *20*, 2183-2189.

Cumano, A., and Godin, I. (2007). Ontogeny of the hematopoietic system. Annual review of immunology *25*, 745-785.

de Bruijn, M. F., Ma, X., Robin, C., Ottersbach, K., Sanchez, M. J., and Dzierzak, E. (2002). Hematopoietic stem cells localize to the endothelial cell layer in the midgestation mouse aorta. Immunity *16*, 673-683.

Dedera, D. A., Waller, E. K., LeBrun, D. P., Sen-Majumdar, A., Stevens, M. E., Barsh, G. S., and Cleary, M. L. (1993). Chimeric homeobox gene E2A-PBX1 induces proliferation, apoptosis, and malignant lymphomas in transgenic mice. Cell *74*, 833-843.

DeKoter, R. P., Lee, H. J., and Singh, H. (2002). PU.1 regulates expression of the interleukin-7 receptor in lymphoid progenitors. Immunity *16*, 297-309.

DeKoter, R. P., and Singh, H. (2000). Regulation of B lymphocyte and macrophage development by graded expression of PU.1. Science *288*, 1439-1441.

Descatoire, M., Weller, S., Irtan, S., Sarnacki, S., Feuillard, J., Storck, S., Guiochon-Mantel, A., Bouligand, J., Morali, A., Cohen, J.*, et al.* (2014). Identification of a human splenic marginal zone B cell precursor with NOTCH2-dependent differentiation properties. The Journal of experimental medicine *211*, 987-1000.

Dieterlen-Lievre, F. (1975). On the origin of haemopoietic stem cells in the avian embryo: an experimental approach. Journal of embryology and experimental morphology *33*, 607-619.

Downs, K. M. (2003). Florence Sabin and the mechanism of blood vessel lumenization during vasculogenesis. Microcirculation *10*, 5-25.

Downs, K. M., Hellman, E. R., McHugh, J., Barrickman, K., and Inman, K. E. (2004). Investigation into a role for the primitive streak in development of the murine allantois. Development (Cambridge, England) *131*, 37-55.

Du, Y. C., Oshima, H., Oguma, K., Kitamura, T., Itadani, H., Fujimura, T., Piao, Y. S., Yoshimoto, T., Minamoto, T., Kotani, H.*, et al.* (2009). Induction and down-regulation of Sox17 and its possible roles during the course of gastrointestinal tumorigenesis. Gastroenterology *137*, 1346-1357.

Dunnick, W., Hertz, G. Z., Scappino, L., and Gritzmacher, C. (1993). DNA sequences at immunoglobulin switch region recombination sites. Nucleic acids research *21*, 365-372.

Duong, T., Proulx, S. T., Luciani, P., Leroux, J. C., Detmar, M., Koopman, P., and Francois, M. (2012). Genetic ablation of SOX18 function suppresses tumor lymphangiogenesis and metastasis of melanoma in mice. Cancer research *72*, 3105-3114.

Dzierzak, E. (2005). The emergence of definitive hematopoietic stem cells in the mammal. Current opinion in hematology *12*, 197-202.

Dzierzak, E., and Speck, N. A. (2008). Of lineage and legacy: the development of mammalian hematopoietic stem cells. Nature immunology *9*, 129-136.

Eilken, H. M., Nishikawa, S., and Schroeder, T. (2009). Continuous single-cell imaging of blood generation from haemogenic endothelium. Nature *457*, 896-900.

Ema, H., Morita, Y., and Suda, T. (2014). Heterogeneity and hierarchy of hematopoietic stem cells. Experimental hematology *42*, 74-82 e72.

Fan, R., Zhang, L. Y., Wang, H., Yang, B., Han, T., Zhao, X. L., Wang, W., Wang, X. Q., and Lin, G. W. (2012). Methylation of the CpG island near SOX7 gene promoter is correlated with the poor prognosis of patients with myelodysplastic syndrome. The Tohoku journal of experimental medicine *227*, 119-128.

Fehling, H. J., Lacaud, G., Kubo, A., Kennedy, M., Robertson, S., Keller, G., and Kouskoff, V. (2003). Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. Development (Cambridge, England) *130*, 4217-4227.

Finicelli, M., Benedetti, G., Squillaro, T., Pistilli, B., Marcellusi, A., Mariani, P., Santinelli, A., Latini, L., Galderisi, U., and Giordano, A. (2014). Expression of stemness genes in primary breast cancer tissues: the role of SOX2 as a prognostic marker for detection of early recurrence. Oncotarget.

Fischer, M., Schwieger, M., Horn, S., Niebuhr, B., Ford, A., Roscher, S., Bergholz, U., Greaves, M., Lohler, J., and Stocking, C. (2005). Defining the oncogenic function of the TEL/AML1 (ETV6/RUNX1) fusion protein in a mouse model. Oncogene *24*, 7579-7591.

Francois, M., Caprini, A., Hosking, B., Orsenigo, F., Wilhelm, D., Browne, C., Paavonen, K., Karnezis, T., Shayan, R., Downes, M.*, et al.* (2008). Sox18 induces development of the lymphatic vasculature in mice. Nature *456*, 643-647.

Fruman, D. A., Snapper, S. B., Yballe, C. M., Davidson, L., Yu, J. Y., Alt, F. W., and Cantley, L. C. (1999). Impaired B cell development and proliferation in absence of phosphoinositide 3-kinase p85alpha. Science *283*, 393-397.

Fung, T. K., Leung, A. Y., and So, C. W. (2013). Sox4you: a new player in C/EBPalpha leukemia. Cancer cell *24*, 557-559.

Futaki, S., Hayashi, Y., Emoto, T., Weber, C. N., and Sekiguchi, K. (2004). Sox7 plays crucial roles in parietal endoderm differentiation in F9 embryonal carcinoma cells through regulating Gata-4 and Gata-6 expression. Molecular and cellular biology *24*, 10492-10503.

Gandillet, A., Serrano, A. G., Pearson, S., Lie, A. L. M., Lacaud, G., and Kouskoff, V. (2009). Sox7-sustained expression alters the balance between proliferation and differentiation of hematopoietic progenitors at the onset of blood specification. Blood *114*, 4813-4822.

Gardner, R. L. (1979). Cell lineage and cell commitment in the early mammalian embryo. Mead Johnson Symp Perinat Dev Med, 18-24.

Gekas, C., Dieterlen-Lievre, F., Orkin, S. H., and Mikkola, H. K. (2005). The placenta is a niche for hematopoietic stem cells. Developmental cell *8*, 365-375.

Georgopoulos, K., Bigby, M., Wang, J. H., Molnar, A., Wu, P., Winandy, S., and Sharpe, A. (1994). The Ikaros gene is required for the development of all lymphoid lineages. Cell *79*, 143-156.

Georgopoulos, K., Winandy, S., and Avitahl, N. (1997). The role of the Ikaros gene in lymphocyte development and homeostasis. Annual review of immunology *15*, 155-176.

Giliani, S., Mori, L., de Saint Basile, G., Le Deist, F., Rodriguez-Perez, C., Forino, C., Mazzolari, E., Dupuis, S., Elhasid, R., Kessel, A.*, et al.* (2005). Interleukin-7 receptor alpha (IL-7Ralpha) deficiency: cellular and molecular bases. Analysis of clinical, immunological, and molecular features in 16 novel patients. Immunological reviews *203*, 110-126.

Gonzalez, D., van der Burg, M., Garcia-Sanz, R., Fenton, J. A., Langerak, A. W., Gonzalez, M., van Dongen, J. J., San Miguel, J. F., and Morgan, G. J. (2007). Immunoglobulin gene rearrangements and the pathogenesis of multiple myeloma. Blood *110*, 3112-3121.

Gordon-Keylock, S., Sobiesiak, M., Rybtsov, S., Moore, K., and Medvinsky, A. (2013). Mouse extraembryonic arterial vessels harbor precursors capable of maturing into definitive HSCs. Blood *122*, 2338-2345.

Greaves, M. (2003). Pre-natal origins of childhood leukemia. Reviews in clinical and experimental hematology *7*, 233-245.

Griffiths, E. A., Gore, S. D., Hooker, C., McDevitt, M. A., Karp, J. E., Smith, B. D., Mohammad, H. P., Ye, Y., Herman, J. G., and Carraway, H. E. (2010). Acute myeloid leukemia is characterized by Wnt pathway inhibitor promoter hypermethylation. Leukemia & lymphoma *51*, 1711-1719.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature *346*, 245-250.

Guo, L., Zhong, D., Lau, S., Liu, X., Dong, X. Y., Sun, X., Yang, V. W., Vertino, P. M., Moreno, C. S., Varma, V.*, et al.* (2008). Sox7 Is an independent checkpoint for beta-catenin function in prostate and colon epithelial cells. Molecular cancer research : MCR *6*, 1421-1430.

Guzman, M. L., and Jordan, C. T. (2004). Considerations for targeting malignant stem cells in leukemia. Cancer control : journal of the Moffitt Cancer Center *11*, 97-104.

Haddad, R., Guardiola, P., Izac, B., Thibault, C., Radich, J., Delezoide, A. L., Baillou, C., Lemoine, F. M., Gluckman, J. C., Pflumio, F., and Canque, B. (2004). Molecular characterization of early human T/NK and B-lymphoid progenitor cells in umbilical cord blood. Blood *104*, 3918-3926.

Hagman, J., and Lukin, K. (2006). Transcription factors drive B cell development. Current opinion in immunology *18*, 127-134.

Hann, I. M., Evans, D. I., Marsden, H. B., Jones, P. M., and Palmer, M. K. (1978). Bone marrow fibrosis in acute lymphoblastic leukaemia of childhood. Journal of clinical pathology *31*, 313-315.

Hardy, R. R., and Hayakawa, K. (1991). A developmental switch in B lymphopoiesis. Proceedings of the National Academy of Sciences of the United States of America *88*, 11550-11554.

Hardy, R. R., Kincade, P. W., and Dorshkind, K. (2007). The protean nature of cells in the B lymphocyte lineage. Immunity *26*, 703-714.

Hastings, J. W. (1996). Chemistries and colors of bioluminescent reactions: a review. Gene *173*, 5-11.

Hayano, T., Garg, M., Yin, D., Sudo, M., Kawamata, N., Shi, S., Chien, W., Ding, L. W., Leong, G., Mori, S.*, et al.* (2013). SOX7 is down-regulated in lung cancer. Journal of experimental & clinical cancer research : CR *32*, 17.

He, S., Kim, I., Lim, M. S., and Morrison, S. J. (2011). Sox17 expression confers self-renewal potential and fetal stem cell characteristics upon adult hematopoietic progenitors. Genes & development *25*, 1613-1627.

Heidenreich, O. (2009). Targeting oncogenes with siRNAs. Methods in molecular biology *487*, 221-242.

Hendriks, R. W., de Bruijn, M. F., Maas, A., Dingjan, G. M., Karis, A., and Grosveld, F. (1996). Inactivation of Btk by insertion of lacZ reveals defects in B cell development only past the pre-B cell stage. The EMBO journal *15*, 4862-4872.

Herpers, R., van de Kamp, E., Duckers, H. J., and Schulte-Merker, S. (2008). Redundant roles for sox7 and sox18 in arteriovenous specification in zebrafish. Circulation research *102*, 12-15.

Holtschke, T., Lohler, J., Kanno, Y., Fehr, T., Giese, N., Rosenbauer, F., Lou, J., Knobeloch, K. P., Gabriele, L., Waring, J. F.*, et al.* (1996). Immunodeficiency and chronic myelogenous leukemia-like syndrome in mice with a targeted mutation of the ICSBP gene. Cell *87*, 307-317.

Honda, H., Oda, H., Suzuki, T., Takahashi, T., Witte, O. N., Ozawa, K., Ishikawa, T., Yazaki, Y., and Hirai, H. (1998). Development of acute lymphoblastic leukemia and myeloproliferative disorder in transgenic mice expressing p210bcr/abl: a novel transgenic model for human Ph1-positive leukemias. Blood *91*, 2067-2075.

Hosking, B., Francois, M., Wilhelm, D., Orsenigo, F., Caprini, A., Svingen, T., Tutt, D., Davidson, T., Browne, C., Dejana, E., and Koopman, P. (2009). Sox7 and Sox17 are strain-specific modifiers of the lymphangiogenic defects caused by Sox18 dysfunction in mice. Development (Cambridge, England) *136*, 2385-2391.

Hosking, B. M., Muscat, G. E., Koopman, P. A., Dowhan, D. H., and Dunn, T. L. (1995). Trans-activation and DNA-binding properties of the transcription factor, Sox-18. Nucleic acids research *23*, 2626-2628.

Hosking, B. M., Wyeth, J. R., Pennisi, D. J., Wang, S. C., Koopman, P., and Muscat, G. E. (2001). Cloning and functional analysis of the Sry-related HMG box gene, Sox18. Gene *262*, 239-247.

Huber, T. L., Kouskoff, V., Fehling, H. J., Palis, J., and Keller, G. (2004). Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. Nature *432*, 625-630.

Hudson, C., Clements, D., Friday, R. V., Stott, D., and Woodland, H. R. (1997). Xsox17alpha and -beta mediate endoderm formation in Xenopus. Cell *91*, 397-405.

Hystad, M. E., Myklebust, J. H., Bo, T. H., Sivertsen, E. A., Rian, E., Forfang, L., Munthe, E., Rosenwald, A., Chiorazzi, M., Jonassen, I.*, et al.* (2007). Characterization of early stages of human B cell development by gene expression profiling. Journal of immunology *179*, 3662-3671.

Ichikawa, M., Asai, T., Saito, T., Seo, S., Yamazaki, I., Yamagata, T., Mitani, K., Chiba, S., Ogawa, S., Kurokawa, M., and Hirai, H. (2004). AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. Nature medicine *10*, 299-304.

Ikushima, H., Todo, T., Ino, Y., Takahashi, M., Miyazawa, K., and Miyazono, K. (2009). Autocrine TGF-beta signaling maintains tumorigenicity of glioma-initiating cells through Sry-related HMG-box factors. Cell stem cell *5*, 504-514.

Inman, K. E., and Downs, K. M. (2007). The murine allantois: emerging paradigms in development of the mammalian umbilical cord and its relation to the fetus. Genesis *45*, 237-258.

Irrthum, A., Devriendt, K., Chitayat, D., Matthijs, G., Glade, C., Steijlen, P. M., Fryns, J. P., Van Steensel, M. A., and Vikkula, M. (2003). Mutations in the transcription factor gene SOX18 underlie recessive and dominant forms of hypotrichosis-lymphedema-telangiectasia. American journal of human genetics *72*, 1470-1478.

Jaatinen, T., and Laine, J. (2007). Isolation of hematopoietic stem cells from human cord blood. Current protocols in stem cell biology *Chapter 2*, Unit 2A 2.

Jaffredo, T., Gautier, R., Brajeul, V., and Dieterlen-Lievre, F. (2000). Tracing the progeny of the aortic hemangioblast in the avian embryo. Developmental biology *224*, 204-214.

Jaffredo, T., Gautier, R., Eichmann, A., and Dieterlen-Lievre, F. (1998). Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny. Development (Cambridge, England) *125*, 4575-4583.

Jagannathan-Bogdan, M., and Zon, L. I. (2013). Hematopoiesis. Development (Cambridge, England) *140*, 2463-2467.

Jia, Y., Yang, Y., Liu, S., Herman, J. G., Lu, F., and Guo, M. (2010). SOX17 antagonizes WNT/beta-catenin signaling pathway in hepatocellular carcinoma. Epigenetics : official journal of the DNA Methylation Society *5*, 743-749.

Jiang, Q., Li, W. Q., Aiello, F. B., Mazzucchelli, R., Asefa, B., Khaled, A. R., and Durum, S. K. (2005). Cell biology of IL-7, a key lymphotrophin. Cytokine & growth factor reviews *16*, 513-533.

Johnstone, R. W. (2002). Histone-deacetylase inhibitors: novel drugs for the treatment of cancer. Nature reviews Drug discovery *1*, 287-299.

Kamachi, Y., Uchikawa, M., and Kondoh, H. (2000). Pairing SOX off: with partners in the regulation of embryonic development. Trends in genetics : TIG *16*, 182-187.

Kamps, M. P., and Baltimore, D. (1993). E2A-Pbx1, the t(1;19) translocation protein of human pre-B-cell acute lymphocytic leukemia, causes acute myeloid leukemia in mice. Molecular and cellular biology *13*, 351-357.

Kanai-Azuma, M., Kanai, Y., Gad, J. M., Tajima, Y., Taya, C., Kurohmaru, M., Sanai, Y., Yonekawa, H., Yazaki, K., Tam, P. P., and Hayashi, Y. (2002). Depletion of definitive gut endoderm in Sox17-null mutant mice. Development (Cambridge, England) *129*, 2367-2379.

Katoh, M. (2002). Expression of human SOX7 in normal tissues and tumors. International journal of molecular medicine *9*, 363-368.

Kee, B. L. (2009). E and ID proteins branch out. Nature reviews Immunology *9*, 175-184.

Keller, G., Kennedy, M., Papayannopoulou, T., and Wiles, M. V. (1993). Hematopoietic commitment during embryonic stem cell differentiation in culture. Molecular and cellular biology *13*, 473-486.

Keller, G. M. (1995). In vitro differentiation of embryonic stem cells. Current opinion in cell biology *7*, 862-869.

Khan, W. N., Wright, J. A., Kleiman, E., Boucher, J. C., Castro, I., and Clark, E. S. (2013). B-lymphocyte tolerance and effector function in immunity and autoimmunity. Immunologic research *57*, 335-353.

Kim, I., Saunders, T. L., and Morrison, S. J. (2007). Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells. Cell *130*, 470-483.

Kitamura, D., Kudo, A., Schaal, S., Muller, W., Melchers, F., and Rajewsky, K. (1992). A critical role of lambda 5 protein in B cell development. Cell *69*, 823-831.

Kitamura, D., Roes, J., Kuhn, R., and Rajewsky, K. (1991). A B cell-deficient mouse by targeted disruption of the membrane exon of the immunoglobulin mu chain gene. Nature *350*, 423-426.

Koethe, S., Zander, L., Koster, S., Annan, A., Ebenfelt, A., Spencer, J., and Bemark, M. (2011). Pivotal advance: CD45RB glycosylation is specifically regulated during human peripheral B cell differentiation. Journal of leukocyte biology *90*, 5-19.

Kondoh, H., and Kamachi, Y. (2010). SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. The international journal of biochemistry & cell biology *42*, 391-399.

Kong, Y., Yoshida, S., Saito, Y., Doi, T., Nagatoshi, Y., Fukata, M., Saito, N., Yang, S. M., Iwamoto, C., Okamura, J., *et al* (2008). CD34+CD38+CD19+ as well as CD34+CD38-CD19+ cells are leukaemia-initiating cells with self-renewal capacity in human B-precursor ALL. Leukemia *22*, 1207-1213.

Kormish, J. D., Sinner, D., and Zorn, A. M. (2010). Interactions between SOX factors and Wnt/beta-catenin signaling in development and disease. Developmental dynamics : an official publication of the American Association of Anatomists *239*, 56-68.

Kraus, M., Pao, L. I., Reichlin, A., Hu, Y., Canono, B., Cambier, J. C., Nussenzweig, M. C., and Rajewsky, K. (2001). Interference with immunoglobulin (Ig)alpha immunoreceptor tyrosine-based activation motif (ITAM) phosphorylation modulates or blocks B cell development, depending on the availability of an Igbeta cytoplasmic tail. The Journal of experimental medicine *194*, 455-469.

Krejci, O., Wunderlich, M., Geiger, H., Chou, F. S., Schleimer, D., Jansen, M., Andreassen, P. R., and Mulloy, J. C. (2008). p53 signaling in response to increased

DNA damage sensitizes AML1-ETO cells to stress-induced death. Blood *111*, 2190-2199.

Kuter, D. J., Bain, B., Mufti, G., Bagg, A., and Hasserjian, R. P. (2007). Bone marrow fibrosis: pathophysiology and clinical significance of increased bone marrow stromal fibres. British journal of haematology *139*, 351-362.

Lacaud, G., Gore, L., Kennedy, M., Kouskoff, V., Kingsley, P., Hogan, C., Carlsson, L., Speck, N., Palis, J., and Keller, G. (2002). Runx1 is essential for hematopoietic commitment at the hemangioblast stage of development in vitro. Blood *100*, 458-466.

Lacaud, G., Keller, G., and Kouskoff, V. (2004a). Tracking mesoderm formation and specification to the hemangioblast in vitro. Trends in cardiovascular medicine *14*, 314-317.

Lacaud, G., Kouskoff, V., Trumble, A., Schwantz, S., and Keller, G. (2004b). Haploinsufficiency of Runx1 results in the acceleration of mesodermal development and hemangioblast specification upon in vitro differentiation of ES cells. Blood *103*, 886-889.

Lacaud, G., Robertson, S., Palis, J., Kennedy, M., and Keller, G. (2001). Regulation of hemangioblast development. Annals of the New York Academy of Sciences *938*, 96-107; discussion 108.

Lam, E. Y., Hall, C. J., Crosier, P. S., Crosier, K. E., and Flores, M. V. (2010). Live imaging of Runx1 expression in the dorsal aorta tracks the emergence of blood progenitors from endothelial cells. Blood *116*, 909-914.

Lancrin, C., Sroczynska, P., Serrano, A. G., Gandillet, A., Ferreras, C., Kouskoff, V., and Lacaud, G. (2010). Blood cell generation from the hemangioblast. Journal of molecular medicine *88*, 167-172.

Lancrin, C., Sroczynska, P., Stephenson, C., Allen, T., Kouskoff, V., and Lacaud, G. (2009). The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. Nature *457*, 892-895.

Lange, A. W., Keiser, A. R., Wells, J. M., Zorn, A. M., and Whitsett, J. A. (2009). Sox17 promotes cell cycle progression and inhibits TGF-beta/Smad3 signaling to initiate progenitor cell behavior in the respiratory epithelium. PloS one *4*, e5711.

Laudet, V., Stehelin, D., and Clevers, H. (1993). Ancestry and diversity of the HMG box superfamily. Nucleic acids research *21*, 2493-2501.

Lawson, K. A., Meneses, J. J., and Pedersen, R. A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. Development (Cambridge, England) *113*, 891-911.

LeBien, T. W. (2000). Fates of human B-cell precursors. Blood *96*, 9-23.

LeBien, T. W., and Tedder, T. F. (2008). B lymphocytes: how they develop and function. Blood *112*, 1570-1580.

Lee, D., Park, C., Lee, H., Lugus, J. J., Kim, S. H., Arentson, E., Chung, Y. S., Gomez, G., Kyba, M., Lin, S.*, et al.* (2008). ER71 acts downstream of BMP, Notch, and Wnt signaling in blood and vessel progenitor specification. Cell stem cell *2*, 497-507.

Lee, E., Choi, H. S., Hwang, J. H., Hoh, J. K., Cho, Y. H., and Baek, E. J. (2014). The RNA in reticulocytes is not just debris: it is necessary for the final stages of erythrocyte formation. Blood cells, molecules & diseases *53*, 1-10.

Le Viseur, C., Hotfilder, M., Bomken, S., Wilson, K., Röttgers, S., Schrauder, A., Rosemann, A., Irving, J., Stam, R. W., Shultz, L. D., Harbott, J., Jürgens, H., Schrappe, M., Pieters, R., and Vormoor, J. (2008).In childhood acute lymphoblastic leukemia, blasts at different stages of immunophenotypic maturation have stem cell properties. Cancer cell *14*, 47-58.

Li, B., Ge, Z., Song, S., Zhang, S., Yan, H., Huang, B., and Zhang, Y. (2012a). Decreased expression of SOX7 is correlated with poor prognosis in lung adenocarcinoma patients. Pathology oncology research : POR *18*, 1039-1045.

Li, M., Jones, L., Gaillard, C., Binnewies, M., Ochoa, R., Garcia, E., Lam, V., Wei, G., Yang, W., Lobe, C.*, et al.* (2013). Initially disadvantaged, TEL-AML1 cells expand and initiate leukemia in response to irradiation and cooperating mutations. Leukemia *27*, 1570-1573.

Li, Y. S., Hayakawa, K., and Hardy, R. R. (1993). The regulated expression of B lineage associated genes during B cell differentiation in bone marrow and fetal liver. The Journal of experimental medicine *178*, 951-960.

Li, Z., Lan, Y., He, W., Chen, D., Wang, J., Zhou, F., Wang, Y., Sun, H., Chen, X., Xu, C.*, et al.* (2012b). Mouse embryonic head as a site for hematopoietic stem cell development. Cell stem cell *11*, 663-675.

Liang, Y., Jansen, M., Aronow, B., Geiger, H., and Van Zant, G. (2007). The quantitative trait gene latexin influences the size of the hematopoietic stem cell population in mice. Nature genetics *39*, 178-188.

Lin, H., and Grosschedl, R. (1995). Failure of B-cell differentiation in mice lacking the transcription factor EBF. Nature *376*, 263-267.

Loder, F., Mutschler, B., Ray, R. J., Paige, C. J., Sideras, P., Torres, R., Lamers, M. C., and Carsetti, R. (1999). B cell development in the spleen takes place in discrete steps and is determined by the quality of B cell receptor-derived signals. The Journal of experimental medicine *190*, 75-89.

Long, F., Wang, X., Fang, S., Xu, Y., Sun, K., Chen, S., and Xu, R. (2013). A potential relationship among beta-defensins haplotype, SOX7 duplication and cardiac defects. PloS one *8*, e72515.

Lu, R., Medina, K. L., Lancki, D. W., and Singh, H. (2003). IRF-4,8 orchestrate the pre-B-to-B transition in lymphocyte development. Genes & development *17*, 1703-1708.

Lundberg, I. V., Lofgren Burstrom, A., Edin, S., Eklof, V., Oberg, A., Stenling, R., Palmqvist, R., and Wikberg, M. L. (2014). SOX2 expression is regulated by BRAF and contributes to poor patient prognosis in colorectal cancer. PloS one *9*, e101957.

Ma, H., Mallampati, S., Lu, Y., Sun, B., Wang, E., Leng, X., Gong, Y., Shen, H., Yin, C. C., Jones, D*., et al.* (2014). Sox4/Tcf7l1 axis promotes progression of BCR- ABL+ acute lymphoblastic leukemia. Haematologica.

Maier, H., Ostraat, R., Gao, H., Fields, S., Shinton, S. A., Medina, K. L., Ikawa, T., Murre, C., Singh, H., Hardy, R. R., and Hagman, J. (2004). Early B cell factor cooperates with Runx1 and mediates epigenetic changes associated with mb-1 transcription. Nature immunology *5*, 1069-1077.

Malki, S., Boizet-Bonhoure, B., and Poulat, F. (2010). Shuttling of SOX proteins. The international journal of biochemistry & cell biology *42*, 411-416.

Manoharan, A., Horsley, R., and Pitney, W. R. (1979). The reticulin content of bone marrow in acute leukaemia in adults. British journal of haematology *43*, 185-190.

Manz, M.G., Traver, D., Miyamoto, T., Weissman, I. L., and Akashi K. (2001). Dendritic cell potentials of early lymphoid and myeloid progenitors. Blood *97*, 3333-41.

McGrath, K. E., and Palis, J. (2005). Hematopoiesis in the yolk sac: more than meets the eye. Experimental hematology *33*, 1021-1028.

McKercher, S. R., Torbett, B. E., Anderson, K. L., Henkel, G. W., Vestal, D. J., Baribault, H., Klemsz, M., Feeney, A. J., Wu, G. E., Paige, C. J., and Maki, R. A. (1996). Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. The EMBO journal *15*, 5647-5658.

Mebius, R. E., and Kraal, G. (2005). Structure and function of the spleen. Nature reviews Immunology *5*, 606-616.

Medina, K. L., Pongubala, J. M., Reddy, K. L., Lancki, D. W., Dekoter, R., Kieslinger, M., Grosschedl, R., and Singh, H. (2004). Assembling a gene regulatory network for specification of the B cell fate. Developmental cell *7*, 607-617.

Medina, P. P., Castillo, S. D., Blanco, S., Sanz-Garcia, M., Largo, C., Alvarez, S., Yokota, J., Gonzalez-Neira, A., Benitez, J., Clevers, H. C*., et al.* (2009). The SRY-HMG box gene, SOX4, is a target of gene amplification at chromosome 6p in lung cancer. Human molecular genetics *18*, 1343-1352.

Medvinsky, A., and Dzierzak, E. (1996). Definitive hematopoiesis is autonomously initiated by the AGM region. Cell *86*, 897-906.

Medvinsky, A., Rybtsov, S., and Taoudi, S. (2011). Embryonic origin of the adult hematopoietic system: advances and questions. Development *138*, 1017-1031.

Meggendorfer, M., Kern, W., Haferlach, C., Haferlach, T., and Schnittger, S. (2013). SOX11 overexpression is a specific marker for mantle cell lymphoma and correlates with t(11;14) translocation, CCND1 expression and an adverse prognosis. Leukemia *27*, 2388-2391.

Melamed, D., Kench, J. A., Grabstein, K., Rolink, A., and Nemazee, D. (1997). A functional B cell receptor transgene allows efficient IL-7-independent maturation of B cell precursors. Journal of immunology *159*, 1233-1239.

Mikkola, I., Heavey, B., Horcher, M., and Busslinger, M. (2002). Reversion of B cell commitment upon loss of Pax5 expression. Science *297*, 110-113.

Miller, R. W., Young, J. L., Jr., and Novakovic, B. (1995). Childhood cancer. Cancer *75*, 395-405.

Miosge, L. A., and Goodnow, C. C. (2005). Genes, pathways and checkpoints in lymphocyte development and homeostasis. Immunology and cell biology *83*, 318-335.

Mittrucker, H. W., Matsuyama, T., Grossman, A., Kundig, T. M., Potter, J., Shahinian, A., Wakeham, A., Patterson, B., Ohashi, P. S., and Mak, T. W. (1997). Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function. Science *275*, 540-543.

Mombaerts, P., Iacomini, J., Johnson, R. S., Herrup, K., Tonegawa, S., and Papaioannou, V. E. (1992). RAG-1-deficient mice have no mature B and T lymphocytes. Cell *68*, 869-877.

Montes, R., Ayllon, V., Gutierrez-Aranda, I., Prat, I., Hernandez-Lamas, M. C., Ponce, L., Bresolin, S., Te Kronnie, G., Greaves, M., Bueno, C., and Menendez, P. (2011). Enforced expression of MLL-AF4 fusion in cord blood CD34+ cells enhances the hematopoietic repopulating cell function and clonogenic potential but is not sufficient to initiate leukemia. Blood *117*, 4746-4758.

Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E.*, et al.* (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature genetics *34*, 267-273.

Mori, S., Rempel, R. E., Chang, J. T., Yao, G., Lagoo, A. S., Potti, A., Bild, A., and Nevins, J. R. (2008). Utilization of pathway signatures to reveal distinct types of B lymphoma in the Emicro-myc model and human diffuse large B-cell lymphoma. Cancer research *68*, 8525-8534.

Mundt, C., Licence, S., Shimizu, T., Melchers, F., and Martensson, I. L. (2001). Loss of precursor B cell expansion but not allelic exclusion in VpreB1/VpreB2 double-deficient mice. The Journal of experimental medicine *193*, 435-445.

Murakami, A., Shen, H., Ishida, S., and Dickson, C. (2004). SOX7 and GATA-4 are competitive activators of Fgf-3 transcription. The Journal of biological chemistry *279*, 28564-28573.

Murphy, A. J., Bijl, N., Yvan-Charvet, L., Welch, C. B., Bhagwat, N., Reheman, A., Wang, Y., Shaw, J. A., Levine, R. L., Ni, H.*, et al.* (2013). Cholesterol efflux in megakaryocyte progenitors suppresses platelet production and thrombocytosis. Nature medicine *19*, 586-594.

Nagasawa, T. (2006). Microenvironmental niches in the bone marrow required for B-cell development. Nature reviews Immunology *6*, 107-116.

Nakajima-Takagi, Y., Osawa, M., Oshima, M., Takagi, H., Miyagi, S., Endoh, M., Endo, T. A., Takayama, N., Eto, K., Toyoda, T.*, et al.* (2013). Role of SOX17 in hematopoietic development from human embryonic stem cells. Blood *121*, 447-458.

Nakano, H., Liu, X., Arshi, A., Nakashima, Y., van Handel, B., Sasidharan, R., Harmon, A. W., Shin, J. H., Schwartz, R. J., Conway, S. J.*, et al.* (2013). Haemogenic endocardium contributes to transient definitive haematopoiesis. Nature communications *4*, 1564.

Nelson, T. J., Chiriac, A., Faustino, R. S., Crespo-Diaz, R. J., Behfar, A., and Terzic, A. (2009). Lineage specification of Flk-1+ progenitors is associated with divergent Sox7 expression in cardiopoiesis. Differentiation; research in biological diversity *77*, 248-255.

Neuberger, M. S., and Milstein, C. (1995). Somatic hypermutation. Current opinion in immunology *7*, 248-254.

Niebuhr, B., Kriebitzsch, N., Fischer, M., Behrens, K., Gunther, T., Alawi, M., Bergholz, U., Muller, U., Roscher, S., Ziegler, M.*, et al.* (2013). Runx1 is essential at two stages of early murine B-cell development. Blood *122*, 413-423.

Nielsen, P. J., Georgiev, O., Lorenz, B., and Schaffner, W. (1996). B lymphocytes are impaired in mice lacking the transcriptional co-activator Bob1/OCA-B/OBF1. European journal of immunology *26*, 3214-3218.

Niimi, T., Hayashi, Y., Futaki, S., and Sekiguchi, K. (2004). SOX7 and SOX17 regulate the parietal endoderm-specific enhancer activity of mouse laminin alpha1 gene. The Journal of biological chemistry *279*, 38055-38061.

Nilson, I., Reichel, M., Ennas, M. G., Greim, R., Knorr, C., Siegler, G., Greil, J., Fey, G. H., and Marschalek, R. (1997). Exon/intron structure of the human AF-4 gene, a member of the AF-4/LAF-4/FMR-2 gene family coding for a nuclear protein with structural alterations in acute leukaemia. British journal of haematology *98*, 157-169.

Nishikawa, S. (2012). Hemangioblast: an in vitro phantom. Wiley interdisciplinary reviews Developmental biology *1*, 603-608.

Nishikawa, S. I., Nishikawa, S., Hirashima, M., Matsuyoshi, N., and Kodama, H. (1998). Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages. Development (Cambridge, England) *125*, 1747-1757.

Nossal, G. J. (1994). Negative selection of lymphocytes. Cell *76*, 229-239.

Nutt, S. L., Heavey, B., Rolink, A. G., and Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature *401*, 556-562.

O'Leary, H., Akers, S. M., Piktel, D., Walton, C., Fortney, J. E., Martin, K. H., Craig, M., Coad, J., and Gibson, L. F. (2010). VE-cadherin Regulates Philadelphia Chromosome Positive Acute Lymphoblastic Leukemia Sensitivity to Apoptosis. Cancer microenvironment : official journal of the International Cancer Microenvironment Society *3*, 67-81.

Oberlin, E., Tavian, M., Blazsek, I., and Peault, B. (2002). Blood-forming potential of vascular endothelium in the human embryo. Development *129*, 4147-4157.

Obino, D., and Lennon-Dumenil, A. M. (2014). A critical role for cell polarity in antigen extraction, processing, and presentation by B lymphocytes. Advances in immunology *123*, 51-67.

Okuda, T., van Deursen, J., Hiebert, S. W., Grosveld, G., and Downing, J. R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell *84*, 321-330.

Omidvar, N., Maunakea, M. L., Jones, L., Sevcikova, S., Yin, B., Himmel, K. L., Tennant, T. R., Le Beau, M. M., Largaespada, D. A., and Kogan, S. C. (2013). PML-RARalpha co-operates with Sox4 in acute myeloid leukemia development in mice. Haematologica *98*, 424-427.

Orkin, S. H. (2000). Diversification of haematopoietic stem cells to specific lineages. Nature reviews Genetics *1*, 57-64.

Orkin, S. H., and Zon, L. I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. Cell *132*, 631-644.

Ottersbach, K., and Dzierzak, E. (2005). The murine placenta contains hematopoietic stem cells within the vascular labyrinth region. Developmental cell *8*, 377-387.

Ottersbach, K., and Dzierzak, E. (2009). Analysis of the mouse placenta as a hematopoietic stem cell niche. Methods in molecular biology *538*, 335-346.

Palis, J., McGrath, K. E., and Kingsley, P. D. (1995). Initiation of hematopoiesis and vasculogenesis in murine yolk sac explants. Blood *86*, 156-163.

Palis, J., Robertson, S., Kennedy, M., Wall, C., and Keller, G. (1999). Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. Development (Cambridge, England) *126*, 5073-5084.

Pang, L., Weiss, M. J., and Poncz, M. (2005). Megakaryocyte biology and related disorders. The Journal of clinical investigation *115*, 3332-3338.

Papathanasiou, P., Perkins, A. C., Cobb, B. S., Ferrini, R., Sridharan, R., Hoyne, G. F., Nelms, K. A., Smale, S. T., and Goodnow, C. C. (2003). Widespread failure of hematolymphoid differentiation caused by a recessive niche-filling allele of the Ikaros transcription factor. Immunity *19*, 131-144.

Pappu, R., Cheng, A. M., Li, B., Gong, Q., Chiu, C., Griffin, N., White, M., Sleckman, B. P., and Chan, A. C. (1999). Requirement for B cell linker protein (BLNK) in B cell development. Science *286*, 1949-1954.

Paukku, K., and Silvennoinen, O. (2004). STATs as critical mediators of signal transduction and transcription: lessons learned from STAT5. Cytokine & growth factor reviews *15*, 435-455.

Peled, J. U., Kuang, F. L., Iglesias-Ussel, M. D., Roa, S., Kalis, S. L., Goodman, M. F., and Scharff, M. D. (2008). The biochemistry of somatic hypermutation. Annual review of immunology *26*, 481-511.

Pencovich, N., Jaschek, R., Dicken, J., Amit, A., Lotem, J., Tanay, A., and Groner, Y. (2013). Cell-autonomous function of Runx1 transcriptionally regulates mouse megakaryocytic maturation. PloS one *8*, e64248.

Pendergast, A. M., Muller, A. J., Havlik, M. H., Maru, Y., and Witte, O. N. (1991). BCR sequences essential for transformation by the BCR-ABL oncogene bind to the ABL SH2 regulatory domain in a non-phosphotyrosine-dependent manner. Cell *66*, 161-171.

Pendeville, H., Winandy, M., Manfroid, I., Nivelles, O., Motte, P., Pasque, V., Peers, B., Struman, I., Martial, J. A., and Voz, M. L. (2008). Zebrafish Sox7 and Sox18 function together to control arterial-venous identity. Developmental biology *317*, 405-416.

Pennisi, D., Bowles, J., Nagy, A., Muscat, G., and Koopman, P. (2000). Mice null for sox18 are viable and display a mild coat defect. Molecular and cellular biology *20*, 9331-9336.

Penzo-Mendez, A. I. (2010). Critical roles for SoxC transcription factors in development and cancer. The international journal of biochemistry & cell biology *42*, 425-428.

Perez-Andres, M., Paiva, B., Nieto, W. G., Caraux, A., Schmitz, A., Almeida, J., Vogt, R. F., Jr., Marti, G. E., Rawstron, A. C., Van Zelm, M. C.*, et al.* (2010). Human peripheral blood B-cell compartments: a crossroad in B-cell traffic. Cytometry Part B, Clinical cytometry *78 Suppl 1*, S47-60.

Peschon, J. J., Morrissey, P. J., Grabstein, K. H., Ramsdell, F. J., Maraskovsky, E., Gliniak, B. C., Park, L. S., Ziegler, S. F., Williams, D. E., Ware, C. B.*, et al.* (1994). Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice. The Journal of experimental medicine *180*, 1955-1960.

Pillai, S., and Cariappa, A. (2009). The follicular versus marginal zone B lymphocyte cell fate decision. Nature reviews Immunology *9*, 767-777.

Puel, A., Ziegler, S. F., Buckley, R. H., and Leonard, W. J. (1998). Defective IL7R expression in T(-)B(+)NK(+) severe combined immunodeficiency. Nature genetics *20*, 394-397.

Pui, C. H., and Evans, W. E. (2006). Treatment of acute lymphoblastic leukemia. The New England journal of medicine *354*, 166-178.

Pui, C. H., Mullighan, C. G., Evans, W. E., and Relling, M. V. (2012). Pediatric acute lymphoblastic leukemia: where are we going and how do we get there? Blood *120*, 1165-1174.

Pui, C. H., Relling, M. V., and Downing, J. R. (2004). Acute lymphoblastic leukemia. The New England journal of medicine *350*, 1535-1548.

Pui, C. H., Robison, L. L., and Look, A. T. (2008). Acute lymphoblastic leukaemia. Lancet *371*, 1030-1043.

Pula, B., Kobierzycki, C., Solinski, D., Olbromski, M., Nowak-Markwitz, E., Spaczynski, M., Kedzia, W., Zabel, M., and Dziegiel, P. (2014). SOX18 Expression Predicts Response to Platinum-based Chemotherapy in Ovarian Cancer. Anticancer research *34*, 4029-4037.

Pula, B., Olbromski, M., Wojnar, A., Gomulkiewicz, A., Witkiewicz, W., Ugorski, M., Dziegiel, P., and Podhorska-Okolow, M. (2013). Impact of SOX18 expression in cancer cells and vessels on the outcome of invasive ductal breast carcinoma. Cellular oncology *36*, 469-483.

Quintas-Cardama, A., and Verstovsek, S. (2013). Molecular pathways: Jak/STAT pathway: mutations, inhibitors, and resistance. Clinical cancer research : an official journal of the American Association for Cancer Research *19*, 1933-1940.

Ramadani, F., Bolland, D. J., Garcon, F., Emery, J. L., Vanhaesebroeck, B., Corcoran, A. E., and Okkenhaug, K. (2010). The PI3K isoforms p110alpha and p110delta are essential for pre-B cell receptor signaling and B cell development. Science signaling *3*, ra60.

Ramezani-Rad, P., Geng, H., Hurtz, C., Chan, L. N., Chen, Z., Jumaa, H., Melnick, A., Paietta, E., Carroll, W. L., Willman, C. L*., et al.* (2013). SOX4 enables oncogenic survival signals in acute lymphoblastic leukemia. Blood *121*, 148-155.

Reichlin, A., Hu, Y., Meffre, E., Nagaoka, H., Gong, S., Kraus, M., Rajewsky, K., and Nussenzweig, M. C. (2001). B cell development is arrested at the immature B cell stage in mice carrying a mutation in the cytoplasmic domain of immunoglobulin beta. The Journal of experimental medicine *193*, 13-23.

Reya, T., O'Riordan, M., Okamura, R., Devaney, E., Willert, K., Nusse, R., and Grosschedl, R. (2000). Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. Immunity *13*, 15-24.

Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A., and Chinnaiyan, A. M. (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia *6*, 1-6.

Robertson, K. D. (2002). DNA methylation and chromatin - unraveling the tangled web. Oncogene *21*, 5361-5379.

Rolink, A. G., Nutt, S. L., Melchers, F., and Busslinger, M. (1999). Long-term in vivo reconstitution of T-cell development by Pax5-deficient B-cell progenitors. Nature *401*, 603-606.

Ryan, M. D., and Drew, J. (1994). Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. The EMBO journal *13*, 928-933.

Rybtsov, S., Sobiesiak, M., Taoudi, S., Souilhol, C., Senserrich, J., Liakhovitskaia, A., Ivanovs, A., Frampton, J., Zhao, S., and Medvinsky, A. (2011). Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. The Journal of experimental medicine *208*, 1305-1315.

Sacilotto, N., Monteiro, R., Fritzsche, M., Becker, P. W., Sanchez-Del-Campo, L., Liu, K., Pinheiro, P., Ratnayaka, I., Davies, B., Goding, C. R.*, et al.* (2013). Analysis of Dll4 regulation reveals a combinatorial role for Sox and Notch in arterial development. Proceedings of the National Academy of Sciences of the United States of America *110*, 11893-11898.

Saegusa, M., Hashimura, M., and Kuwata, T. (2012). Sox4 functions as a positive regulator of beta-catenin signaling through upregulation of TCF4 during morular differentiation of endometrial carcinomas. Laboratory investigation; a journal of technical methods and pathology *92*, 511-521.

Sakamoto, Y., Hara, K., Kanai-Azuma, M., Matsui, T., Miura, Y., Tsunekawa, N., Kurohmaru, M., Saijoh, Y., Koopman, P., and Kanai, Y. (2007). Redundant roles of Sox17 and Sox18 in early cardiovascular development of mouse embryos. Biochemical and biophysical research communications *360*, 539-544.

Schepers, G. E., Teasdale, R. D., and Koopman, P. (2002). Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. Developmental cell *3*, 167-170.

Scherr, M., Battmer, K., Winkler, T., Heidenreich, O., Ganser, A., and Eder, M. (2003). Specific inhibition of bcr-abl gene expression by small interfering RNA. Blood *101*, 1566-1569.

Schilham, M. W., Oosterwegel, M. A., Moerer, P., Ya, J., de Boer, P. A., van de Wetering, M., Verbeek, S., Lamers, W. H., Kruisbeek, A. M., Cumano, A., and Clevers,

H. (1996). Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. Nature *380*, 711-714.

Schlierf, B., Friedrich, R. P., Roerig, P., Felsberg, J., Reifenberger, G., and Wegner, M. (2007). Expression of SoxE and SoxD genes in human gliomas. Neuropathology and applied neurobiology *33*, 621-630.

Schorl, C., and Sedivy, J. M. (2007). Analysis of cell cycle phases and progression in cultured mammalian cells. Methods *41*, 143-150.

Schweighoffer, E., Vanes, L., Mathiot, A., Nakamura, T., and Tybulewicz, V. L. (2003). Unexpected requirement for ZAP-70 in pre-B cell development and allelic exclusion. Immunity *18*, 523-533.

Scott, E. W., Simon, M. C., Anastasi, J., and Singh, H. (1994). Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science *265*, 1573-1577.

Seguin, C. A., Draper, J. S., Nagy, A., and Rossant, J. (2008). Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. Cell stem cell *3*, 182-195.

Seita, J., and Weissman, I. L. (2010). Hematopoietic stem cell: self-renewal versus differentiation. Wiley interdisciplinary reviews Systems biology and medicine *2*, 640-653.

Seo, W., Ikawa, T., Kawamoto, H., and Taniuchi, I. (2012). Runx1-Cbfbeta facilitates early B lymphocyte development by regulating expression of Ebf1. The Journal of experimental medicine *209*, 1255-1262.

Serrano, A. G., Gandillet, A., Pearson, S., Lacaud, G., and Kouskoff, V. (2010). Contrasting effects of Sox17- and Sox18-sustained expression at the onset of blood specification. Blood *115*, 3895-3898.

Shalaby, F., Rossant, J., Yamaguchi, T. P., Gertsenstein, M., Wu, X. F., Breitman, M. L., and Schuh, A. C. (1995). Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. Nature *376*, 62-66.

Shinkai, Y., Rathbun, G., Lam, K. P., Oltz, E. M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, F., Stall, A. M., and et al. (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. Cell *68*, 855-867.

Shirai, T., Miyagi, S., Horiuchi, D., Okuda-Katayanagi, T., Nishimoto, M., Muramatsu, M., Sakamoto, Y., Nagata, M., Hagiwara, K., and Okuda, A. (2005). Identification of an enhancer that controls up-regulation of fibronectin during differentiation of embryonic stem cells into extraembryonic endoderm. The Journal of biological chemistry *280*, 7244-7252.

Siegle, J. M., Basin, A., Sastre-Perona, A., Yonekubo, Y., Brown, J., Sennett, R., Rendl, M., Tsirigos, A., Carucci, J. A., and Schober, M. (2014). SOX2 is a cancer-specific regulator of tumour initiating potential in cutaneous squamous cell carcinoma. Nature communications *5*, 4511.

Siino, J. S., Nissen, M. S., and Reeves, R. (1995). Replacement of conserved threonines by alanine residues in high mobility group protein HMG-I(Y): effect on DNA binding affinity. Biochemical and biophysical research communications *207*, 497-507.

Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., Foster, J. W., Frischauf, A. M., Lovell-Badge, R., and Goodfellow, P. N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature *346*, 240-244.

Sinner, D., Rankin, S., Lee, M., and Zorn, A. M. (2004). Sox17 and beta-catenin cooperate to regulate the transcription of endodermal genes. Development (Cambridge, England) *131*, 3069-3080.

Sonnemann, J., Gruhn, B., Wittig, S., Becker, S., and Beck, J. F. (2012). Increased activity of histone deacetylases in childhood acute lymphoblastic leukaemia and acute myeloid leukaemia: support for histone deacetylase inhibitors as antileukaemic agents. British journal of haematology *158*, 664-666.

Soullier, S., Jay, P., Poulat, F., Vanacker, J. M., Berta, P., and Laudet, V. (1999). Diversification pattern of the HMG and SOX family members during evolution. Journal of molecular evolution *48*, 517-527.

Spangrude, G. J. (1994). Biological and clinical aspects of hematopoietic stem cells. Annual review of medicine *45*, 93-104.

Speck, N. A. (2001). Core binding factor and its role in normal hematopoietic development. Current opinion in hematology *8*, 192-196.

Sprent, J., and Tough, D. F. (1994). Lymphocyte life-span and memory. Science *265*, 1395-1400.

Srivastava, B., Quinn, W. J., 3rd, Hazard, K., Erikson, J., and Allman, D. (2005). Characterization of marginal zone B cell precursors. The Journal of experimental medicine *202*, 1225-1234.

Stavnezer, J. (1996). Immunoglobulin class switching. Current opinion in immunology *8*, 199-205.

Stavnezer, J., Guikema, J. E., and Schrader, C. E. (2008). Mechanism and regulation of class switch recombination. Annual review of immunology *26*, 261-292.

Stovall, D. B., Cao, P., and Sui, G. (2014). SOX7: from a developmental regulator to an emerging tumor suppressor. Histology and histopathology *29*, 439-445.

Stovall, D. B., Wan, M., Miller, L. D., Cao, P., Maglic, D., Zhang, Q., Stampfer, M. R., Liu, W., Xu, J., and Sui, G. (2013). The regulation of SOX7 and its tumor suppressive role in breast cancer. The American journal of pathology *183*, 1645-1653.

Stros, M., Launholt, D., and Grasser, K. D. (2007). The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. Cellular and molecular life sciences : CMLS *64*, 2590-2606.

Studniak, E., Maloney, E., Ociepa, T., Urasinski, T., Skonieczka, K., Haus, O., Poluha, A., Kowalczyk, J., and Zajaczek, S. (2013). Allelic loss of selected tumor suppressor genes in acute lymphoblastic leukemia in children. Polish journal of pathology : official journal of the Polish Society of Pathologists *64*, 121-128.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545-15550.

Sudbeck, P., and Scherer, G. (1997). Two independent nuclear localization signals are present in the DNA-binding high-mobility group domains of SRY and SOX9. The Journal of biological chemistry *272*, 27848-27852.

Sun, J., Matthias, G., Mihatsch, M. J., Georgopoulos, K., and Matthias, P. (2003). Lack of the transcriptional coactivator OBF-1 prevents the development of systemic lupus erythematosus-like phenotypes in Aiolos mutant mice. Journal of immunology *170*, 1699-1706.

Suttie, A. W. (2006). Histopathology of the spleen. Toxicologic pathology *34*, 466-503.

Takash, W., Canizares, J., Bonneaud, N., Poulat, F., Mattei, M. G., Jay, P., and Berta, P. (2001). SOX7 transcription factor: sequence, chromosomal localisation, expression, transactivation and interference with Wnt signalling. Nucleic acids research *29*, 4274-4283.

Takata, M., and Kurosaki, T. (1996). A role for Bruton's tyrosine kinase in B cell antigen receptor-mediated activation of phospholipase C-gamma 2. The Journal of experimental medicine *184*, 31-40.

Tam, P. P., and Behringer, R. R. (1997). Mouse gastrulation: the formation of a mammalian body plan. Mech Dev *68*, 3-25.

Tang, C. Y., Lin, J., Qian, W., Yang, J., Ma, J. C., Deng, Z. Q., Yang, L., An, C., Wen, X. M., Zhang, Y. Y., and Qian, J. (2014). Low SOX17 expression: prognostic significance in de novo acute myeloid leukemia with normal cytogenetics. Clinical chemistry and laboratory medicine : CCLM / FESCC.

Taniguchi, K., Hiraoka, Y., Ogawa, M., Sakai, Y., Kido, S., and Aiso, S. (1999). Isolation and characterization of a mouse SRY-related cDNA, mSox7. Biochimica et biophysica acta *1445*, 225-231.

Taoudi, S., and Medvinsky, A. (2007). Functional identification of the hematopoietic stem cell niche in the ventral domain of the embryonic dorsal aorta. Proceedings of the National Academy of Sciences of the United States of America *104*, 9399-9403.

Thiele, J., and Kvasnicka, H. M. (2005). Hematopathologic findings in chronic idiopathic myelofibrosis. Seminars in oncology *32*, 380-394.

Thomas, L., Franck, S., Messinger, M., Linssen, J., Thome, M., and Thomas, C. (2005a). Reticulocyte hemoglobin measurement--comparison of two methods in the diagnosis of iron-restricted erythropoiesis. Clinical chemistry and laboratory medicine : CCLM / FESCC *43*, 1193-1202.

Thomas, M., Gessner, A., Vornlocher, H. P., Hadwiger, P., Greil, J., and Heidenreich, O. (2005b). Targeting MLL-AF4 with short interfering RNAs inhibits clonogenicity and engraftment of t(4;11)-positive human leukemic cells. Blood *106*, 3559-3566.

Thomas, M., Greil, J., and Heidenreich, O. (2006). Targeting leukemic fusion proteins with small interfering RNAs: recent advances and therapeutic potentials. Acta pharmacologica Sinica *27*, 273-281.

Traver, D., Akashi, K., Manz, M., Merad, M., Miyamoto, T., Engleman, E. G., Weissman, I. L. (2000). Development of CD8alpha-positive dendritic cells from a common myeloid progenitor. Science *290*, 2152-4.

Trentin, L., Giordan, M., Dingermann, T., Basso, G., Te Kronnie, G., and Marschalek, R. (2009). Two independent gene signatures in pediatric t(4;11) acute lymphoblastic leukemia patients. European journal of haematology *83*, 406-419.

Tschernatsch, M., Gross, O., Kneifel, N., Kaps, M., and Blaes, F. (2009). SOX-1 autoantibodies in patients with paraneoplastic neurological syndromes. Autoimmunity reviews *8*, 549-551.

Turner, M., Mee, P. J., Costello, P. S., Williams, O., Price, A. A., Duddy, L. P., Furlong, M. T., Geahlen, R. L., and Tybulewicz, V. L. (1995). Perinatal lethality and blocked B-cell development in mice lacking the tyrosine kinase Syk. Nature *378*, 298-302.

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

van der Weyden, L., Giotopoulos, G., Rust, A. G., Matheson, L. S., van Delft, F. W., Kong, J., Corcoran, A. E., Greaves, M. F., Mullighan, C. G., Huntly, B. J., and Adams, D. J. (2011). Modeling the evolution of ETV6-RUNX1-induced B-cell precursor acute lymphoblastic leukemia in mice. Blood *118*, 1041-1051.

Vardiman, J. W. (2010). The World Health Organization (WHO) classification of tumors of the hematopoietic and lymphoid tissues: an overview with emphasis on the myeloid neoplasms. Chemico-biological interactions *184*, 16-20.

Vieira, P., and Cumano, A. (2004). Differentiation of B lymphocytes from hematopoietic stem cells. Methods in molecular biology *271*, 67-76.

Voncken, J. W., Kaartinen, V., Pattengale, P. K., Germeraad, W. T., Groffen, J., and Heisterkamp, N. (1995). BCR/ABL P210 and P190 cause distinct leukemia in transgenic mice. Blood *86*, 4603-4611.

Vousden, K. H., and Lu, X. (2002). Live or let die: the cell's response to p53. Nature reviews Cancer *2*, 594-604.

Vural, B., Chen, L. C., Saip, P., Chen, Y. T., Ustuner, Z., Gonen, M., Simpson, A. J., Old, L. J., Ozbek, U., and Gure, A. O. (2005). Frequency of SOX Group B (SOX1, 2, 3) and ZIC2 antibodies in Turkish patients with small cell lung carcinoma and their correlation with clinical parameters. Cancer *103*, 2575-2583.

Wang, J. H., Avitahl, N., Cariappa, A., Friedrich, C., Ikeda, T., Renold, A., Andrikopoulos, K., Liang, L., Pillai, S., Morgan, B. A., and Georgopoulos, K. (1998). Aiolos regulates B cell activation and maturation to effector state. Immunity *9*, 543-553.

Wang, L., Guo, H., Yang, L., Dong, L., Lin, C., Zhang, J., Lin, P., and Wang, X. (2013). Morusin inhibits human cervical cancer stem cell growth and migration through attenuation of NF-kappaB activity and apoptosis induction. Molecular and cellular biochemistry *379*, 7-18.

Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A. H., and Speck, N. A. (1996). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. Proceedings of the National Academy of Sciences of the United States of America *93*, 3444-3449.

Wareing, S., Mazan, A., Pearson, S., Gottgens, B., Lacaud, G., and Kouskoff, V. (2012). The Flk1-Cre-mediated deletion of ETV2 defines its narrow temporal requirement during embryonic hematopoietic development. Stem cells *30*, 1521-1531.

Wasik, A. M., Priebe, V., Lord, M., Jeppsson-Ahlberg, A., Christensson, B., and Sander, B. (2014). Flow cytometric analysis of SOX11; a new diagnostic method for distinguishing B-CLL from MCL. Leukemia & lymphoma, 1-16.

Wat, M. J., Beck, T. F., Hernandez-Garcia, A., Yu, Z., Veenma, D., Garcia, M., Holder, A. M., Wat, J. J., Chen, Y., Mohila, C. A.*, et al.* (2012). Mouse model reveals the role of SOX7 in the development of congenital diaphragmatic hernia associated with recurrent deletions of 8p23.1. Human molecular genetics *21*, 4115-4125.

Wegner, M. (1999). From head to toes: the multiple facets of Sox proteins. Nucleic acids research *27*, 1409-1420.

Wohlbold, L., van der Kuip, H., Miething, C., Vornlocher, H. P., Knabbe, C., Duyster, J., and Aulitzky, W. E. (2003). Inhibition of bcr-abl gene expression by small interfering RNA sensitizes for imatinib mesylate (STI571). Blood *102*, 2236-2239.

Woof, J. M., and Kerr, M. A. (2004). IgA function--variations on a theme. Immunology *113*, 175-177.

Wu, Q., Tidmarsh, G. F., Welch, P. A., Pierce, J. H., Weissman, I. L., and Cooper, M. D. (1989). The early B lineage antigen BP-1 and the transformation-associated antigen 6C3 are on the same molecule. Journal of immunology *143*, 3303-3308.

Xu, W., and Li, J. Y. (2010). SOX11 expression in mantle cell lymphoma. Leukemia & lymphoma *51*, 1962-1967.

Yao, Z., Cui, Y., Watford, W. T., Bream, J. H., Yamaoka, K., Hissong, B. D., Li, D., Durum, S. K., Jiang, Q., Bhandoola, A.*, et al.* (2006). Stat5a/b are essential for normal lymphoid development and differentiation. Proceedings of the National Academy of Sciences of the United States of America *103*, 1000-1005.

Yoshimoto, M., Montecino-Rodriguez, E., Ferkowicz, M. J., Porayette, P., Shelley, W. C., Conway, S. J., Dorshkind, K., and Yoder, M. C. (2011). Embryonic day 9 yolk sac and intra-embryonic hemogenic endothelium independently generate a B-1 and marginal zone progenitor lacking B-2 potential. Proceedings of the National Academy of Sciences of the United States of America *108*, 1468-1473.

Young, N., Hahn, C. N., Poh, A., Dong, C., Wilhelm, D., Olsson, J., Muscat, G. E., Parsons, P., Gamble, J. R., and Koopman, P. (2006). Effect of disrupted SOX18 transcription factor function on tumor growth, vascularization, and endothelial development. Journal of the National Cancer Institute *98*, 1060-1067.

Yu, B. D., Hanson, R. D., Hess, J. L., Horning, S. E., and Korsmeyer, S. J. (1998). MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. Proceedings of the National Academy of Sciences of the United States of America *95*, 10632-10636.

Zelent, A., Greaves, M., and Enver, T. (2004). Role of the TEL-AML1 fusion gene in the molecular pathogenesis of childhood acute lymphoblastic leukaemia. Oncogene *23*, 4275-4283.

Zhang, C., Basta, T., Fawcett, S. R., and Klymkowsky, M. W. (2005a). SOX7 is an immediate-early target of VegT and regulates Nodal-related gene expression in Xenopus. Developmental biology *278*, 526-541.

Zhang, C., Basta, T., and Klymkowsky, M. W. (2005b). SOX7 and SOX18 are essential for cardiogenesis in Xenopus. Developmental dynamics : an official publication of the American Association of Anatomists *234*, 878-891.

Zhang, H., Alberich-Jorda, M., Amabile, G., Yang, H., Staber, P. B., Di Ruscio, A., Welner, R. S., Ebralidze, A., Zhang, J., Levantini, E.*, et al.* (2013). Sox4 is a key

oncogenic target in C/EBPalpha mutant acute myeloid leukemia. Cancer cell *24*, 575-588.

Zhang, Y., Huang, S., Dong, W., Li, L., Feng, Y., Pan, L., Han, Z., Wang, X., Ren, G., Su, D*., et al.* (2009). SOX7, down-regulated in colorectal cancer, induces apoptosis and inhibits proliferation of colorectal cancer cells. Cancer letters *277*, 29-37.

Zhao, X., Ghaffari, S., Lodish, H., Malashkevich, V. N., and Kim, P. S. (2002). Structure of the Bcr-Abl oncoprotein oligomerization domain. Nature structural biology *9*, 117-120.

Zhuang, Y., Soriano, P., and Weintraub, H. (1994). The helix-loop-helix gene E2A is required for B cell formation. Cell *79*, 875-884.

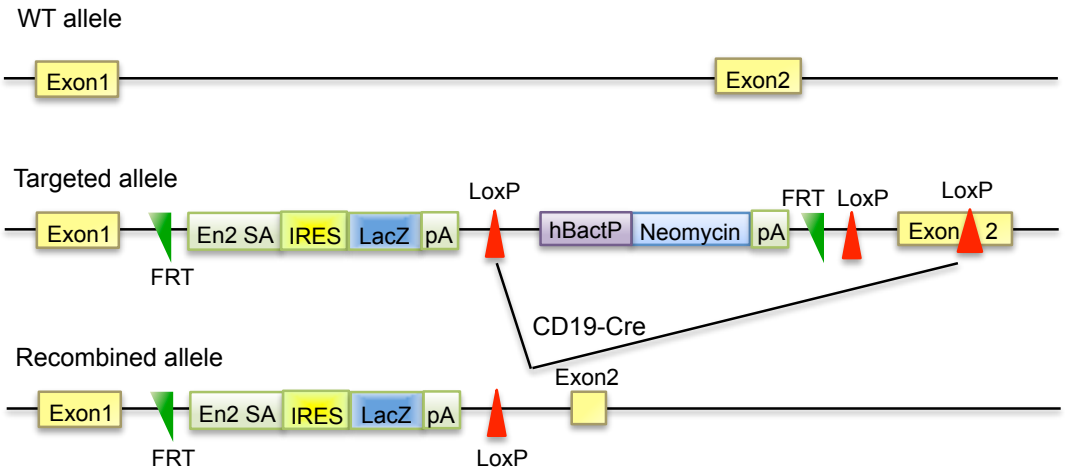
Zovein, A. C., Hofmann, J. J., Lynch, M., French, W. J., Turlo, K. A., Yang, Y., Becker, M. S., Zanetta, L., Dejana, E., Gasson, J. C*., et al.* (2008). Fate tracing reveals the endothelial origin of hematopoietic stem cells. Cell stem cell *3*, 625-636.

Zuber, J., Radtke, I., Pardee, T. S., Zhao, Z., Rappaport, A. R., Luo, W., McCurrach, M. E., Yang, M. M., Dolan, M. E., Kogan, S. C*., et al.* (2009). Mouse models of human AML accurately predict chemotherapy response. Genes & development *23*, 877-889.

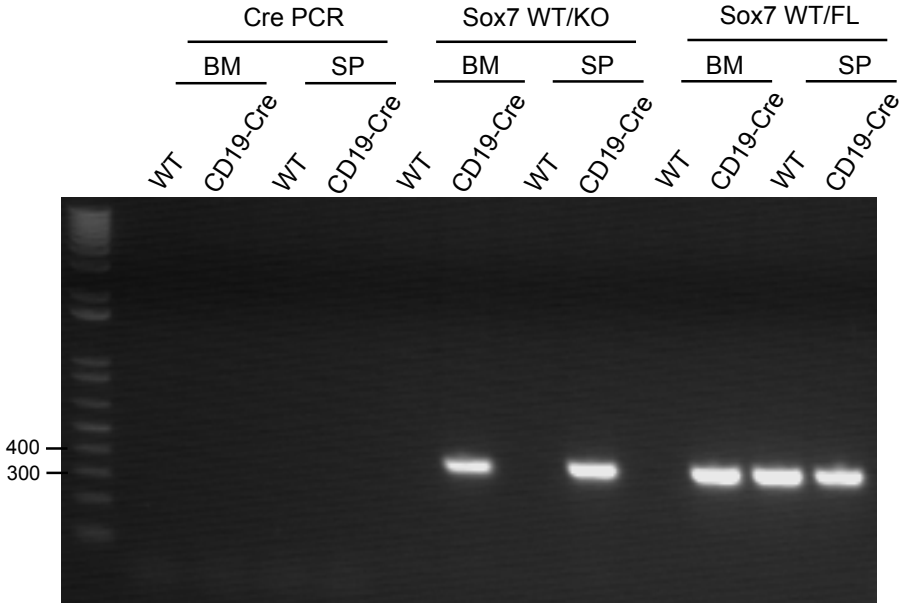
Appendices

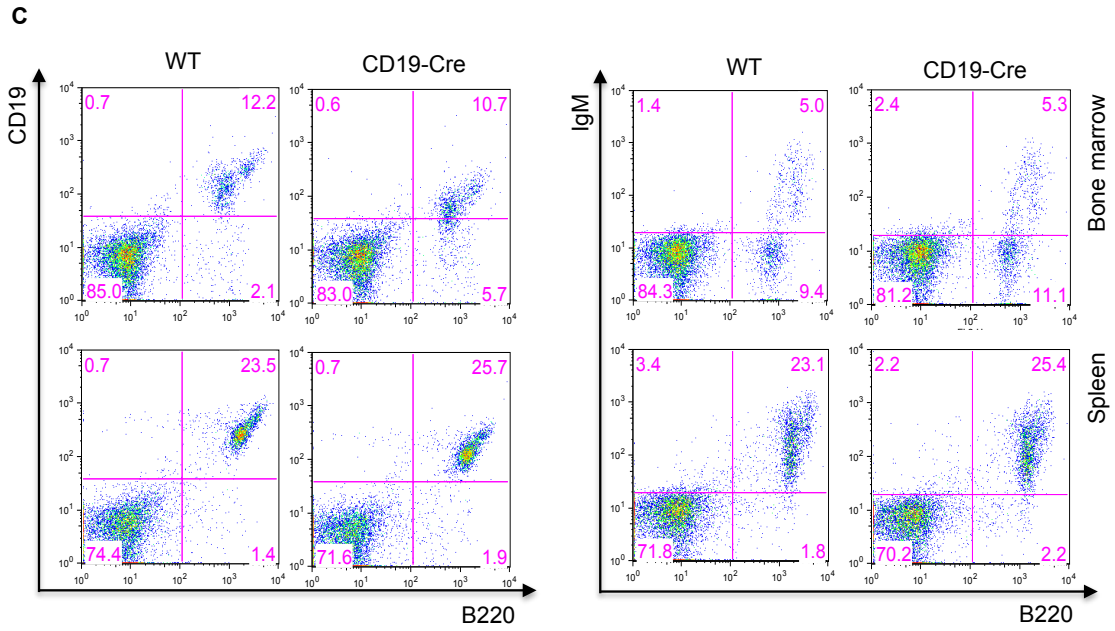
Appendix 1.

A



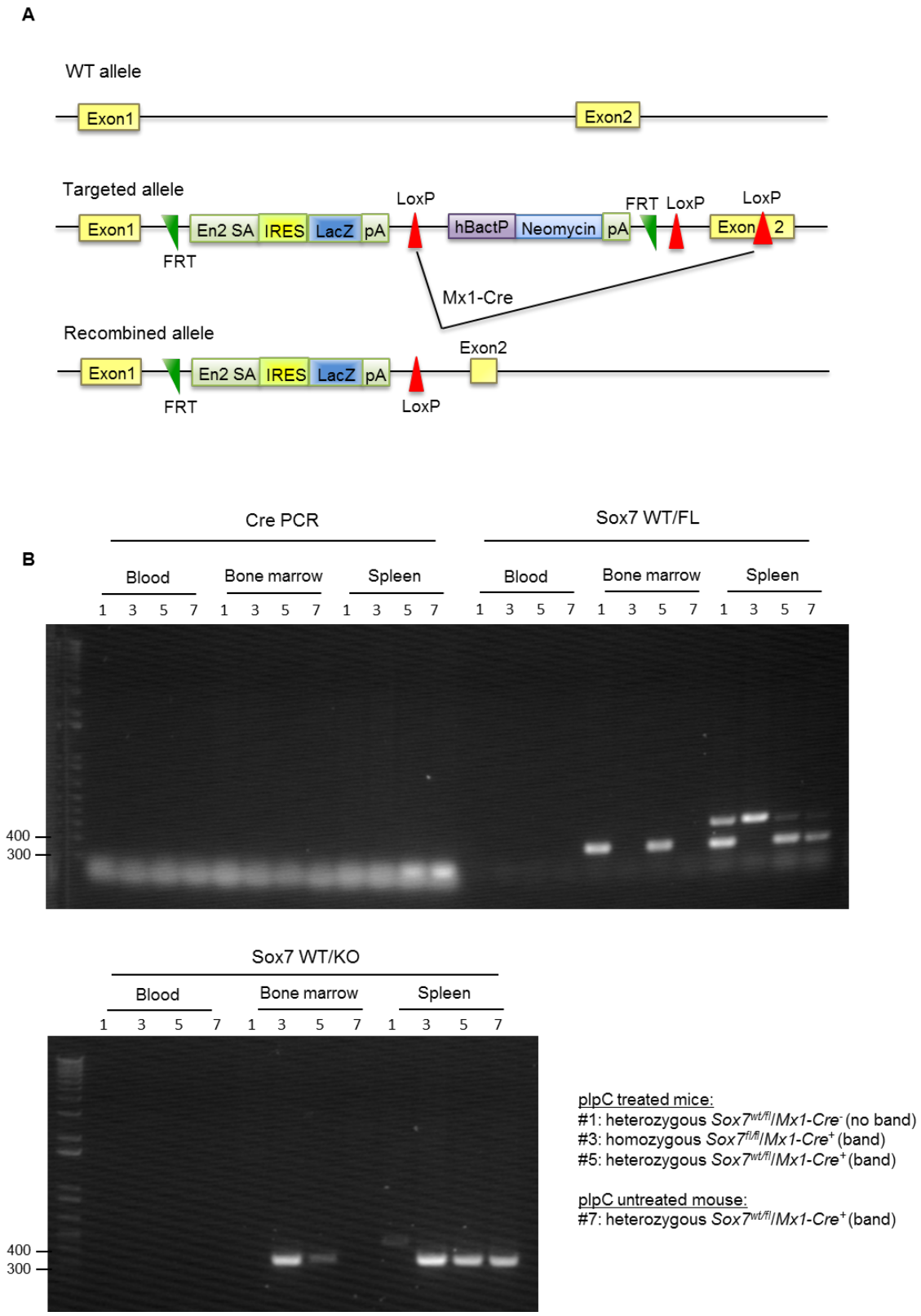
B

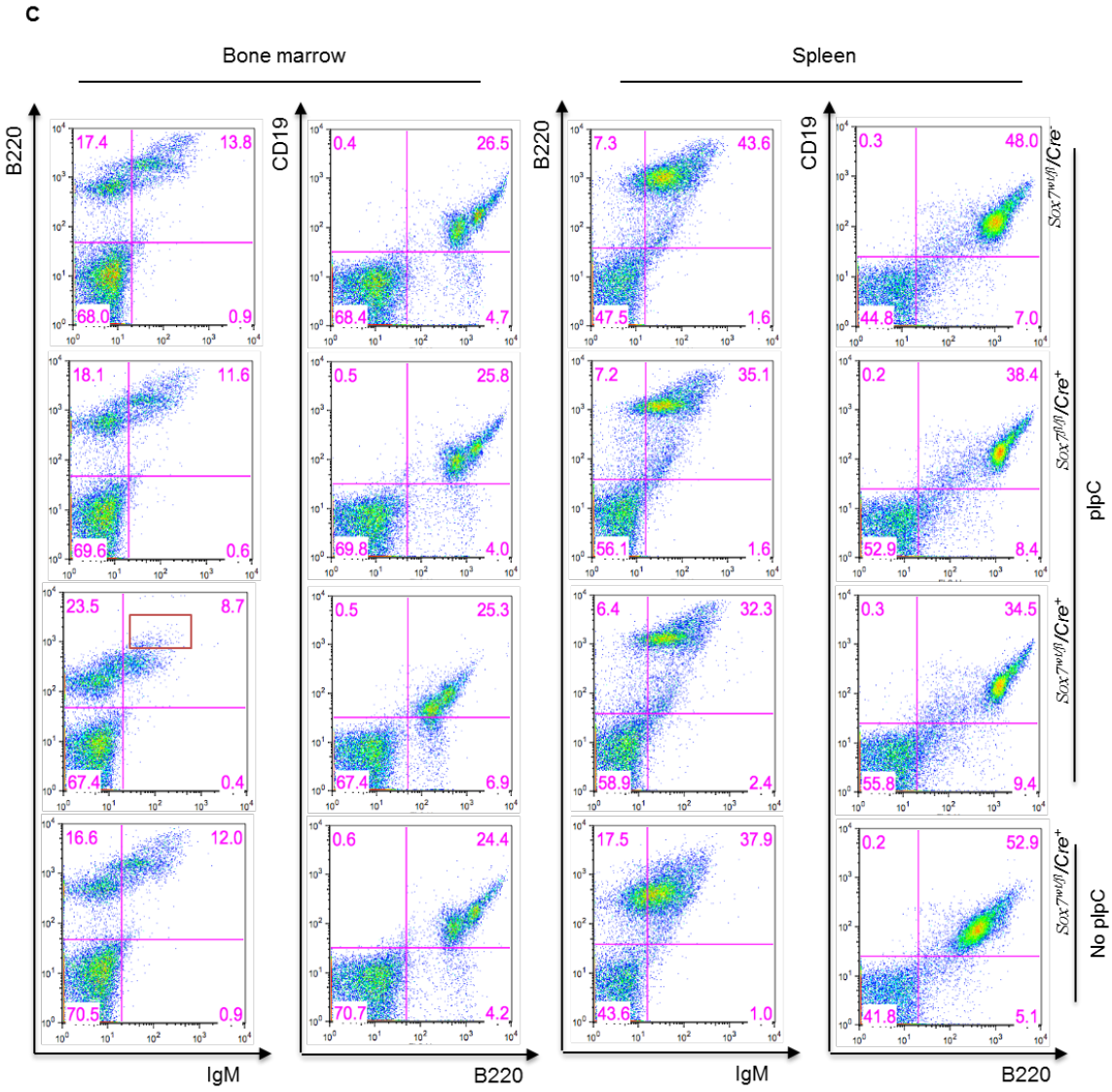




Deletion of *Sox7* using a costitutive Cre/LoxP recombinase system.
A) Schematic representation of the *Sox7* WT, targeted and recombined alleles. Cre recombinase, expressed in $CD19^{+}$ cells, deletes the region between the two LoxP sites. B) Bone marrow and spleen genotyping PCR. Agarose gel electrophoresis detects 357 bp (KO) and 1832 bp (WT) PCR products originating form $Sox7^{wt/ko}$ and 345 bp (fl) and 195 bp (WT) PCR products originating form $Sox7^{wt/fl}$; Cre band: 726 bp; BM: bone marrow; SP: spleen C) Representative FACS plots of CD19, B220 and IgM expression profiles from bone marrow and spleen cells of *wild-type* and *CD19-Cre* mice. Percentage cells are shown in the indicated quadrants. FACS plots are representative of 3 independent experiments.

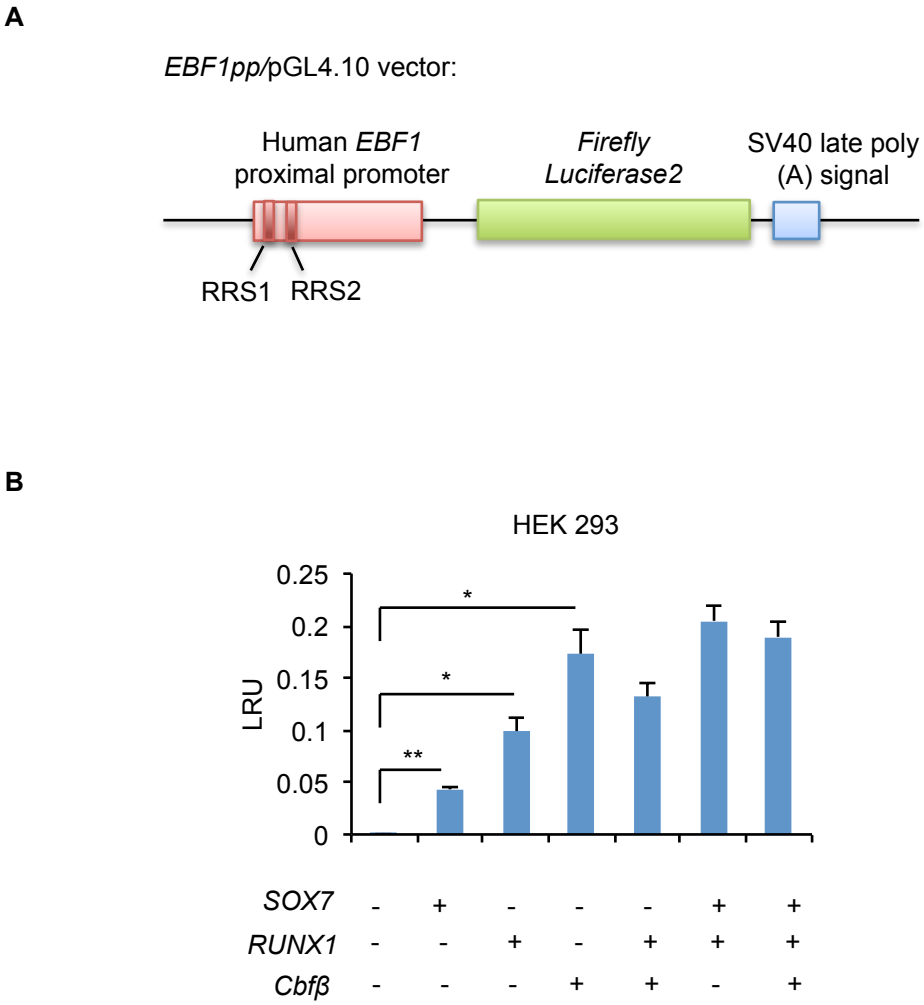
Appendix 2.


A
WT allele
Exon1
Exon2
Targeted allele
Exon1
En2 SA
IRES
LacZ
pA
LoxP
hBactP
Neomycin
pA
FRT
LoxP
LoxP
Exon 2
FRT
Mx1-Cre
Recombined allele
Exon1
En2 SA
IRES
LacZ
pA
Exon2
FRT
LoxP
B
Cre PCR
Sox7 WT/FL
Blood
Bone marrow
Spleen
Blood
Bone marrow
Spleen
1 3 5 7 1 3 5 7 1 3 5 7 1 3 5 7 1 3 5 7 1 3 5 7
400
300
Sox7 WT/KO
Blood
Bone marrow
Spleen
1 3 5 7 1 3 5 7 1 3 5 7
400
300
plpC treated mice:
#1: heterozygous Sox7wt/fl/Mx1-Cre- (no band)
#3: homozygous Sox7fl/fl/Mx1-Cre+ (band)
#5: heterozygous Sox7wt/fl/Mx1-Cre+ (band)
plpC untreated mouse:
#7: heterozygous Sox7wt/fl/Mx1-Cre+ (band)



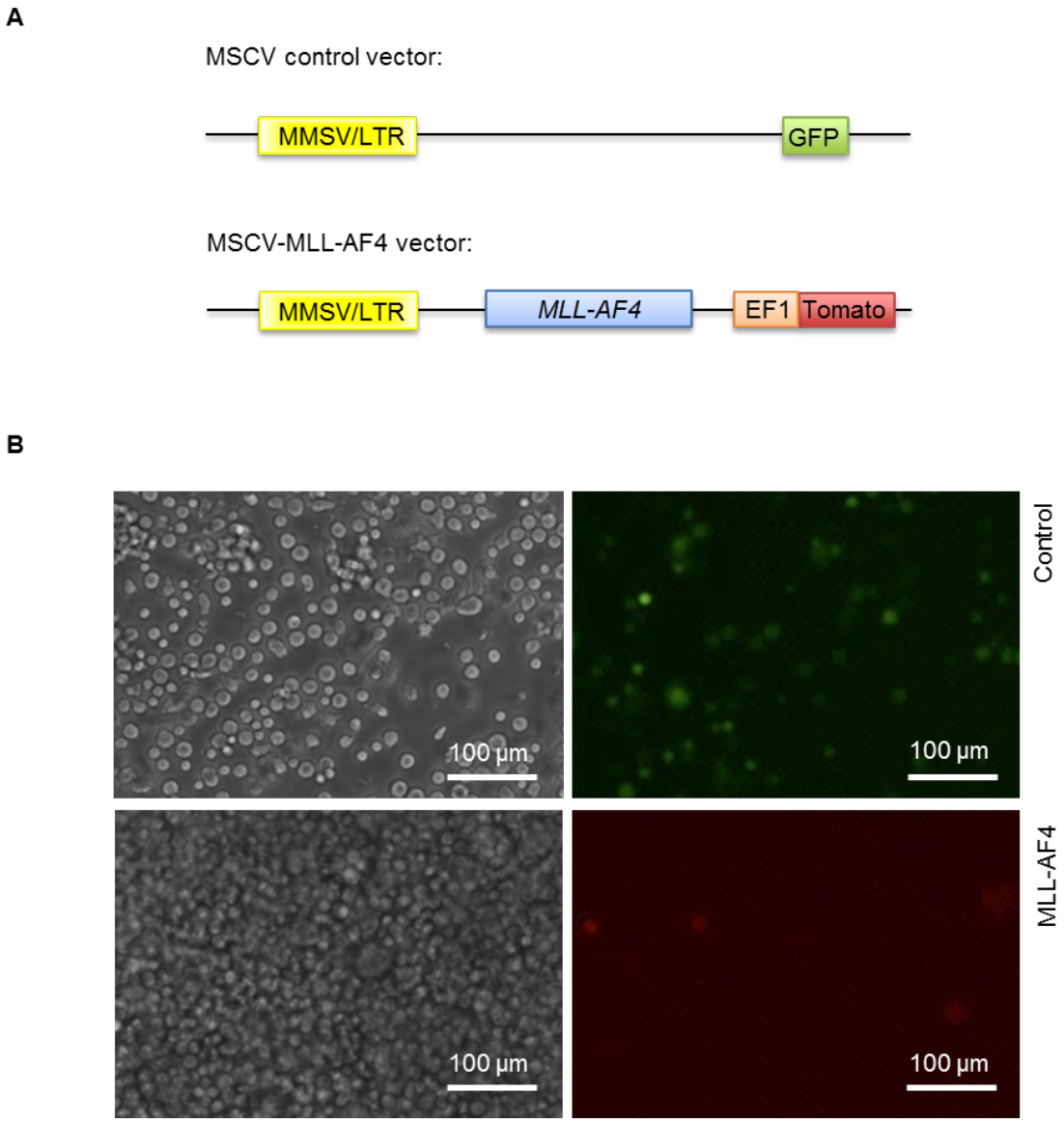
Deletion of *Sox7* using an inducible Cre/LoxP recombinase system.
A) Schematic representation of the *Sox7* WT, targeted and recombined alleles. Cre recombinase is induced by injecting intra-peritoneally polyinosinic-polycytidylic acid (plpC) and deletes the region between the two *LoxP* sites in haematopoietic stem cells. B) Blood, bone marrow and spleen genotyping PCR. Agarose gel electrophoresis detects 357 bp (KO) and 1832 bp (WT) PCR products originating form $Sox7^{wt/ko}$ and 345 bp (fl) and 195 bp (WT) PCR products originating form $Sox7^{wt/fl}$; Cre band: 726 bp; (plpC treated mice: #1: heterozygous $Sox7^{wt/fl}$/*Mx1-Cre*$^{-}$ (no band), #3: homozygous $Sox7^{fl/fl}$/*Mx1-Cre*$^{+}$ (band), #5: heterozygous $Sox7^{wt/fl}$/*Mx1-Cre*$^{+}$ (band); plpC untreated mouse: #7: heterozygous $Sox7^{wt/fl}$/*Mx1-Cre*$^{+}$ (band)). C) Representative FACS plots of CD19, B220 and IgM expression profiles from bone marrow and spleen cells of heterozygote or homozygote *Sox7*/*Mx1-Cre* mice treated or not with plpC. Percentage cells are shown in the indicated quadrants. FACS plots are representative of 1 independent experiment.

Appendix 3.



Cloning of *EBF1* proximal promoter and Luciferase assay.
A) Representative scheme of the Luciferase vector. The pGL4.10 vector contains the human *EBF1* proximal promoter sequence and the FireFly-Luciferase sequence. B) Luciferase assays show the activity of EBF1pp/pGL4.10 construct activated by human *SOX7*, human *RUNX1* and *Cbfβ* plasmids alone or in combination in HEK293 cells. Bars represent the mean Luciferase intensity relative to control ± S.E.M (n=3; * P< 0.05; ** P< 0.01). Bar graph is representative of 2 independent experiments. RLU means relative luciferase units.

Appendix 4.



Cloning of MLL-AF4 translocation.
A) Representative scheme of the MSCV vectors. The MSCV control vector contains the GFP sequence while the MSCV-MLL-AF4 vector contains the human MLL-AF4 sequence together with Tomato reporter gene under the control of EF1 promoter. B) Representative picture of $cKIT^{+}$ bone marrow cells transduced with MLL-AF4 or control retrovirus. Cells infected with the control retrovirus, expressing GFP protein, result green while cells infected with the MLL-AF4 retrovirus, containing Tomato protein, result red (n=1; scale bar: 100µm).