

Washington University in St. Louis

Washington University Open Scholarship

All Theses and Dissertations (ETDs)

Spring 2-27-2014

Cold Shock Domain Proteins in Normal and Leukemic Hematopoiesis

Cheng Li

Washington University in St. Louis

Follow this and additional works at: <https://openscholarship.wustl.edu/etd>

Recommended Citation

Li, Cheng, "Cold Shock Domain Proteins in Normal and Leukemic Hematopoiesis" (2014). *All Theses and Dissertations (ETDs)*. 1247.

<https://openscholarship.wustl.edu/etd/1247>

This Dissertation is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Molecular Cell Biology

Dissertation Examination Committee:

Timothy Ley, Chair

Kyunghee Choi

Michael Tomasson

Matthew Walter

Jason Weber

John Welch

Cold Shock Domain Proteins in Normal and Leukemic Hematopoiesis

by

Cheng Li

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2014

St. Louis, Missouri

TABLE OF CONTENTS

	Page
List of Figures	iii
List of Tables	v
Acknowledgement	vii
Abstract of the Dissertation	xi
Chapter 1	
Introduction	1
References	25
Chapter 2	
Cold Shock Domain Proteins in Normal and Leukemic Hematopoiesis	43
References	67
Chapter 3	
Genetic and Functional Heterogeneity of Induced Pluripotent Stem Cells Derived from Adult Skin Fibroblasts	89
References	111
Chapter 4	
Summaries and Future Directions	152
References	158
Resume	160

LIST OF FIGURES

Chapter 1: Introduction

Figure Legends	39
Figure 1-1. Expression of CSD proteins in human and mouse cell lines	40
Figure 1-2. Expression of CSD genes in mouse and human flow-sorted bone marrow cells and leukemia samples	41
Figure 1-3. Developmental stage-specific and tissue-specific expression patterns of CSD proteins in wildtype mice	42

Chapter 2: Cold Shock Domain Proteins in Normal and Leukemic Hematopoiesis

Figure Legends	69
Figure 2-1. Resting hematopoiesis in <i>Ybx1</i> ^{+/-} and <i>Msy4</i> ^{-/-} mice	73
Figure 2-2. Competitive repopulation of <i>Msy4</i> ^{-/-} mice	74
Figure 2-3. Expression array data comparing KLS cells from <i>Msy4</i> ^{-/-} and <i>Msy4</i> ^{+/+} mice	75
Figure 2-4. Characterization of <i>Ybx1</i> ^{-/-} E14.5 fetal livers	76
Figure 2-5. Hematopoiesis in mice engrafted with <i>Ybx1</i> ^{-/-} E14.5 fetal liver cells	77
Figure 2-6. Characterization of <i>Ybx1</i> ^{lox/+} mice	79
Figure 2-7. Effects of the <i>Ybx1</i> floxed allele on adult hematopoiesis	80

Figure 2-8. Effects of <i>Msy4</i> deficiency on serial replating by <i>Ctsg</i> - <i>PML-RARA</i> bone marrow cells	81
Figure 2-9. Effects of <i>Msy4</i> or <i>Ybx1</i> deficiency on a retroviral MLL-AF9 leukemia model	82
Figure 2-10. Optimal induction of floxing by TAT-cre	84
Figure 2-11. Effects of <i>Ybx1</i> and <i>Msy4</i> deficiency on aberrant replating of bone marrow cells expressing MLL-AF9	85
Figure 2-12. Expression of the truncated Ybx1 protein predicted from the floxing of <i>Ybx1</i> conditional allele	86

Chapter 3: Genetic and Functional Heterogeneity of Induced Pluripotent Stem Cells Derived from Adult Skin Fibroblasts

Figure Legends	116
Figure 3-1. Hematopoietic differentiation from mESCs/miPSCs	118
Figure 3-2. Hematopoietic potential of the 24 miPSC clones	119
Figure 3-3. Mutation landscapes of the 24 miPSC clones	121
Figure 3-4. Expression array comparing six miPSC clones with normal vs low hematopoietic potential	123

LIST OF TABLES

Chapter 1: Introduction

Table 1-1. Protein binding partners of YBX1	5
Table 1-2. YBX1 in transcription regulation	6
Table 1-3. CSD protein expression in clinical specimen of human tumors	13

Chapter 2: Cold Shock Domain Proteins in Normal and Leukemic Hematopoiesis

Table 2-1. Genotype distribution from <i>Ybx1^{lox/+}</i> crosses	87
Table 2-2. TAT-cre testing on <i>ROSA-lox-STOP-lox-YFP</i> bone marrow cells. Floxing efficiency is shown in YFP percentage by flow cytometry	88

Chapter 3: Genetic and Functional Heterogeneity of Induced Pluripotent Stem Cells Derived from Adult Skin Fibroblasts

Table 3-1. Summary of the iPSC heterogeneity literature	125
Table 3-2. OP9 and FBS lot compatibility (Based on number of CFUs produced by B6 ESC-derived hematopoietic progenitors) with the hematopoietic differentiation assay	126
Table 3-3. Whole exome sequencing coverage	127
Table 3-4. Validation array coverage	128

Table 3-5. GFP and Oct3/4 expression of miPSC clones	129
Table 3-6. Private mutations in all 24 miPSC clones	130
Table 3-7. Common mutations shared among miPSC clones	144
Table 3-8. OSK lentiviral integration sites	148
Table 3-9. Integration “hotspots”	150
Table 3-10. Expression changes in genes located, upstream, or downstream of integration sites	151

ACKNOWLEDGEMENTS

I owe a tremendous debt of gratitude to my thesis mentor Dr. Timothy Ley. Simply said, Tim taught me everything I know about being a scientist. His achievements in better understanding the genomics of acute myeloid leukemia and his career-long dedication to understand leukemia biology in an effort to help improve the treatment of leukemia patients inspire me to make an impact in improving patients' lives. His patience and encouragement as a mentor supported me through many difficult times when my projects were not going as planned. He taught me how to be optimistic but critical; how to think beyond the immediate implications of the results; and how to write and present a story. More than just teaching me science, I would not have become the person I am without Tim.

Next, I would like to acknowledge the guidance and help from many members of the Ley Lab, both past and present. I worked closely with John Welch during my first year in the lab. He directly, and very patiently, taught me every essential laboratory technique and much about the biology of cold shock domain proteins. Over the last several years, I have been fortunate to benefit from John's formal mentorship as a member of my thesis committee. I have never stopped being amazed by his brilliance and creativity. I was beyond lucky to have shared the same bay for three years with Lukas Wartman, who also became a great friend. Besides being one of the smartest people I know, his courage and strength

throughout his own unimaginable struggle with leukemia and his commitment to fight back as a cancer scientist continue to remind me of why I am in this fight, and to never give up. Jeff Klco became my go-to person for everything as soon as he showed up in lab. Brilliant as every other fellow who has been in the Ley Lab, Jeff has also been the most optimistic and encouraging. I would like to thank former graduate students, Sheng Cai and Maggie Young, and our indefatigable lab manager, Dan George, for their friendship and significant support throughout my graduate studies. I would also like to thank the other members of the Ley Lab: Mieke Hoock, Tami Lamprecht, David Russler-Germain, Angela Verdoni, David Spencer, and Chris Cole. All of my lab mates have provided me much help over the course of my studies, and I am deeply appreciative for their efforts. I would also like to acknowledge our extended lab family: the other labs in our Section, the members of the Embryonic Stem Cell Core, and the members of the Siteman Flow Cytometry Core. They have all been an essential part of my graduate school experience and made my own studies possible.

I would like to thank the members of my thesis committee: Jason Weber, Matt Walter, John Welch, Michael Tomasson, KC Choi, and former member Jim Hsieh, for their generous guidance and advice. I would like to especially thank Jason Weber for being a part of my graduate school life from the very beginning-- he interviewed me for graduate school via an Internet connection while I was in Beijing. I am very thankful for his unwavering support.

Graduate school would not have been the same without the wonderful friends I have met in DBBS, CSSA, and BALSAs: Jie, Shuo, Dan, Fei, Muhan, Rei, Giriand many others. I would also like to thank my dance family throughout St. Louis, and my fellow DBBS dance warriors, Annie and Scott, for sharing the art of movement that kept me centered for the past four years, and for introducing me to Cliff who kept me happy and sane through my thesis writing.

Finally, I would like to thank my mom, Min, for supporting me unconditionally for all my life and for all the sacrifices she made to raise me as a single mom for the past 20 years.

DEDICATION

I dedicate this thesis to the memory of my dad, Ningwu, who lost his life to lung cancer 20 years ago.

ABSTRACT OF THE DISSERTATION

Cold Shock Domain Proteins in Normal and Leukemic Hematopoiesis

by

Cheng Li

Doctor of Philosophy in Biology and Biological Sciences

Molecular Cell Biology

Washington University in St. Louis, 2014

Professor Timothy Ley, Chairperson

Cold shock domain (CSD) proteins are the most evolutionarily conserved family of nucleic-acid binding proteins. There are four functional genes that contain CSDs in humans and mice: *YBX1*, *MSY4*, *MSY2*, and *CSDE1*. *YBX1* is overexpressed in most cancers, and is frequently associated with poor outcomes and chemotherapy resistance. Both *YBX1* and *MSY4* are highly expressed in normal hematopoietic progenitors, and both are downregulated with terminal myeloid differentiation; both genes are highly expressed in virtually all cases of acute myeloid leukemia (AML). These two genes are functionally redundant as well: *Msy4* has been shown to complement *Ybx1* function in late-stage embryogenesis in mouse knockout models. Nevertheless, most studies that have sought to clarify the role of *YBX1* in cancer have failed to consider a possible role

for *MSY4* complementation in cells where both genes are expressed. In mice deficient for either *Ybx1* or *Msy4*, hematopoiesis is not altered. However, the loss of both proteins leads to a reduced ability for *MLL-AF9* (a potent leukemia-initiating gene) expressing bone marrow cells to proliferate and serially replat *in vitro*, suggesting that *Ybx1* and *Msy4* have redundant functions in this model system. Since these proteins are involved in the prevention of senescence during proliferative stress, the inhibition of both proteins may provide a novel strategy for the treatment of AML and other cancers.

To create a system to study the roles of *Ybx1* and *Msy4* in the earliest stages of hematopoiesis (i.e. progression from pluripotency to committed hematopoietic stem cells), we developed a system to study the production of hematopoietic stem and progenitor cells from murine embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs). By comparing the hematopoietic potentials of ESCs, and 24 independent iPSC clones obtained from a single adult mouse, we discovered considerable functional heterogeneity among the clones. To determine whether the basis of this heterogeneity was genetic, we sequenced the exomes of all 24 clones. Although each had a set of private mutations that defined its clonal origins, none of the mutations readily explained why some clones had a reduced potential to form hematopoietic progenitors *in vitro*. Finally, we compared the expression profiles of clones with extreme outlier phenotypes for hematopoiesis *in vitro*; this study yielded a small set of candidate genes (including *Wt1* and *Lef1*) that could be relevant for the hematopoietic differentiation potential of mouse iPSCs. These findings have provided new

insights into the origins of genetic heterogeneity among iPSC clones, and may ultimately provide new information about the genes that govern the earliest steps of hematopoietic commitment.

Chapter 1

Introduction

1.1 Cold shock domain proteins are evolutionarily conserved

Cold shock domain (CSD) proteins belong to one of the most evolutionarily conserved nucleic-acid-binding protein families. Structurally, CSD proteins are highly conserved among bacteria, archaea, plants, and vertebrates. Functionally, CSD proteins are involved in various stress response throughout evolution. Interestingly, CSD proteins are not found in yeast.

1.1.1 Bacterial cold shock proteins

Bacterial cold shock proteins (Csps) are small proteins (~7kDa) containing a single CSD, and are found in more than 50 bacterial species.¹ They bind single-stranded RNA and DNA (ssRNA/ssDNA), but not double-stranded DNA (dsDNA).² Three-dimensional (3-D) structural analysis of CspA in *Escherichia coli* revealed a β barrel formed by five antiparallel β strands, adopting an oligonucleotide and/or oligosaccharide binding (OB) fold.^{1,2} Strands β 2 and β 3 contain conserved ribonucleoprotein1 (RNP1) and RNP2 motifs respectively, which are thought to be responsible for the protein-nucleic acid interaction of CspA.^{2,3}

Csps have been shown to respond to different environmental stresses, and also thought to be important during normal growth conditions for bacteria. Among the 11 family members of Csps found in *E. coli*, CspA, CspB, CspG, and CspI respond to temperature downshift, CspD is induced by nutrition starvation, while CspC and CspE are constitutively expressed.^{1,4} The cold-shock response leads to cessation of growth and reduction in protein synthesis in bacteria. Csps are thought to mediate this process by functioning as RNA chaperones to destabilize and recover functionality of the RNAs, which tend to be folded into unfavorable structures at low temperatures. Csps are also involved in transcription, translation, and RNA turnover.^{3,5,6}

1.1.2 Archaeal Csps

In archaea, Csp homologues have been identified in a small number of psychrophiles, including *Methanogenium frigidum* (which have the lowest known

upper growth temperature limit at 18°C); they do not appear to be present in any thermophiles thriving at relatively high temperatures (45-80°C). Archaeal Csps have a similar structure as bacterial Csps, with a higher content of solvent-exposed basic residues located on the nucleic acid binding surface. They have been shown to complement a cold-sensitive growth defect in *E. coli*. Further, purified *M. frigidum* Csp was found to bind *E. coli* ssRNA.⁷

1.1.3 CSD proteins in plants

Eukaryotic CSD proteins contain C- and/or N-terminal auxiliary domains in addition to a CSD. CSD proteins are presented across lower and higher plants. They all contain a highly conserved CSD and a diverse combination of Glycine-rich regions and CCHC zinc fingers on the C-terminus. CSD proteins have been shown to be essential for acquisition of freezing tolerance in plants,⁸ and are involved in other stress responses, such as rapid cell division, dehydration, and salt stress.⁹

1.1.4 CSD proteins in nematodes

Caenorhabditis elegans Lin-28 contains one CSD and two CCHC zinc-finger motifs. These domains are thought to cooperate in RNA target recognition. Lin-28 can regulate gene expression associated with development, differentiation, and cancer progression by directly binding to target mRNAs or indirectly regulating microRNA (miRNA) processing.³

1.1.5 Vertebrate CSD proteins

Human CSD proteins Y-box binding protein 1 (YBX1), Cold shock domain protein A (MSY4), and Y-box binding protein 2 (YBX2) share about 40% CSD homology with bacteria, and more than 84% overall amino acid homology with their mouse homologues. These proteins all contain a divergent alanine- and proline-rich N-terminal domain and a structurally similar C-terminal domain with basic/acidic amino acid repeats (B/A repeat)⁶.

Vertebrate CSD proteins have been associated with cold-adaptation¹⁰, but more importantly, they are associated with stress responses more common in vertebrates, i.e. maintaining rapid cell growth. Both Ybx1 and Msy4 are highly expressed throughout murine embryogenesis.¹¹ High levels of Ybx1 are detected *in vivo* in regenerating liver tissue following chemical-induced damage¹² or hepatectomy¹³ and actively proliferating colorectal epithelial glands¹⁴. In various human cell types, YBX1 levels increase in response to mitogenic stimuli, including cytokine-stimulation of T cells¹⁵, serum-activation of fibroblasts¹³, and agonist-stimulation of endothelial cells¹⁶.

Another family member, Cold Shock Domain Containing E1 (CSDE1, also known as Upstream of N-ras, or UNR), has a unique structure containing five copies of the CSD and, unlike other family members, has not been associated with stress response. It is involved in the regulation of translation and mRNA stability in translation.³ For example, CSDE1 is required for efficient initiation of translation from the internal ribosome entry sites (IRES) of both Rhinovirus and Poliovirus¹⁷; it also plays a key role in translationally coupled mRNA decay mediated by the *c-fos* major protein coding-region determinant¹⁸.

1.2 Structural and functional organization of the YBX1 protein

Ybx1 is the most well known member among all the vertebrate CSD proteins, and has been the most widely studied, both structurally and functionally. Through its interactions with DNA, RNA, and proteins, Ybx1 is thought to play an essential role in normal cellular functions and stress responses, including transcription regulation, translation regulation, and DNA repair.

1.2.1 Structure and properties of YBX1 protein

The cold shock domain (CSD) of YBX1 contains a β -barrel structure comparable to bacterial cold shock proteins, with a similar arrangement of RNA-binding motif. It binds only weakly and non-specifically to DNA⁶, but has a strong affinity to single-stranded pyrimidine-rich sequences and triplex/single-stranded H-DNA.^{12,19} The CSD is thought to be required for nuclear import of the Ybx1

protein³ and RNA target sequence recognition²⁰, and has been shown to assist the formation of Ybx1 fibrils.²⁰

The C-terminal domain consists of a B/A repeat, with 30 amino acids of alternate regions of basic or acidic amino acids. It is highly divergent among vertebrate CSD proteins and is used to distinguish between germ and somatic cell types.⁹ It has been shown to have a strong but non-specific affinity for ssDNA/ssRNA *in vitro*, as it interacts with negatively charged phosphate groups of nucleic acids.⁶ It is thought to be involved in protein binding, including dimer formation and homomultimerization.^{9,20} It also contains the nuclear localization signal (NLS) and the cytoplasmic retention site (CRS).²⁰

Finally, the alanine- and proline-rich N-terminal Domain is thought to be a *trans*-activation domain involved in protein interaction.²⁰

YBX1 binds to a variety of proteins (**Table 1-1**), and it can bind to a number of proteins using more than one domain.²⁰ YBX1 also binds to both DNA and RNA, in both sequence-specific and sequence-nonspecific fashions.

Table 1-1. Protein binding partners of YBX1

Functional category	Protein binding partners
Transcription Factors	Activating Protein-2 (AP2) ²¹ CCCTC-binding factor (CTCF) ²² Interferon Regulatory Factor-1 (IRF-1) ²³ p53 ^{21,24} p65 ²⁵ Smad3 ²⁶ Sox21 ²⁷ Yin Yang 1 (YY1) ²⁸
RNA-binding Proteins	Heterogeneous nuclear ribonuclearprotein K (HnRNPK) ^{29,30} SRp30c ³¹ U2 auxiliary factor (U2AF) ³²
Viral Proteins	Human Polyomavirus JC Virus Large T Antigen (JCV LT) ³³ Tat ³⁴

DNA Repair Proteins	Human apurinic/aprimidinic (AP) endonuclease (APE1/Ref-1) ^{35,36} Endonuclease III ³⁷ DNA ligase III α ³⁸ DNA polymerase β ³⁸ DNA polymerase δ ³⁹ Ku80 ³⁹ MutS protein homolog (MSH2) ³⁹ Nei endonuclease VIII-like 2 (NEIL2) ³⁸ Proliferating Cell Nuclear Antigen (PCNA) ⁴⁰ Werner syndrome, RecQ helicase-like (WRN) ^{39,41}
Cytoskeletal Proteins	Actin filaments ⁴² Microtubules ⁴³
Oncogenes	Ewings sarcoma breakpoint region 1 (EWS) ⁴⁴ Fused in sarcoma (FUS/TLS) ⁴⁴
Others	Ankyrin repeat domain 2 (Ankrd2) ⁴⁵ Cardiac Ankyrin Repeat Protein (CARP) ^{45,46} Cyclin D1 ⁴⁷ Histone deacetylase (HDAC2) ⁴⁸ Heat shock protein (HSP60) ⁴⁹ Iron-Regulatory Protein 2 (IRP2) ³⁰ Karyopherin β 2 ⁵⁰ Purine-rich element binding protein A (Pura) ^{51,52} Pur β ⁵² Ybx1: dimer ³⁰ or fibril ⁵³ Y-box Protein-associated Acidic Protein (YBAP1) ⁵⁴

1.2.2 YBX1 in transcriptional regulation

YBX1 regulates transcription of a number of genes involved in normal cellular functions, including cell cycle, apoptosis, immune responses, as well as stress responses, tumor growth, and multidrug resistance. It has also been shown to regulate the transcription of some viral genes. (**Table 1-2**)

Table 1-2. YBX1 in transcription regulation ^{6,9,20}

Regulation	Regulated gene
Activation	Adenovirus late genes under control of promoter E2 <i>Chemokine (C-C motif) ligand 5 (CCL5)</i> <i>CD44</i>

	<p> <i>CD49f</i> <i>cyclin A and cyclin B1</i> <i>DNA polymerase α</i> <i>Epidermal growth factor receptor (EGFR)</i> <i>v-erb-b2 erythroblastic leukemia viral oncogene homolog 2 (ERBB2/HER-2)</i> <i>Fas</i> <i>Gelatinase A/matrix metalloproteinase 2</i> <i>Genes under control of HIV-1 TAR-promoter</i> <i>GM-CSF</i> <i>LRP/MVP</i> <i>Human Multidrug transporter (MDR1)</i> <i>MET proto-oncogene (MET)</i> <i>myosin light-chain 2v (MLC-2v)</i> <i>p21</i> <i>PDGF B</i> <i>Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha (PI3KCA)</i> <i>Polyomavirus JCV promoters (late)</i> <i>Protein tyrosine phosphatase, non-receptor type 1 (PTP1B)</i> <i>Smad7</i> </p>
Repression	<p> <i>ATP-binding cassette, sub-family C (CFTR/MRP), member 2 (Abcc2/Mrp2)</i> <i>α-actin</i> <i>c-Myc</i> <i>COLα1</i> <i>COLα2</i> <i>Carbamoyl-phosphate synthase 1 (CPS-1)</i> <i>Fas</i> <i>γ-globin</i> <i>GM-CSF</i> <i>Heat shock 70kDa protein 5 (HSPA5/GRP78/BiP)</i> <i>MVP major vault protein (MVP/LRP)</i> <i>MMP12</i> <i>MMP13</i> <i>MHC I</i> <i>MHC II (HLA DRα, I-Aβ)</i> <i>p21</i> <i>p53</i> <i>Polyomavirus JCV promoters (early)</i> <i>Thyrotropin receptor</i> <i>VEGF</i> </p>

YBX1 was first identified as a transcription factor by its ability to bind to the inverted CCAAT motifs (Y-box sequence) of *Major Histocompatibility Complex (MHC) class II* promoter.⁵⁵ Y-box-binding and recruitment of other transcription factors has been shown to upregulate *myosin light-chain 2v* and human *Multidrug Transporter (MDR1)*, and to downregulate human *collagen α 2 (COL α 2)*.⁶ However, YBX1 has also been reported to regulate transcription independent of Y-box binding.

YBX1 can bind to other transcription factors to co-activate gene transcription. For example, Ybx1 has been reported to interact with TP53 and activate transcription of *P21*²⁴; it has also been shown to activate the *Granulocyte-Macrophage Colony-stimulating Factor (GM-CSF)* promoter in Jurkat T cells through interaction with transcription factors such as RELA/NF- κ B p65⁵⁶.

YBX1 can also bind to single-stranded regions of promoters (S1-nuclease hypersensitive sites), with a higher affinity to pyrimidine-rich regions. For some genes, including *Collagen, type I alpha 2 (COL α 2)* and *Vascular endothelial growth factor A (VEGF)*, binding of YBX1 to pyrimidine-rich promoter regions directly correlates to inhibition of their transcription. It is hypothesized that binding of YBX1 inhibits transcription factors from binding to the DNA. For other genes, such as *c-myc*, *metalloproteinase 2*, and human JCV polyoma virus early and late genes, binding of YBX1 to pyrimidine-rich promoter regions stimulates transcription. It is thought that YBX1 can recruit transcription factors to these promoters and as a result activates their transcription.²⁰ The mechanism for both transcription inhibition and activation through YBX1-pyrimidine-rich-DNA binding remains unclear.

1.2.3 YBX1 in translation regulation

YBX1 plays a vital role in various stages of mRNA life cycle, including stabilizing mRNA as a chaperone in cytoplasm, pre-mRNA splicing, cap-dependent and independent translation initiation, in both sequence-specific and non-specific fashions.

YBX1 binds to nascent pre-mRNA on chromosomes⁹ and acts as an mRNA chaperone in messenger ribonucleoprotein particles (mRNPs)²⁰. It has been shown to be a global cap-dependent mRNA stabilizer that protects mRNA from degradation.⁶ The CSD is required for mRNA recognition³, while the first half of the C-terminal domain has been proposed to interact with the cap structure (or its adjacent region) and prevent Eukaryotic translation initiation factor 4G (eIF4G) from binding. mRNA stabilization through sequence-specific binding has also been reported, such as the CU-rich elements in the 3' untranslated region (UTR) of *renin* mRNA and the c-Jun N-terminal kinases (JNK)-response element in the 5' UTR of *Interleukin 2 (IL-2)* mRNA.²⁰

The ratio of YBX1/mRNA decides whether YBX1 stabilizes mRNA or initiates its activation. When the YBX1/mRNA ratio is high, YBX1 stabilizes mRNA, protecting it from degradation while suppressing translation initiation. When the YBX1/mRNA ratio is low, YBX1 has been proposed to stimulate translation initiation by releasing the cap structure from the C-terminal domain and allowing eIF4G to bind.^{6,9,20} YBX1 is also involved in cap-independent translation regulation for mRNAs such as *Snail1* and the *myc* family proto-oncogenes.²⁰ It has not been shown to play a role in translation elongation or termination. Nevertheless, YBX1 could also influence translation through its interaction with other RNA-binding proteins.³⁰

YBX1 has also been shown to be a spliceosome-associated factor and may play a role in alternative splicing of pre-mRNA. It has been reported to be involved in exon skipping in *MDM2* mRNA, as well as exon inclusion in *CD44* and *NF1* mRNA. YBX1 is proposed to recognize specific sequences in certain mRNAs and potentially recruit splicing factors to them (**Table 1-1**).^{20,32} However, it is not thought to be a core component of the spliceosome.²⁰

1.2.4 YBX1 in DNA repair

YBX1 has been implicated in DNA repair due to its ability to interact with many DNA repair proteins (**Table 1-1**) that are involved in base excision repair,

nucleotide excision repair, mismatch repair, repair of DNA single-stranded and double stranded breaks, as well as recombination repair.²⁰ It has also been reported to have stronger affinity towards secondary structures in damaged DNA, such as duplex DNA containing mispaired bases, abasic sites, or cisplatin modifications.^{39,40,57,58}

1.3 Regulation of YBX1

As a protein with pleiotropic functions, YBX1 is highly regulated on transcriptional, translational, and post-translational level. It shuttles between cytoplasm and nucleus upon various signals. Moreover, it can be secreted to act as an extracellular ligand.

1.3.1 Transcriptional and translational regulation of YBX1

The human *YBX1* promoter does not contain typical RNA polymerase II regulatory sequences, such as a TATA box or the CCAAT element. Its transcriptional regulation is largely dependent on several E-boxes (CATCTG) containing GC-repeats and GATA motifs located at the beginning of the first exon.⁵⁹ Several E-box binding proteins, including MYC, P73⁶⁰, TWIST⁶¹, GATA-1 and GATA-2^{20,62}, have been shown to activate the transcription of *YBX1* in different cell lines. The expression level of some of these transcriptional factors, such as GATA-1 and GATA-2, has been correlated with mRNA level of *YBX1*.²⁰

YBX1 is expressed differently in different tissues. It is consistently found to be abundant in testis^{13,63}, fetal liver^{12,13}, and early precursors of hematopoietic cells: (erythroid, lymphoid, [[Gene Report/BioGPS](#)] and myeloid progenitors⁶⁴). Murine *Ybx1* is expressed at lower levels in kidneys^{13,63}, adult liver, and lungs.¹³ Interestingly, the YBX1 protein abundance in different tissues is not tightly correlated with the mRNA levels^{11,65}, indicating it is regulated at the post-transcriptional level as well.

YBX1 translation can be regulated through both 3'UTR and 5'UTR. There is an evolutionarily conserved AC-rich sequence on the 3'UTR. In a cell-free system, it has been shown that the binding of YBX1 to this sequence will suppress the

translation of *YBX1* mRNA⁶⁶, while the binding of Poly (A)-binding protein (PABP) will activate the translation.⁶⁷ *miR-216a* microRNA may also inhibit translation of murine *Ybx1* mRNA through its 3' UTR.⁶⁸ Further, YBX1 has also been shown to bind to 5'UTR of *YBX1* mRNA and inhibit its translation.⁶⁹

1.3.2 Post-translational regulation of YBX1

YBX1 is post translationally modified by phosphorylation, acetylation, ubiquitination, and limited proteolysis with the 20S proteasome. These modifications are highly correlated with its pleiotropic functions.

YBX1 can be phosphorylated at Ser102 by AKT^{70,71}, which decreases its ability to bind mRNA cap regions,⁷² and minimizes cap-dependent translation⁷¹. It is also phosphorylated by the kinase RSK⁷³. YBX1 can also be phosphorylated by the kinases ERK2 and GSK3 β , which enhance its binding to the *VEGF* promoter.⁷⁴ Other phosphorylation sites revealed by mass spectrometric studies include Ser165 and/or Ser167, Ser174 and/or Ser176, Ser313 and/or Ser314, and Tyr162.²⁰

YBX1 is proposed to be acetylated at Lys301 and Lys304, which may be important for its secretion from cells.⁷⁵

Ubiquitination of YBX1 is mediated by the F-Box protein 33 (FBX33), which leads to its complete degradation by the 26S proteasome.⁷⁶

YBX1 also undergoes ATP- and ubiquitin-independent limited proteolysis by the 20S proteasome, which cleaves *Ybx1* into two fragments. This modification has been reported to occur with genotoxic stresses⁷⁷ and upon treatment of endothelial cells with thrombin.⁷⁸

1.3.3 YBX1 shuttles to the cell nucleus in response to stress signals

YBX1 is predominantly found in the cytoplasm, where it is associated with mRNA, stress granules, and processing bodies. Upon various cellular signals, including stress responses, it can be shuttled to the nucleus where it is most frequently located on chromatin.²⁰

YBX1 normally shuttles from the cytoplasm to the nucleus during the G1/S interphase. This transition is dependent on its NLS sequence and is associated with transcriptional activation of *CCNA2* (*Cyclin A2*) and *CCNB1* (*Cyclin B1*).⁷⁹ YBX1 also moves to the nucleus in response to growth factors and cytokines^{26,80}, through interaction with other proteins like SRSF9³¹ and TP53^{41,81,82}, and in response to various stress signals (including UV-radiation⁸³, DNA-damaging agent treatments⁸⁴, oxidative stress³⁸, and hyperthermia⁸⁵). YBX1 has been shown to interact with KPNB2⁵⁰, which is a member of the transporter family Import-Karyopherin that is responsible for the nuclear import of numerous RNA binding proteins. However, this is not experimentally proven to be the nuclear import mechanism for YBX1.

YBX1 nuclear export has also been reported upon signaling with Platelet-Derived Growth Factor beta (PDGF-B), which also leads to upregulation of *YBX1* transcription. This is observed in kidney cells with mesangioproliferative glomerular disease, in which PDGF-B is a key mediator.⁸⁶ The export of YBX1 has been shown to be independent of CRM1⁸⁷, which is an exportin protein that controls the nuclear export of many proteins, including MDM2/TP53 and CCNB1. The mechanism for the nuclear export of YBX1 remains unidentified.

1.3.4 YBX1 secretion and extracellular functions

YBX1 secretion is mediated via a non-classical mechanism within endolysosomal vesicles, instead of the Golgi apparatus and the endoplasmic reticulum. Secretion presumably requires acetylated Lys301 and Lys304, as the substitution of these residues by alanine completely inhibits secretion of YBX1.⁸⁸ YBX1 is secreted when cells are exposed to lipopolysaccharide, hydrogen peroxide, PDGF-B, or TGF β .⁷⁵ Increased levels of secreted YBX1 have been detected in the sera of sepsis patients⁸⁹, suggesting that the secretion is a consequence of inflammatory stress. Secreted Ybx1 can act as a ligand for the Notch-3 receptor, which is able to activate the transcription of multiple target genes, including members of the HES gene family.⁸⁸

1.4 CSD proteins YBX1 and MSY4 are highly expressed in various cancers

YBX1 is expressed at high levels in almost all types of cancer, including Acute Myeloid Leukemia (AML). It is frequently associated with advanced stages, poor prognosis, and drug resistance, and suggested to be a biomarker for diagnosis, as outlined below. The role of YBX1 in cancer biology has also been explored extensively using cell lines. More recently, MSY4 has been reported to be expressed in various cancer types, and has been suggested to play a role in tumor progression.

1.4.1 CSD protein expression in malignant tumors and cancer cell lines

YBX1 is highly expressed in almost all major types of solid tumors and in leukemia samples. In most reports of clinical specimens, protein levels have been detected using immunohistochemistry, while mRNA abundance has been estimated using microarrays and/or RT-PCR. Both high expression level (mRNA and protein) and nuclear localization have been associated with advanced stages of cancer, including metastatic disease, relapsed disease, drug resistance, and poor outcome (**Table 1-3**). Many groups have suggested that these findings are markers of poor prognosis.

MSY4 has also been detected in primary samples of gastric cancer, hepatocellular carcinoma, and AML. Its expression has not been consistently correlated with poor prognosis. (**Table 1-3**)

Table 1-3. CSD protein expression in clinical specimen of human tumors

CSD protein	Cancer type	Overexpressed species	Prognosis association
YBX1	Breast cancer	Protein ⁹⁰	Nuclear YBX1 associated with intrinsic multidrug resistance
		mRNA ⁹¹	Highly associated with relapse and poor survival
		mRNA ²⁰	Associated with metastasis
		Protein ⁹²	Significantly associated with poor outcome and relapse in patients with or without postoperative chemotherapy

		Protein ²⁰	High expression level associated with estrogen receptor-negative and lymph node positive breast tumors; low expression level associated with lower recurrence risk among patients who did not receive adjuvant chemotherapy
		Protein ⁹³	Correlate with reduced expression of E-cadherin (indication of metastasis) and poor patient survival
		Protein ²⁰	Correlate with large tumors (more than 5 cm) and metastasis of small neoplasms
		Protein ⁹⁴	N/A
		Protein ⁹⁵	Nuclear YBX1 associated with poor survival in early breast cancer
	Ovarian cancer	Protein ⁹⁶	Nuclear YBX1 associated with poor survival
		Protein ⁹⁷	Poor survival associated with co-expression of P-glycoprotein
		Protein ⁹⁸	Nuclear YBX1 associated with poor survival
		mRNA ⁹⁹	Associated with primary tumor and metastatic sites in epithelial ovarian cancer
		Protein ⁹⁶	Nuclear YBX1 associated with recurrent lesions
	Colorectal carcinoma	Protein ¹⁴	N/A
	Gastric cancer	Protein ¹⁰⁰	Not correlated with overall survival, but vascular invasion, liver metastasis, and shortened disease-free survival
	Melanoma	mRNA ¹⁰¹	N/A
	Lung cancer	Protein ¹⁰²	Nuclear YBX1 associated with poor prognosis in non-small cell lung cancer and squamous cell carcinoma
		Protein ¹⁰³	Nuclear YBX1 associated with reduced survival times in non-small cell lung cancer

		Protein ¹⁰⁴	Nuclear YBX1 associated with poor prognosis in non-small cell lung cancer and adenocarcinoma
		Protein ⁵⁴	Associated with T3-4 and Stage II-IV tumors
	Synovial sarcoma	Protein ¹⁰⁵	Nuclear YBX1 associated with poor prognosis
	Osteosarcoma	Protein ¹⁰⁶	Nuclear YBX1 associated with poor overall survival
	Glioblastoma	mRNA (Faury 2007)	N/A
		Protein (Gao 2009)	N/A
		Protein (Fotovati 2011)	YBX1 expression increases with tumor grade
	Neuroblastoma	mRNA (Wachowiak 2010)	Not correlated with poor prognosis
	Prostate cancer	Protein (Gimenez-Bonafe 2004)	N/A
	Diffuse large B-cell lymphoma	Xu 2009	Nuclear YBX1 associated with poor prognosis
	Non-Hodgkin's Lymphoma	Protein (Szczuraszek 2011)	Associated with poor prognosis
	Thyroid anaplastic carcinoma	Protein (Ito 2003)	N/A
	Dialysis-associated renal cell carcinoma	Protein (Fushimi 2013)	N/A
	Bladder cancer	mRNA and protein (Song 2013)	mRNA level correlated with grade and invasiveness; protein level correlated with invasiveness and co-expression level with Twist correlated with poor overall survival rate
	Resectable esophageal squamous cell	Protein (Li 2011)	Associated with high recurrence and lower survival rate

	carcinoma		
	AML FAB type M0-M7	mRNA (Payton 2009, cancer genome research network 2013)	Not correlated with patient outcome
MSY4	AML FAB type M0-M7	mRNA (Payton 2009)	N/A
	hepatocellular carcinoma	mRNA (Yasen 2012)	Not correlated with poor prognosis
		Protein (Yasen 2005)	Nuclear MSY4 associated with poor prognosis
	Gastric cancer	Protein (wang 2009)	N/A

YBX1 protein is also highly expressed in various cancer cell lines, including the human leukemia cell lines K562 and Kasumi, the human lymphoma cell line U937, the human breast cancer cell line MDA (MDA-MB-231), the human melanoma cell line A375, and the human cervical cancer cell line HeLa. It is also highly expressed in the mouse lymphoma cell lines Yac-1, Wehi, and RMAS, the mouse mastocytoma cell line P815, and the mouse skin melanoma cell line B16. MSY4 is expressed in most human cancer cell lines, but has only been detected in mouse P815 cells. In contrast, the CSD family member MSY2 protein is not detected in any of the cell lines noted above.¹¹ (**Figure 1-1**)

CSDE1 is expressed in cancer cell lines such as MCF-7¹⁰⁷. However, its expression is thought to be co-regulated with oncoprotein N-ras^{107,108}, which is often overexpressed in cancer.

1.4.2 CSD protein expression in AML patients and mouse models

AML is a heterogeneous hematopoietic malignancy in which hematopoietic progenitor cells fail to differentiate normally, resulting in the accumulation of immature cells in the bone marrow and blood.¹⁰⁹ The French-American-British (FAB) classification divides AML into eight major subtypes (M0-M7) based on morphology and cytochemical staining patterns.¹¹⁰

In a cohort of 200 AML samples from all eight FAB subtypes, *YBX1* and *MSY4* mRNA expression levels were found to be very high compared to mature neutrophils isolated from healthy bone marrow donors. *MSY2* was not expressed in any AML samples, nor in normal healthy marrow samples. Another CSD family member, *UNR/CSDE1*, is expressed the same in AML patients and mature neutrophils.⁶⁴ (**Figure 1-2A**) No correlation has been found between the expression of either *Ybx1* or *Msy4* and patient outcomes.¹¹¹

Ybx1 and *Msy4* mRNA expression levels have also been measured in a mouse model of Acute Promyelocytic Leukemia (APL), also known as FAB subtype M3. The chromosomal translocation t(15;17)(q21;q22) is found in 98% of APL patients¹¹², which results in the fusion of the promyelocytic leukemia (PML) gene on chromosome 15 with the retinoic acid receptor alpha (RARA) gene on chromosome 17.^{113,114} In *Ctsg-PML-RAR* transgenic mice, a human *PML-RARA* cDNA is knocked into the endogenous murine *cathepsin G* locus, which is maximally expressed in early myeloid progenitor cells, Common Myeloid Progenitors (CMPs) and Granulocyte-Macrophage Progenitors (GMPs).¹¹⁵ About 60% of the *Ctsg-PML-RAR* mice develop a fatal myeloid leukemia that closely resembles human APL.¹¹⁶

Expression profiling has revealed that *Ybx1* mRNA is highly expressed in all the myeloid progenitor compartments, including SLAM, KLS, Promyelocytes, and Neutrophils of both 6 week old *Ctsg-PML-RARA* mice and wildtype mice, and in all 15 tested APL tumors arose in these mice. *Msy4/Csda1* expression is not significantly different in the myeloid progenitors derived from *Ctsg-PML-RARA* vs. wildtype mice, but *Msy4* expression is uniformly very high in mouse APL samples. *Msy2* was not detected in any mouse APL samples or progenitor cell populations, and *Csde1* expression level was not altered in APL tumors compared to healthy myeloid compartments.¹¹⁵ (**Figure 1-2B**)

1.4.3 YBX1 has been implicated in several cancer pathways

Because YBX1 is expressed in many human cancers, and associated with poor outcomes, numerous studies have been carried out to reveal the mechanism by which Ybx1 acts to promote cancer. Most of the data were acquired in cell lines, and many of the studies have limitations due to experimental design.

First, the most significant and frequently reported phenotype of YBX1 is its role in proliferation. Knockdown of YBX1 has been shown to induce apoptosis and/or inhibit cell proliferation in many cancer cell lines, including cell lines derived from patients with melanoma¹¹⁷, liver cancer¹¹⁷, lung cancer^{117–119}, bladder cancer¹²⁰, multiple myeloma^{121,122}, glioblastoma¹²³, breast cancer^{118–120,124,125}, prostate cancer¹²⁰, colon cancer^{117,118}, and leukemia¹²⁶. Several technical issues exist with these published studies including the use of empty vector controls rather than scrambled siRNAs^{121,126}, and single siRNA or shRNAs used without controlling for potential off-target effects^{119,121–123}. Among studies with appropriate controls, only 3 of 5 found a correlation between YBX1 knockdown and more than 50% reduced cell division rates.^{120,125 118,124,127}

Secondly, YBX1 has been implicated in key signaling pathways that are crucial for cancer development, including the E2F pathway, the PI3K/Akt/mTor pathway, MAPK pathways, and the p53 signaling pathway, among others; all of these studies have used cancer cell lines.¹²⁸ Nuclear YBX1 has been shown to be a negative regulator of transcription of *TP53*, leading to a reduced level of TP53 protein.¹¹⁷ It also physically interacts with TP53 and selectively facilitates TP53-induced transactivation of genes such as Matrix metalloproteinase-2 (MMP2), Cyclin-dependent kinase inhibitor 1A (CDKN1A), and MDM2. YBX1 also prevents TP53 from transactivating cell death genes such as BAX, which is a Bcl-like protein that promotes apoptosis.^{21,24,48,81,82,129} In addition, TP53 is required for the nuclear localization of YBX1⁸², which completes a negative feedback loop. Dysregulation of the TP53 signaling by YBX1 could lead not only to limitless replicative potential, self-sufficiency in growth signals, and escape from normal programmed cell death, but also genomic instability; all are all hallmarks of cancer.¹³⁰ Similarly, dysregulation of the PI3K/Akt/mTor pathway by

YBX1 overexpression is associated with dysregulated energy metabolism, since this pathway regulates many components of glycolysis. By drawing conclusions from weak data and overreaching correlations, YBX1 has been proposed to be a master regulator of malignancy, similar to that of Myc and Ras¹²⁸. However, its role in oncogenesis has not yet been convincingly demonstrated in a mouse model of cancer.

Thirdly, YBX1 overexpression is frequently correlated with metastasis in human solid tumors. E-cadherin (CDH1), which is involved in maintaining cell-cell adhesion within tumors, is frequently inactivated in metastatic cancers. YBX1 has been reported to promote the translation of the mRNAs of several proteins that can lead to transcriptional repression of *CDH1*, including SNAIL, LEF1, and TWIST1.^{93,128} Interestingly, TWIST1, which is an E-box binding protein, has been shown to regulate the expression of YBX1 in cancer cell lines.^{120,131–133}

Finally, YBX1 was associated with drug-resistance when it was identified as a protein that is bound to the regulatory region of the *MDR1* gene promoter.¹³⁴ The gene product of *MDR1*, P-glycoprotein, is one of the transporter proteins from the ABC family (ATP-Binding Cassette), and is known to be important for the development of multidrug resistance. Since YBX1 overexpression is frequently associated with high levels of P-glycoprotein in various cancers^{20,135}, it has been hypothesized to activate the transcription of *MDR1*. The evidence for this, however, is controversial. Although the expression levels of YBX1 and *MDR1* have been correlated in some settings, no experiments have yet proven that *MDR1* is a direct transcriptional target of YBX1. Further, YBX1 is not found in the complex bound to the *MDR1* promoter, and its expression level does not correlate with P-glycoprotein or chemo-sensitivity in several cell lines tested.^{20,136}

Interestingly, YBX1 has also been found to be a tumor suppressor that can prevent transformation of chicken embryo fibroblasts induced by PI3K or Akt, but not other oncoproteins such as Src, Jun, or Qin. This phenotype disappears when the RNA binding motif within the CSD is mutated.¹³⁷ However, the mechanism by which YBX1 could act as a tumor suppressor remains unclear.

In conclusion, YBX1 is highly expressed in various cancers and is frequently associated with poor outcomes and chemotherapy resistance. Several lines of evidence suggest a role in cell cycle kinetics, tumor suppression-function, and MDR regulation. However, the mechanism by which it influences cancer development is yet clear. A mouse model is needed to better define the role of YBX1 in the development of cancer.

1.4.4 MSY4 may also play a role in cancer

MSY4 overexpression has been shown to promote tumor cell growth and metastasis in squamous cell carcinoma cell lines, Chronic Myeloid Leukemia (CML) cell lines, and human gastric cancer cell lines^{138–140}. Conversely, it has been shown to inhibit angiogenesis and lymphangiogenesis in aortic and lymphatic endothelial cells.¹⁴¹ Its exact role in cancer remains unclear.

1.5 Murine cold shock domain protein Ybx1 and Msy4 are required for late-stage embryogenesis

There are four mouse CSD genes. Ybx1 is expressed throughout embryogenesis and in virtually all adult tissues¹¹(**Figure 1-3A-B**). Msy4 is expressed at high levels in mid-stage embryos; its expression declines in late-stage embryos(**Figure 1-3A**), and it is expressed only in testis(**Figure 1-3B**) and CD34⁺ hematopoietic progenitor cells in adults^{11,64} (**Figure 1-2**). Ybx2 is only expressed in germ cells and the testis^{11,142}. (**Figure 1-3A-B**) *Csde1* mRNA is expressed in all tissues and developmentally regulated in testis.¹⁰⁸ (**Figure 1-3C-D**)

Thus far, only constitutive knockout animals have been reported for the four CSD genes. Ybx2 expression is limited to the testis in adult animals, and the knockout has a testis-specific phenotype.¹⁴³ Although *Csde1*^{-/-} mice are embryonic lethal¹⁴⁴, *Csde1* is structurally and functionally distinct from the other three family members as discussed in **1.1.5**, and it is not developmentally regulated in hematopoietic cells as discussed in **1.4.2**. Due to these reasons, only Ybx1 and Msy4 knockout animals will be discussed in detail in this section

1.5.1 Ybx1 deficient mice

Two knockout Ybx1 models were generated by independent groups in 2005. Lu *et al.*¹⁴⁵ created a targeted deletion of exon 3 of the *Ybx1* gene, which encodes a portion of the Ybx1 CSD. This targeting strategy creates a deletion of exon 3 and a frame shift that leads to termination of translation within exon 4, completely disrupting the cold shock domain. *Ybx1*^{-/-} embryos developed normally up to embryonic day 13.5 (E13.5). These embryos exhibited severe growth retardation starting from E13.5, and died between E18.5 and gestation. However, Ybx1 deficiency did not result in “global” changes in the transcriptome, proteome, or rates of protein synthesis in Murine Embryonic Fibroblasts (MEFs) derived from *Ybx1*^{-/-} embryos.¹⁴⁵ In addition, it did not cause alterations in gene expression patterns in E13.5 embryos (Li and Ley, unpublished). *Ybx1*^{-/-} MEFs exhibit an elevated sensitivity to oxidative (20% O₂ *in vitro* culture), genotoxic (mitomycin C and cisplatin), and oncogene-induced stress (c-Myc overexpression) compared to WT and *Ybx1*^{+/-} MEFs. Importantly, sensitivity to oxidative stress was found to be due to premature senescence, not increased apoptosis. Re-expression of exogenous Ybx1 rescued the normal growth phenotype, suggesting that it was caused directly by the lack of Ybx1, and not a nearby gene.¹⁴⁵ *Ybx1*^{-/-} MEFs have been reported to produce significantly more viral particles when infected by Dengue Virus, and lead to higher expression level of Dengue Virus proteins, suggesting that Ybx1 has an antiviral effect.¹⁴⁶

Ybx1^{+/-} mice are phenotypically indistinguishable from their wildtype (WT) littermates in the resting state.¹⁴⁵ However, when challenged with lipopolysaccharide (LPS), *Ybx1*^{+/-} mice were protected from LPS-associated mortality (20% vs. 80% in *Ybx1*^{+/+} controls). Immunosuppression of *Ybx1*^{+/-} mice resulted in 50% mortality (0% in *Ybx1*^{+/+} controls). These data suggest that Ybx1 may be an important mediator of bacterial and sterile inflammation.⁸⁹ The targeted mutation in the mouse model used in these studies was generated in 129/SvJ ES cells, and after germline transmission, was backcrossed to C57BL/6 mice for more than ten generations. The phenotype that Lu *et al.* reported from

129/SvJ genetic background remained consistent after the ten generations of backcrossing to B6 mice (Li and Ley, unpublished data).

In a separate study, Uchiumi *et al.*¹⁴⁷ created a targeted mutation involving exon 5 and 6 of the *Ybx1* gene. This targeting strategy creates a deletion of exon 5 and 6 and a frame shift that leads to termination of translation within exon 7, leaving an intact cold shock domain in the truncated mutant protein. The Uchiumi *Ybx1*^{-/-} embryos developed normally up to E11.5. These embryos exhibited growth retardation, hemorrhage, and severe anemia starting from E11.5, and died between E18.5 and gestation, similar to that of the Lu *Ybx1*^{-/-} mice. Neural tube closure was impaired in the Uchiumi *Ybx1*^{-/-} embryos examined from E10.5-E13.5. Translational elongation factor-1 (EF-1) was found to be overexpressed in the Uchiumi *Ybx1*^{-/-} embryos on E11.5. Meanwhile, the Uchiumi *Ybx1*^{-/-} MEFs exhibited decreased growth, which was rescued by re-expression of *Ybx1*. Spontaneous transformation activity was reduced in immortalized Uchiumi *Ybx1*^{-/-} MEF lines, which was rescued by introduction of recombinant *Ybx1*.¹⁴⁷ The Uchiumi *Ybx1*^{-/-} embryos had reduced neural stem cell markers (including Sox-2, nestin, and musashi-1) that are normally expressed in the subventricular zone of fetal brain.¹⁴⁸ The growth of Uchiumi *Ybx1*^{+/-} MEFs and ES cells were similar to that of wildtype cells.^{147,149} However, the Uchiumi *Ybx1*^{+/-} ES cells showed increased sensitivity to cisplatin and mitomycin C, but not to etoposide, X-irradiation, or UV irradiation.¹⁴⁹

In summary, the two *Ybx1* knockout models showed the same embryonic lethality. The growth phenotype in normal and stress conditions for both *Ybx1*^{+/-} and *Ybx1*^{-/-} MEFs were different between the two models, which could potentially be explained by differences between the targeted mutations generated in the two models.

1.5.2 *Msy4* knockout mice

Msy4 knockout mice were generated by deleting exons 2 to 5, including a part of the *Msy4* CSD, resulting in a subsequent frame-shift.¹¹ *Msy4*^{+/-} mice are viable,

fertile, and phenotypically indistinguishable from their wildtype littermates. *Msy4*^{-/-} mice are viable, but the testes of *Msy4*^{-/-} males displayed excessive spermatocyte apoptosis and seminiferous tubule degeneration. *Msy4*^{-/-} males were less fertile than their *Msy4*^{+/-} and wildtype littermates¹¹, and become infertile at 3 to 6 months of age (unpublished data, Lu and Ley). Interestingly, the otherwise phenotypically normal *Msy2*^{-/-} mice exhibited infertility in both male and females¹⁴³, suggesting that *Msy4* and *Msy2* have non-overlapping functions.

Mice deficient for both *Ybx1* and *Msy4* died between E8.5 and E11.5. The expression pattern of *Msy4* during embryogenesis suggests that *Msy4* can compensate for *Ybx1* deficiency during early embryogenesis. However, *Msy4* expression declines starting from E17.5, and the absence of both *Ybx1* and *Msy4* is associated with the onset of severe runting and death.¹¹

In conclusion, a *Ybx1* conditional knockout model is needed to study the roles of the *Ybx1* and *Msy4* proteins in cancer development in adult animals.

1.6 Summary

Cold shock domain (CSD) proteins are the most evolutionarily conserved family of nucleic-acid binding proteins. Studies from the knockouts of *Ybx1* revealed that it plays a crucial role in cellular stress responses, and the prevention of senescence in rapidly dividing cells. *YBX1* overexpression is associated with many types of cancer in humans, and has been implicated in metastasis, drug resistance, and poor outcomes. Unfortunately, most papers that have sought to clarify the role of *YBX1* in cancer have provided incomplete datasets that have been inconclusive. Most investigations to date have also ignored the fact that *MSY4* can clearly complement *YBX1* functions in cells where they are both expressed. Lu *et al.* first described the finding that *Msy4* can compensate for *Ybx1* deficiency: *Ybx1* knockout embryos display severe runting starting from E13.5, and die on E18.5, while mice deficient for both *Ybx1* and *Msy4* die much earlier at E8.5.¹¹ In 2009, our studies showed for the first time that *Msy4* was not only expressed in the testes of adult animals, but also in hematopoietic

progenitor cells of the myeloid lineage.⁶⁴ Therefore, to better understand the role of YBX1 in normal development and cancer, a conditional knockout mouse model of Ybx1 is absolutely required. Further, to fully understand the roles of CSD proteins in hematopoietic cells, it is predicted that both YBX1 and MSY4 would need to be knocked out in adult cells, since they evidently have redundant functions.

References

1. Ermolenko, D. N. & Makhatadze, G. I. Bacterial cold-shock proteins. *Cell. Mol. Life Sci.* **59**, 1902–1913 (2002).
2. Horn, G. & Hofweber, R. Structure and function of bacterial cold shock proteins. *Cell. Mol. Life Sci.* **64**, 1457–70 (2007).
3. Mihailovich, M., Militti, C., Gabaldón, T. & Gebauer, F. Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. *Bioessays* **32**, 109–18 (2010).
4. Yamanaka, K., Fang, L. & Inouye, M. MicroReview The CspA family in *Escherichia coli* : multiple gene duplication for stress adaptation. *Mol. Microbiol.* **27**, 247–255 (1998).
5. Jiang, W., Hou, Y. & Inouye, M. CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. *J. Biol. Chem.* **272**, 196–202 (1997).
6. Kohno K, Izumi H, Uchiumi T, Ashizuka M, K. M. The pleiotropic functions of the Y-box-binding protein, YB-1. *Bioessays* **25**, 691–698 (2003).
7. Giaquinto L, Curmi PM, Siddiqui KS, Poljak A, DeLong E, DasSarma S, C. R. Structure and function of cold shock proteins in archaea. *J. Bacteriol.* **189**, 5738–48 (2007).
8. Sasaki, K. & Imai, R. Pleiotropic roles of cold shock domain proteins in plants. *Front. Plant Sci.* **2**, 116 (2011).
9. Chaikam, V. & Karlson, D. T. Comparison of structure , function and regulation of plant cold shock domain proteins to bacterial and animal cold shock domain proteins. *BMB Rep.* **43**, 1–8 (2009).
10. Matsumoto K, Tanaka KJ, T. M. An acidic protein, YBAP1, mediates the release of YB-1 from mRNA and relieves the translational repression activity of YB-1. *Mol. Cell. Biol.* **25**, 1779–1792 (2005).
11. Lu, Z. H., Books, J. T. & Ley, T. J. Cold shock domain family members YB-1 and MSY4 share essential functions during murine embryogenesis. *Mol. Cell. Biol.* **26**, 8410–7 (2006).
12. Grant, C. E. & Deeley, R. G. Cloning and Characterization of Chicken YB-1 : Regulation of Expression in the Liver. *Mol. Cell. Biol.* **13**, (1993).

13. Ito, K., Tsutsumil, K., Kuzumaki, T., Gomez, P. F. & Otsu, K. A novel growth-inducible gene that encodes a conserved cold-shock domain a protein with. *Nucleic Acids Reserch* **22**, 2036–2041 (1994).
14. Shibao K, Takano H, Nakayama Y, Okazaki K, Nagata N, Izumi H, Uchiumi T, Kuwano M, Kohno K, I. H. Enhanced coexpression of YB-1 and DNA topoisomerase II alpha genes in human colorectal carcinomas. *Int J Cancer*. **83(6)**, 732–737 (1999).
15. Sabath DE, Podolin PL, Comber PG, P. M. cDNA Cloning and Characterization of Interleukin 2-induced Genes in a Cloned T Helper Lymphocyte. *J. Biol. Chem.* **265**, 12671–12678 (1990).
16. Stenina, O. I., Poptic, E. J. & DiCorleto, P. E. Thrombin activates a Y box-binding protein (DNA-binding protein B) in endothelial cells. *J. Clin. Invest.* **106**, 579–87 (2000).
17. Boussadia, O. *et al.* Unr Is Required In Vivo for Efficient Initiation of Translation from the Internal Ribosome Entry Sites of both Rhinovirus and Poliovirus Unr Is Required In Vivo for Efficient Initiation of Translation from the Internal Ribosome Entry Sites of both Rhinovi. *J. Virol.* **77**, 3353–3359 (2003).
18. Chang, T.-C. *et al.* UNR, a new partner of poly(A)-binding protein, plays a key role in translationally coupled mRNA turnover mediated by the c-fos major coding-region determinant. *Genes Dev.* **18**, 2010–23 (2004).
19. Horwitz, E., Maloney, K. & Ley, T. A human protein containing a“ cold shock” domain binds specifically to H-DNA upstream from the human gamma-globin genes. *J. Biol. Chem.* **269**, 14130–14139 (1994).
20. Eliseeva, I. & Kim, E. Y-box-binding protein 1 (YB-1) and its functions. *Biochemistry* **76**, 1402–1433 (2011).
21. Mertens, P. R. *et al.* Combinatorial interactions of p53, activating protein-2, and YB-1 with a single enhancer element regulate gelatinase A expression in neoplastic cells. *J. Biol. Chem.* **277**, 24875–82 (2002).
22. Chernukhin, I. V *et al.* Physical and functional interaction between two pluripotent proteins, the Y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF. *J. Biol. Chem.* **275**, 29915–21 (2000).
23. Narayan, V. *et al.* A multiprotein binding interface in an intrinsically disordered region of the tumor suppressor protein interferon regulatory factor-1. *J. Biol. Chem.* **286**, 14291–303 (2011).

24. Okamoto, T. *et al.* Direct interaction of p53 with the Y-box binding protein, YB-1: a mechanism for regulation of human gene expression. *Oncogene* **19**, 6194–202 (2000).
25. Raj, G. V, Safak, M., MacDonald, G. H. & Khalili, K. Transcriptional regulation of human polyomavirus JC: evidence for a functional interaction between RelA (p65) and the Y-box-binding protein, YB-1. *J. Virol.* **70**, 5944–53 (1996).
26. Higashi, K. *et al.* Interferon-gamma interferes with transforming growth factor-beta signaling through direct interaction of YB-1 with Smad3. *J. Biol. Chem.* **278**, 43470–9 (2003).
27. Ohba, H. *et al.* Sox21 is a repressor of neuronal differentiation and is antagonized by YB-1. *Neurosci. Lett.* **358**, 157–60 (2004).
28. Li, W. W. *et al.* Suppression of grp78 core promoter element-mediated stress induction by the dbpA and dbpB (YB-1) cold shock domain proteins. *Mol. Cell. Biol.* **17**, 61–8 (1997).
29. Shnyreva, M. Interaction of Two Multifunctional Proteins. HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN K AND Y-BOX-BINDING PROTEIN. *J. Biol. Chem.* **275**, 15498–15503 (2000).
30. Ashizuka, M. & Fukuda, T. Novel translational control through an iron-responsive element by interaction of multifunctional protein YB-1 and IRP2. *Mol. Cell. Biol.* **22**, 6375–6383 (2002).
31. Raffetseder, U. *et al.* Splicing factor SRp30c interaction with Y-box protein-1 confers nuclear YB-1 shuttling and alternative splice site selection. *J. Biol. Chem.* **278**, 18241–8 (2003).
32. Wei, W.-J. *et al.* YB-1 binds to CAUC motifs and stimulates exon inclusion by enhancing the recruitment of U2AF to weak polypyrimidine tracts. *Nucleic Acids Res.* **40**, 8622–36 (2012).
33. Safak, M., Gallia, G. L., Ansari, S. a & Khalili, K. Physical and functional interaction between the Y-box binding protein YB-1 and human polyomavirus JC virus large T antigen. *J. Virol.* **73**, 10146–57 (1999).
34. Ansari, S. a *et al.* Interaction of YB-1 with human immunodeficiency virus type 1 Tat and TAR RNA modulates viral promoter activity. *J. Gen. Virol.* **80** (Pt 10), 2629–38 (1999).

35. Chattopadhyay, R. *et al.* Regulatory role of human AP-endonuclease (APE1/Ref-1) in YB-1-mediated activation of the multidrug resistance gene MDR1. *Mol. Cell. Biol.* **28**, 7066–80 (2008).
36. Sengupta S, Mantha AK, Mitra S, B. K. Human AP endonuclease (APE1/Ref-1) and its acetylation regulate YB-1-p300 recruitment and RNA polymerase II loading in the drug-induced activation of multidrug resistance gene MDR1. *Oncogene* **30**, 482–493 (2011).
37. Marenstein, D. R. *et al.* Stimulation of human endonuclease III by Y box-binding protein 1 (DNA-binding protein B). Interaction between a base excision repair enzyme and a transcription factor. *J. Biol. Chem.* **276**, 21242–9 (2001).
38. Das, S. *et al.* Stimulation of NEIL2-mediated oxidized base excision repair via YB-1 interaction during oxidative stress. *J. Biol. Chem.* **282**, 28474–84 (2007).
39. Gaudreault, I., Guay, D. & Lebel, M. YB-1 promotes strand separation in vitro of duplex DNA containing either mispaired bases or cisplatin modifications, exhibits endonucleolytic activities and binds several DNA repair proteins. *Nucleic Acids Res.* **32**, 316–27 (2004).
40. Ise, T., Nagatani, G., Imamura, T. & Kato, K. Transcription Factor Y-Box Binding Protein 1 Binds Preferentially to Cisplatin-modified DNA and Interacts with Proliferating Cell. *Cancer Res.* 342–346 (1999). at <<http://cancerres.aacrjournals.org/content/59/2/342.short>>
41. Guay, D., Gaudreault, I., Massip, L. & Lebel, M. Formation of a nuclear complex containing the p53 tumor suppressor, YB-1, and the Werner syndrome gene product in cells treated with UV light. *Int. J. Biochem. Cell Biol.* **38**, 1300–13 (2006).
42. Ruzanov, P. V, Evdokimova, V. M., Korneeva, N. L., Hershey, J. W. & Ovchinnikov, L. P. Interaction of the universal mRNA-binding protein, p50, with actin: a possible link between mRNA and microfilaments. *J. Cell Sci.* **112 (Pt 2)**, 3487–96 (1999).
43. Chernov, K. G. *et al.* YB-1 promotes microtubule assembly in vitro through interaction with tubulin and microtubules. *BMC Biochem.* **9**, 23 (2008).
44. Chansky, H. A., Hu, M. & Hickstein, D. D. Oncogenic TLS / ERG and EWS / Fli-1 Fusion Proteins Inhibit RNA Splicing Mediated by YB-1 Protein Mediated by YB-1 Protein 1. *Cancer Res.* 3586–3590 (2001).

45. Kojic, S. *et al.* The Ankrd2 protein, a link between the sarcomere and the nucleus in skeletal muscle. *J. Mol. Biol.* **339**, 313–25 (2004).
46. Zou, Y. *et al.* CARP, a cardiac ankyrin repeat protein, is downstream in the Nkx2-5 homeobox gene pathway. *Development* **124**, 793–804 (1997).
47. Khandelwal, P., Padala, M. K., Cox, J. & Guntaka, R. V. The N-terminal domain of y-box binding protein-1 induces cell cycle arrest in g2/m phase by binding to cyclin d1. *Int. J. Cell Biol.* **2009**, 243532 (2009).
48. Tian, B. *et al.* p53 Suppresses lung resistance-related protein expression through Y-box binding protein 1 in the MCF-7 breast tumor cell line. *J. Cell. Physiol.* **226**, 3433–41 (2011).
49. Ohashi, S., Atsumi, M. & Kobayashi, S. HSP60 interacts with YB-1 and affects its polysome association and subcellular localization. *Biochem. Biophys. Res. Commun.* **385**, 545–50 (2009).
50. Lee, B. J. *et al.* Rules for Nuclear Localization Sequence Recognition by Karyopherin β 2. *Cell* **126**, 543–558 (2006).
51. Safak, M., Gallia, G. L. & Khalili, K. Reciprocal interaction between two cellular proteins, Puralpha and YB-1, modulates transcriptional activity of JCVCY in glial cells. *Mol. Cell. Biol.* **19**, 2712–23 (1999).
52. Kelm, R. J. The Single-stranded DNA-binding Proteins, Puralpha , Purbeta , and MSY1 Specifically Interact with an Exon 3-derived Mouse Vascular Smooth Muscle alpha -Actin Messenger RNA Sequence. *J. Biol. Chem.* **274**, 38268–38275 (1999).
53. Guryanov, S. G. *et al.* Formation of amyloid-like fibrils by Y-box binding protein 1 (YB-1) is mediated by its cold shock domain and modulated by disordered terminal domains. *PLoS One* **7**, e36969 (2012).
54. Matsumoto, K. & Bay, B.-H. Significance of the Y-box proteins in human cancers. *J. Mol. Genet. Med.* **1**, 11–7 (2005).
55. Didier, D. K., Schiffenbauer, J., Woulfe, S. L., Zacheis, M. & Schwartz, B. D. Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 7322–6 (1988).
56. Diamond, P., Shannon, M. F., Vadas, M. a & Coles, L. S. Cold shock domain factors activate the granulocyte-macrophage colony-stimulating factor promoter in stimulated Jurkat T cells. *J. Biol. Chem.* **276**, 7943–51 (2001).

57. Lenz, J., Okenquist, S. a, LoSardo, J. E., Hamilton, K. K. & Doetsch, P. W. Identification of a mammalian nuclear factor and human cDNA-encoded proteins that recognize DNA containing apurinic sites. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 3396–400 (1990).
58. Hasegawa, S. L. *et al.* DNA binding properties of YB-1 and dbpA: binding to double-stranded, single-stranded, and abasic site containing DNAs. *Nucleic Acids Res.* **19**, 4915–20 (1991).
59. Makino, Y. *et al.* Structural and functional analysis of the human Y-box binding protein (YB-1) gene promoter. *Nucleic Acids Res.* **24**, 1873–8 (1996).
60. Uramoto, H. *et al.* p73 Interacts with c-Myc to regulate Y-box-binding protein-1 expression. *J. Biol. Chem.* **277**, 31694–702 (2002).
61. Shiota, M. *et al.* Twist promotes tumor cell growth through YB-1 expression. *Cancer Res.* **68**, 98–105 (2008).
62. Yokoyama, H. *et al.* Regulation of YB-1 gene expression by GATA transcription factors. *Biochem. Biophys. Res. Commun.* **303**, 140–145 (2003).
63. Mastrangelo, M. a & Kleene, K. C. Developmental expression of Y-box protein 1 mRNA and alternatively spliced Y-box protein 3 mRNAs in spermatogenic cells in mice. *Mol. Hum. Reprod.* **6**, 779–88 (2000).
64. Payton JE, Grieselhuber NR, Chang LW, Murakami M, Geiss GK, Link DC, Nagarajan R, Watson MA, L. T. High throughput digital quantification of mRNA abundance in primary human acute myeloid leukemia samples. *J. Clin. Invest.* **119**, (2009).
65. Miwa, A., Higuchi, T. & Kobayashi, S. Expression and polysome association of YB-1 in various tissues at different stages in the lifespan of mice. *Biochim. Biophys. Acta* **1760**, 1675–81 (2006).
66. Skabkina, O. V, Lyabin, D. N., Skabkin, M. A. & Ovchinnikov, L. P. YB-1 Autoregulates Translation of Its Own mRNA at or prior to the Step of 40S Ribosomal Subunit Joining. *Mol. Cell. Biol.* **25**, 3317–3323 (2005).
67. Skabkina, O. V *et al.* Poly(A)-binding protein positively affects YB-1 mRNA translation through specific interaction with YB-1 mRNA. *J. Biol. Chem.* **278**, 18191–8 (2003).

68. Kato, M. *et al.* Post-transcriptional up-regulation of Tsc-22 by Ybx1, a target of miR-216a, mediates TGF- β -induced collagen expression in kidney cells. *J. Biol. Chem.* **285**, 34004–15 (2010).
69. Fukuda, T. *et al.* Characterization of the 5'-untranslated region of YB-1 mRNA and autoregulation of translation by YB-1 protein. *Nucleic Acids Res.* **32**, 611–22 (2004).
70. Sutherland, B. W. *et al.* Akt phosphorylates the Y-box binding protein 1 at Ser102 located in the cold shock domain and affects the anchorage-independent growth of breast cancer cells. *Oncogene* **24**, 4281–92 (2005).
71. Evdokimova, V., Ovchinnikov, L. & Sorensen, P. Y-box binding protein 1: providing a new angle on translational regulation. *Cell Cycle* 1143–1147 (2006). at <<http://www.landesbioscience.com/journals/6/article/2784/>>
72. Skabkin, M. a, Evdokimova, V., Thomas, a a & Ovchinnikov, L. P. The major messenger ribonucleoprotein particle protein p50 (YB-1) promotes nucleic acid strand annealing. *J. Biol. Chem.* **276**, 44841–7 (2001).
73. Stratford, A. L. *et al.* Y-box binding protein-1 serine 102 is a downstream target of p90 ribosomal S6 kinase in basal-like breast cancer cells. *Breast cancer Res.* **10**, R99 (2008).
74. Coles, L. S. *et al.* Phosphorylation of cold shock domain/Y-box proteins by ERK2 and GSK3 β and repression of the human VEGF promoter. *FEBS Lett.* **579**, 5372–8 (2005).
75. Frye, B. C. *et al.* Y-box protein-1 is actively secreted through a non-classical pathway and acts as an extracellular mitogen. *EMBO Rep.* **10**, 783–9 (2009).
76. Lutz, M., Wempe, F., Bahr, I., Zopf, D. & von Melchner, H. Proteasomal degradation of the multifunctional regulator YB-1 is mediated by an F-Box protein induced during programmed cell death. *FEBS Lett.* **580**, 3921–30 (2006).
77. Sorokin, A. V *et al.* Proteasome-mediated cleavage of the Y-box-binding protein 1 is linked to DNA-damage stress response. *EMBO J.* **24**, 3602–12 (2005).
78. Stenina, O. I., Shaneyfelt, K. M. & DiCorleto, P. E. Thrombin induces the release of the Y-box protein dbpB from mRNA: a mechanism of transcriptional activation. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 7277–82 (2001).

79. Jurchott, K. *et al.* YB-1 as a cell cycle-regulated transcription factor facilitating cyclin A and cyclin B1 gene expression. *J. Biol. Chem.* **278**, 27988–96 (2003).
80. Basaki, Y. *et al.* Akt-dependent nuclear localization of Y-box-binding protein 1 in acquisition of malignant characteristics by human ovarian cancer cells. *Oncogene* **26**, 2736–46 (2007).
81. Zhang, Y. F. *et al.* Nuclear localization of Y-box factor YB1 requires wild-type p53. *Oncogene* **22**, 2782–94 (2003).
82. Homer, C. *et al.* Y-box factor YB1 controls p53 apoptotic function. *Oncogene* **24**, 8314–25 (2005).
83. Koike, K. *et al.* Nuclear translocation of the Y-box binding protein by ultraviolet irradiation. *FEBS Lett.* **417**, 390–394 (1997).
84. Fujita, T. *et al.* Increased nuclear localization of transcription factor Y-box binding protein 1 accompanied by up-regulation of P-glycoprotein in breast cancer pretreated with paclitaxel. *Clin. cancer Res.* **11**, 8837–44 (2005).
85. Stein, U. *et al.* Hyperthermia-induced nuclear translocation of transcription factor YB-1 leads to enhanced expression of multidrug resistance-related ABC transporters. *J. Biol. Chem.* **276**, 28562–9 (2001).
86. Van Roeyen, C. R. C. *et al.* Y-box protein 1 mediates PDGF-B effects in mesangioproliferative glomerular disease. *J. Am. Soc. Nephrol.* **16**, 2985–96 (2005).
87. Bader, A. & Vogt, P. Inhibition of protein synthesis by Y box-binding protein 1 blocks oncogenic cell transformation. *Mol. Cell. Biol.* **25**, 2095–2106 (2005).
88. Rauen T, Raffetseder U, Frye BC, Djurdjaj S, Mühlenberg PJ, Eitner F, Lendahl U, Bernhagen J, Dooley S, M. P. YB-1 acts as a ligand for Notch-3 receptors and modulates receptor activation. *J. Biol. Chem.* **284**, 26928–26940 (2009).
89. Hanssen, L. *et al.* YB-1 Is an Early and Central Mediator of Bacterial and Sterile Inflammation In Vivo. *J. Immunol.* **191**, 2604–13 (2013).
90. Bargou RC, Jürchott K, Wagener C, Bergmann S, Metzner S, Bommert K, Mapara MY, Winzer KJ, Dietel M, Dörken B, R. H. Nuclear localization and increased levels of transcription factor YB-1 in primary human breast cancers are associated with intrinsic MDR1 gene expression. *Nat. Med.* **3**, 447–450 (1997).

91. Habibi, G. *et al.* Redefining prognostic factors for breast cancer: YB-1 is a stronger predictor of relapse and disease-specific survival than estrogen receptor or HER-2 across all tumor subtypes. *Breast cancer Res.* **10**, R86 (2008).
92. Janz M, Harbeck N, Dettmar P, Berger U, Schmidt A, Jürchott K, Schmitt M, R. H. Y-BOX FACTOR YB-1 PREDICTS DRUG RESISTANCE AND PATIENT OUTCOME IN BREAST CANCER INDEPENDENT OF CLINICALLY RELEVANT tumor biologic factors her2, upa and pai-1. *Int. J. cancer J. cancer* **282**, 278–282 (2002).
93. Evdokimova, V. *et al.* Translational activation of snail1 and other developmentally regulated transcription factors by YB-1 promotes an epithelial-mesenchymal transition. *Cancer Cell* **15**, 402–15 (2009).
94. Rubinstein DB, Stortchevoi A, Boosalis M, Ashfaq R, G. T. Overexpression of DNA-binding Protein B Gene Product in Breast Cancer as Detected by in Vitro -generated Combinatorial Human Immunoglobulin Libraries. *Cancer Res.* **62**, 4985–4991 (2002).
95. Maciejczyk, A. *et al.* Elevated nuclear YB1 expression is associated with poor survival of patients with early breast cancer. *Anticancer Res.* **32**, 3177–84 (2012).
96. Kamura, T. *et al.* Is nuclear expression of Y box-binding protein-1 a new prognostic factor in ovarian serous adenocarcinoma? *Cancer* **85**, 2450–4 (1999).
97. Huang, X. *et al.* Co-expression of Y box-binding protein-1 and P-glycoprotein as a prognostic marker for survival in epithelial ovarian cancer. *Gynecol. Oncol.* **93**, 287–91 (2004).
98. Oda, Y. *et al.* Prognostic implications of the nuclear localization of Y-box-binding protein-1 and CXCR4 expression in ovarian cancer: their correlation with activated Akt, LRP/MVP and P-glycoprotein expression. *Cancer Sci.* **98**, 1020–6 (2007).
99. Iborra, S. *et al.* Alterations in expression pattern of splicing factors in epithelial ovarian cancer and its clinical impact. *Int. J. Gynecol. Cancer* **23**, 990–6 (2013).
100. Wu, Y. *et al.* Strong YB-1 expression is associated with liver metastasis progression and predicts shorter disease-free survival in advanced gastric cancer. *J. Surg. Oncol.* **105**, 724–30 (2012).

101. Hipfel, R., Schitteck, B., Bodingbauer, Y. & Garbe, C. Specifically regulated genes in malignant melanoma tissues identified by subtractive hybridization. *Br. J. Cancer* **82**, 1149–57 (2000).
102. Shibahara K, Sugio K, Osaki T, Uchiumi T, Maehara Y, Kohno K, Yasumoto K, Sugimachi K, K. M. Nuclear Expression of the Y-Box Binding Protein , YB-1 , as a Novel Marker of Disease Progression in Non-Small Cell Lung Cancer. *Clin. cancer Res.* **7**, 3151–3155 (2001).
103. Gessner, C. *et al.* Nuclear YB-1 expression as a negative prognostic marker in nonsmall cell lung cancer. *Eur. Respir. J.* **23**, 14–19 (2004).
104. Kashihara M, Azuma K, Kawahara A, Basaki Y, Hattori S, Yanagawa T, Terazaki Y, Takamori S, Shirouzu K, Aizawa H, Nakano K, Kage M, Kuwano M, O. M. Nuclear Y-Box Binding Protein-1 , a Predictive Marker of Prognosis , Is Correlated with Expression of HER2 / ErbB2 and HER3/ErbB3 in Non-small Cell Lung Cancer. *J. Thorac. Oncol.* **4**, 1066–1074 (2009).
105. Oda, Y. *et al.* Nuclear expression of Y-box-binding protein-1 correlates with P-glycoprotein and topoisomerase II alpha expression, and with poor prognosis in synovial sarcoma. *J. Pathol.* **199**, 251–8 (2003).
106. Fujiwara-Okada, Y. *et al.* Y-box binding protein-1 regulates cell proliferation and is associated with clinical outcomes of osteosarcoma. *Br. J. Cancer* **108**, 836–47 (2013).
107. Fang, H., Yue, X., Li, X. & Taylor, J.-S. Identification and characterization of high affinity antisense PNAs for the human unr (upstream of N-ras) mRNA which is uniquely overexpressed in MCF-7 breast cancer cells. *Nucleic Acids Res.* **33**, 6700–11 (2005).
108. Jeffers, M., Paciucci, R. & Pellicer, a. Characterization of unr; a gene closely linked to N-ras. *Nucleic Acids Res.* **18**, 4891–9 (1990).
109. Estey, E. & Döhner, H. Acute myeloid leukaemia. *Lancet* **368**, 1894–907 (2006).
110. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, S. C. Proposed Revised Criteria for the Classification of Acute Myeloid Leukemia. *Ann. Intern. Med.* **103**, 620–625 (1985).
111. Network, T. C. G. A. R. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–74 (2013).

112. Douer, D. The epidemiology of acute promyelocytic leukaemia. *Best Pract. Res. Clin. Haematol.* **16**, 357–367 (2003).
113. De Thé H, Chomienne C, Lanotte M, Degos L, D. A. The t(15;17) translocation of acute promyelocytic leukaemia fuses the retinoic acid receptor alpha gene to a novel transcribed locus. *Nature* **347**, 588–561 (1990).
114. De Thé, H. *et al.* The PML-RAR alpha fusion mRNA generated by the t(15;17) translocation in acute promyelocytic leukemia encodes a functionally altered RAR. *Cell* **66**, 675–84 (1991).
115. Wartman, L. D. *et al.* Expression and function of PML-RARA in the hematopoietic progenitor cells of Ctsg-PML-RARA mice. *PLoS One* **7**, e46529 (2012).
116. Westervelt, P. *et al.* High-penetrance mouse model of acute promyelocytic leukemia with very low levels of PML-RARalpha expression. *Blood* **102**, 1857–65 (2003).
117. Lasham, a. The Y-box-binding Protein, YB1, Is a Potential Negative Regulator of the p53 Tumor Suppressor. *J. Biol. Chem.* **278**, 35516–35523 (2003).
118. Lasham, A. *et al.* YB-1, the E2F pathway, and regulation of tumor cell growth. *J. Natl. Cancer Inst.* **104**, 133–46 (2012).
119. Basaki, Y. *et al.* Y-box binding protein-1 (YB-1) promotes cell cycle progression through CDC6-dependent pathway in human cancer cells. *Eur. J. Cancer* **46**, 954–65 (2010).
120. Shiota, M. *et al.* Twist1 and Y-box-binding protein-1 promote malignant potential in bladder cancer cells. *BJU Int.* **108**, E142–9 (2011).
121. Chatterjee, M. *et al.* The Y-box binding protein YB-1 is associated with progressive disease and mediates survival and drug resistance in multiple myeloma. *Blood* **111**, 3714–22 (2008).
122. Cobbold, L. C. *et al.* Upregulated c-myc expression in multiple myeloma by internal ribosome entry results from increased interactions with and expression of PTB-1 and YB-1. *Oncogene* **29**, 2884–91 (2010).
123. Gao, Y. *et al.* Inhibition of Y-box binding protein-1 slows the growth of glioblastoma multiforme and sensitizes to temozolomide independent O6-methylguanine-DNA methyltransferase. *Mol. Cancer Ther.* **8**, 3276–84 (2009).

124. Lee, C. *et al.* Targeting YB-1 in HER-2 overexpressing breast cancer cells induces apoptosis via the mTOR/STAT3 pathway and suppresses tumor growth in mice. *Cancer Res.* **68**, 8661–6 (2008).
125. Wu, J. *et al.* Disruption of the Y-box binding protein-1 results in suppression of the epidermal growth factor receptor and HER-2. *Cancer Res.* **66**, 4872–9 (2006).
126. Bhullar, J. & Sollars, V. E. YBX1 expression and function in early hematopoiesis and leukemic cells. *Immunogenetics* **63**, 337–50 (2011).
127. Dhillon, J. *et al.* The expression of activated Y-box binding protein-1 serine 102 mediates trastuzumab resistance in breast cancer cells by increasing CD44+ cells. *Oncogene* **29**, 6294–300 (2010).
128. Lasham, A., Print, C. G., Woolley, A. G., Dunn, S. E. & Braithwaite, A. W. YB-1: oncoprotein, prognostic marker and therapeutic target? *Biochem. J.* **449**, 11–23 (2013).
129. Kim, K. *et al.* Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription. *J. Biol. Chem.* **283**, 9113–26 (2008).
130. Hanahan, D. & Weinberg, R. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
131. Shiota, M. *et al.* Twist and p53 reciprocally regulate target genes via direct interaction. *Oncogene* **27**, 5543–53 (2008).
132. Shiota, M. *et al.* Programmed cell death protein 4 down-regulates Y-box binding protein-1 expression via a direct interaction with Twist1 to suppress cancer cell growth. *Cancer Res.* **69**, 3148–56 (2009).
133. Shiota, M. *et al.* Foxo3a suppression of urothelial cancer invasiveness through Twist1, Y-box-binding protein 1, and E-cadherin regulation. *Clin. Cancer Res.* **16**, 5654–63 (2010).
134. Ohga T, Koike K, Ono M, Makino Y, Itagaki Y, Tanimoto M, Kuwano M, K. K. Role of the Human Y Box-binding Protein YB-1 in Cellular Sensitivity to the DNA-damaging Agents Cisplatin , Mitomycin C , and Ultraviolet Light. *Cancer Res.* **56**, 4224–4228 (1996).
135. Inoue, I., Matsumoto, K., Yu, Y. & Bay, B.-H. Surmounting chemoresistance by targeting the Y-box binding protein-1. *Anat. Rec.* **295**, 215–22 (2012).

136. Dolfini, D. & Mantovani, R. Targeting the Y/CCAAT box in cancer: YB-1 (YBX1) or NF-Y? *Cell Death Differ.* **20**, 676–85 (2013).
137. Bader, A. G., Felts, K. a, Jiang, N., Chang, H. W. & Vogt, P. K. Y box-binding protein 1 induces resistance to oncogenic transformation by the phosphatidylinositol 3-kinase pathway. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12384–9 (2003).
138. Matsumoto, G. *et al.* Cold shock domain protein A (CSDA) overexpression inhibits tumor growth and lymph node metastasis in a mouse model of squamous cell carcinoma. *Clin. Exp. Metastasis* **27**, 539–47 (2010).
139. Sears, D. *et al.* Functional phosphoproteomic analysis reveals cold-shock domain protein A to be a Bcr-Abl effector-regulating proliferation and transformation in chronic myeloid leukemia. *Cell Death Dis.* **1**, e93 (2010).
140. Wang, G. *et al.* Upregulation of human DNA binding protein A (dbpA) in gastric cancer cells. *Acta Pharmacol. Sin.* **30**, 1436–42 (2009).
141. Saito, Y. *et al.* Cold shock domain protein A represses angiogenesis and lymphangiogenesis via inhibition of serum response element. *Oncogene* **27**, 1821–33 (2008).
142. Gu, W. *et al.* Mammalian male and female germ cells express a germ cell-specific Y-Box protein, MSY2. *Biol. Reprod.* **59**, 1266–74 (1998).
143. Yang, J. *et al.* Absence of the DNA-/RNA-binding protein MSY2 results in male and female infertility. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5755–60 (2005).
144. Boussadia, O. *et al.* Transcription of unr (upstream of N-ras) down-modulates N-ras expression in vivo. *FEBS Lett.* **420**, 20–4 (1997).
145. Lu, Z. H., Books, J. T. & Ley, T. J. YB-1 Is Important for Late-Stage Embryonic Development , Optimal Cellular Stress Responses , and the Prevention of Premature Senescence. *Mol. Cell. Biol.* **25**, 4625–4637 (2005).
146. Paranjape, S. M. & Harris, E. Y box-binding protein-1 binds to the dengue virus 3'-untranslated region and mediates antiviral effects. *J. Biol. Chem.* **282**, 30497–508 (2007).
147. Uchiumi, T. *et al.* YB-1 is important for an early stage embryonic development: neural tube formation and cell proliferation. *J. Biol. Chem.* **281**, 40440–9 (2006).

148. Fotovati, A. *et al.* YB-1 bridges neural stem cells and brain tumor-initiating cells via its roles in differentiation and cell growth. *Cancer Res.* **71**, 5569–78 (2011).
149. Shibahara, K. *et al.* Targeted disruption of one allele of the Y-box binding protein-1 (YB-1) gene in mouse embryonic stem cells and increased sensitivity to cisplatin and mitomycin C. *Cancer Sci.* **95**, 348–53 (2004).

Figure legends.

Figure 1-1. Expression of CSD proteins in human and mouse cell lines¹¹

Western Blotting of various human and mouse cancer or immortal cell lines, using rabbit Msy4/CSDA, Yb-1, and Msy2 antibody against human/mouse antigen. Actin and Tubulin antibodies are used as loading controls. Testis sample is used as a positive control for all three CSD proteins. Figure is from Lu *et al*, reference 11.

Figure 1-2. Expression of CSD genes in mouse and human flow-sorted bone marrow cells and leukemia samples

- A. Expression data for four CSD gene family members from 8 FAB subtypes (M0-M7, and also not classified [nc]) of human AML samples, as well as flow-sorted CD34⁺ cells (CD34), promyelocytes (Pros), and neutrophils (PMNs) from healthy donors, all analyzed with the Affymetrix U133 Plus 2 platform.^{64,111}
- B. Expression data for four CSD gene family members in indicated wildtype and *Ctsg-PML-RARA* flow-sorted bone marrow cells, and 15 *Ctsg-PML-RARA* splenic leukemia samples (Mouse APL), all using Nugen amplified mRNA and Affymetrix Mouse Exon 1.0ST arrays.¹¹⁵ The KLS (Kit⁺Lin⁻Sca⁺) population is highly enriched for hematopoietic stem cells (HSPCs), and the SLAM (Kit⁺Lin⁻Sca⁺CD150⁺CD41⁻CD48⁻) compartment is comprised of nearly all HSPCs.

Figure 1-3. Developmental stage-specific and tissue-specific expression patterns of CSD proteins in wildtype mice^{11,108}

- A. Western blotting of lysates from whole-embryos (E9.5-E17.5) and one day old neonates (P1). Figure is from Lu *et al*, Reference 11.
- B. Western blotting of whole-tissue lysates from organs of 2-month-old mice. Figure is from Lu *et al*, Reference 11.
- C. Northern blotting of total RNA of various tissues from 7-week old mice against a *unr/csde* probe consisted of the entire *unr/csde* cDNA. Figure is from Jeffers *et al*, Reference 108.
- D. Northern blotting of total RNA isolated from the testis of 13 day old immature (I) and 49 day old mature (M) mice using a *unr/csde* probe consisted of the entire *unr/csde* cDNA. Figure is from Jeffers *et al*, Reference 108.

Figure 1-1. Expression of CSD proteins in human and mouse cell lines

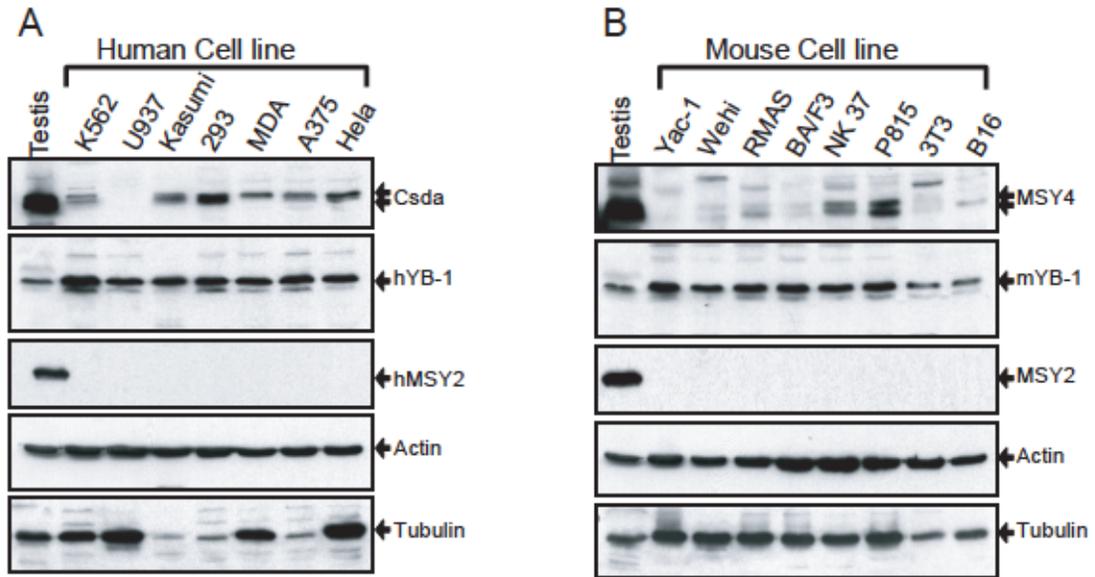


Figure 1-2. Expression of CSD genes in mouse and human flow-sorted bone marrow cells and leukemia samples

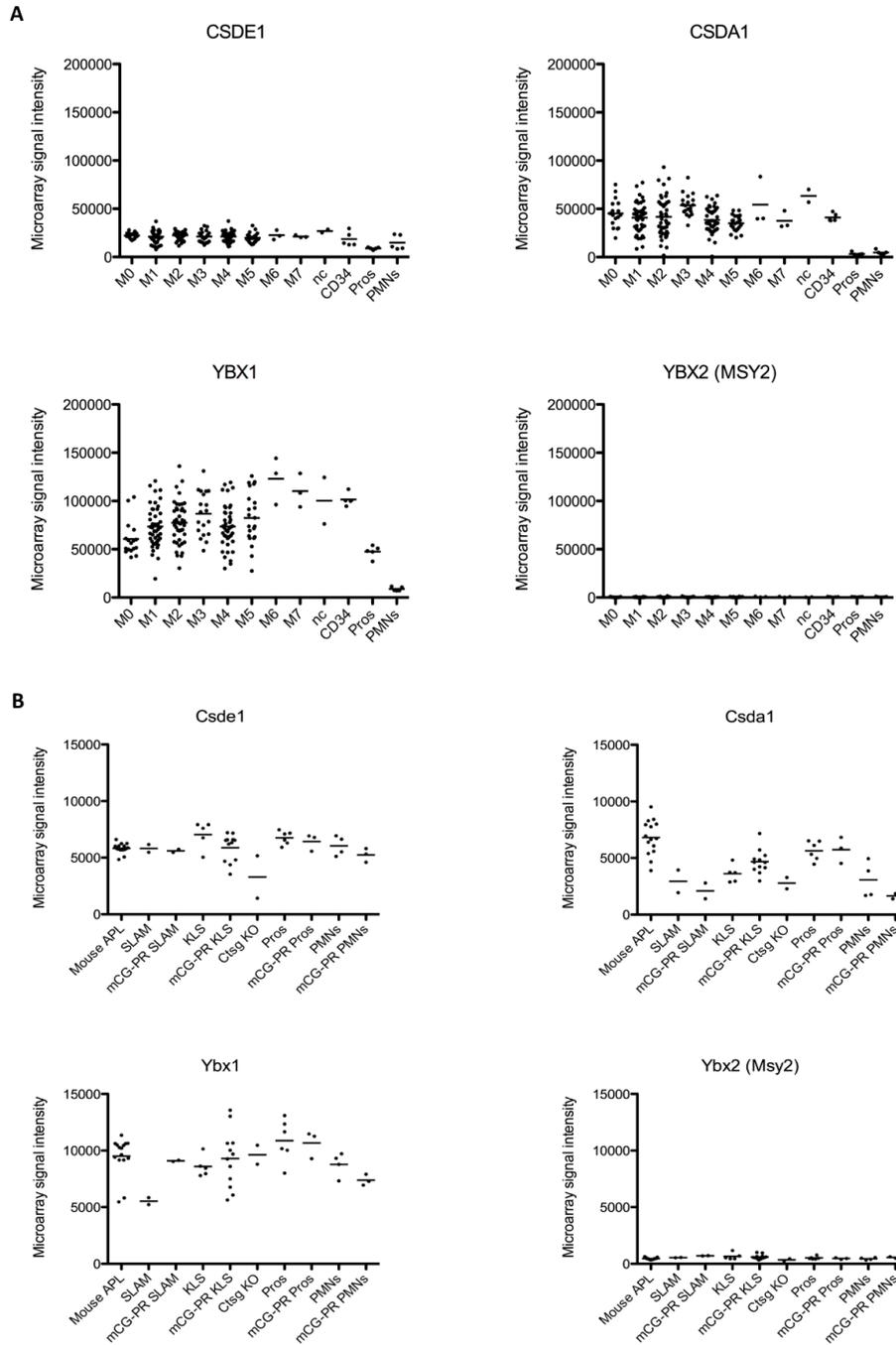
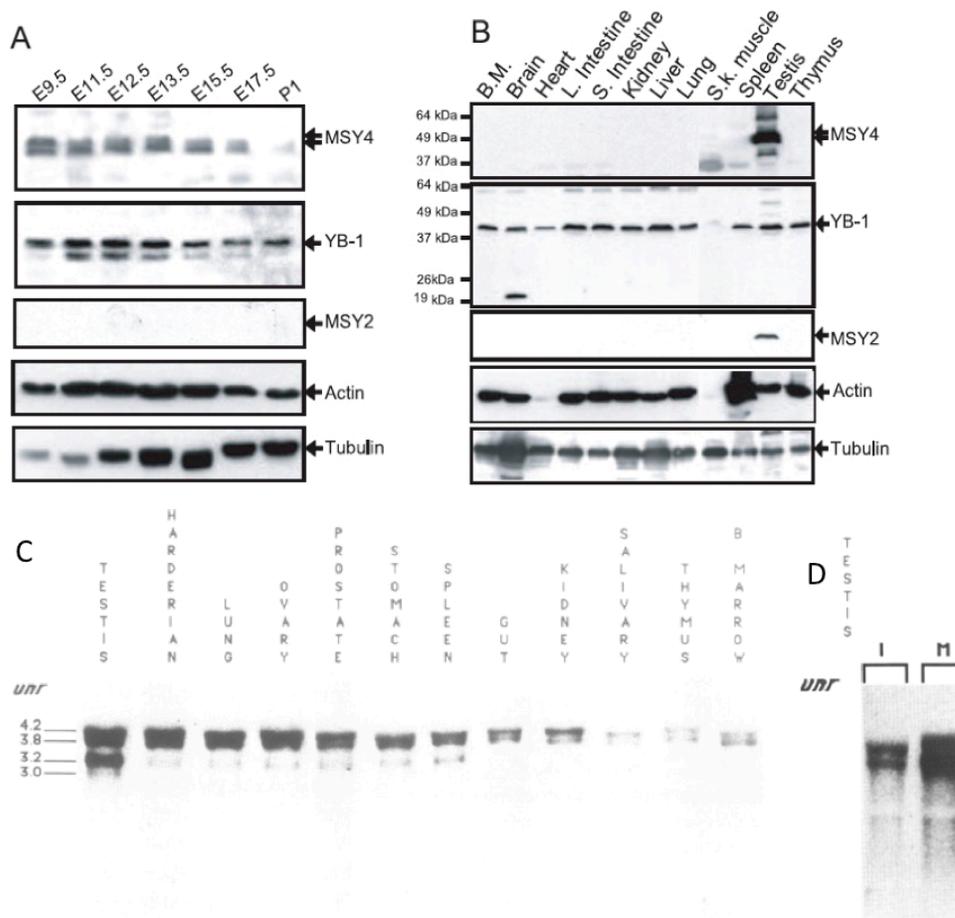


Figure 1-3. Developmental stage-specific and tissue-specific expression patterns of CSD proteins in wildtype mice^{11,108}



Chapter 2

Cold Shock Domain Proteins in Normal and Leukemic Hematopoiesis

ABSTRACT

Genes containing cold shock domains (CSD) encode the most evolutionarily conserved family of nucleic-acid binding proteins. There are four family members: *YBX1*, *MSY4*, *MSY2*, and *CSDE1*. *YBX1* is commonly overexpressed in many cancer types, and overexpression is frequently associated with poor outcomes and chemotherapy resistance. Studies in knockout mice have shown that *Ybx1* is important for stress responses and the prevention of cellular senescence; *Msy4* has overlapping, redundant functions in embryonic development. In virtually all Acute Myeloid Leukemia (AML) samples, *YBX1* and *MSY4* are highly expressed; they are also highly expressed in the myeloid progenitors of healthy donors. Here we report the hematopoietic phenotypes of mice deficient for *Ybx1*, *Msy4*, or both. The loss of *Msy4* does not alter normal adult hematopoiesis. *Msy4*^{-/-} mice have normal complete blood counts (CBCs), normal hematopoietic stem and progenitor cell (HSPC) numbers, and normal numbers of progenitors/colony-forming units (CFUs); competitive repopulation assays showed that loss of *Msy4* does not alter the long-term repopulating potential of hematopoietic stem cells (HSCs). Expression arrays performed on Lin⁻Kit⁺Sca⁺ (KLS) cells (which are highly enriched in HSPCs) showed that deficiency of *Msy4* does not lead to significant changes in mRNA expression patterns. Loss of *Msy4* did not alter hematopoiesis in *Ctsg-PML-RARA* mice, and did not reduce the ability of MLL-AF9 to induce AML in mice. Similarly, we demonstrated that a complete loss of *Ybx1* does not alter hematopoiesis in *Ybx1*^{-/-} fetal liver cells (used since *Ybx1* deficiency causes perinatal lethality). *Ybx1*^{-/-} E14.5 fetal liver cells contained normal numbers of hematopoietic CFUs, and displayed normal engraftment in lethally irradiated recipients. Mice engrafted with *Ybx1* deficient fetal liver cells have normal CBCs and normal numbers of CFUs in their marrow; their HSCs do not have a defect in competitive repopulation assays. In contrast, a heterozygous, conditional truncating mutation of *Ybx1* (*Ybx1*^{lox/+}) was found to cause a severe defect in competitive repopulation assays upon floxing of the conditional allele, suggesting that it has a dominant negative activity. Finally, we demonstrated that *Ybx1*^{lox/-} x *Msy4*^{-/-} bone marrow cells that

express *MLL-AF9* fail to serially replat *in vitro* upon floxing of the conditional allele. In conclusion, Ybx1 and Msy4 appear to have redundant functions in normal and leukemic hematopoietic cells. Expression of one or the other of these cold shock domain proteins is adequate to prevent senescence in the hematopoietic compartment. Inhibition of both, however, may represent a novel approach for limiting the survival of AML cells.

INTRODUCTION

Cold shock domain (CSD) proteins are the most evolutionarily conserved family of nucleic-acid binding proteins. Among the four CSD protein family members, YBX1 overexpression is associated with many types of cancer in humans, and has been implicated in metastasis, drug resistance, and poor outcomes.^{1,2} To date, most studies that have sought to clarify the role of YBX1 in cancer have generated *in vitro* datasets that have not been validated in mouse models, or have not considered the potential role of MSY4 complementation of YBX1 functions in cells where they are both expressed.

In mammals, *Ybx1* is expressed throughout embryogenesis, and in virtually all adult tissues. *Ybx1*^{-/-} mouse embryos display severe runting starting from embryonic day 13.5 (E13.5), and die on E18.5³. *Msy4* is expressed at high levels in mid-stage embryos, but its expression declines in late-stage embryos, and it was thought to be expressed only in the testis of adults.⁴ *Msy4*^{-/-} mice survive into adulthood, but exhibit progressive male infertility. Importantly, mice deficient for both *Ybx1* and *Msy4* die at E8.5, suggesting that *Msy4* can substitute for *Ybx1* during late embryogenesis in *Msy4*^{+/+} x *Ybx1*^{-/-} mice. When *Msy4* expression declines at E17.5 in *Msy4*^{+/+} x *Ybx1*^{-/-} embryos, the lack of both *Ybx1* and *Msy4* causes runting and death.⁴ Based on these knockout mouse studies, and mechanistic studies in cells derived from these mice, *Ybx1* is now known to play a major role in the stress response signaling pathway, where it helps to suppress senescence³.

YBX1 and *MSY4* are both highly expressed in many human cancer cell lines, such as the human leukemia cell lines K562 and Kasumi, the human breast cancer cell line MDA (MDA-MB-231), the human melanoma cell line A375, and the human cervical cancer cell line HeLa.⁴ They are also highly expressed in virtually all primary Acute Myeloid Leukemia (AML) samples from all eight FAB

subtypes. Both genes *are* minimally expressed in mature myeloid cells (promyelocytes and neutrophils) isolated from healthy bone marrow donors.^{5,6} Surprisingly, *YBX1* and *MSY4* are *both* highly expressed in CD34⁺ myeloid progenitors from healthy bone marrow donors. This was unexpected, since previous studies did not detect expression of *Msy4* in Western blotting studies of whole bone marrow lysates from wildtype mice.⁴ It is therefore possible that *MSY4* can complement *YBX1* functions in hematopoietic progenitors and leukemia cells, since they are both expressed in these cells.

In this study, we investigated the impact of *Ybx1* and *Msy4* deficiency in normal and leukemic hematopoiesis. Lu *et al.* showed that *Msy4*^{-/-} mice are viable, but the testes of *Msy4*^{-/-} males displayed excessive spermatocyte apoptosis and seminiferous tubule degeneration.⁴ Hematopoiesis was not characterized in these animals, since *Msy4* was not detected in lysates of the whole bone marrow or spleen of wildtype mice, and since CBCs were normal in *Msy4*^{-/-} mice. In this report, we show that loss of *Msy4* does not alter normal or leukemic hematopoiesis. Although *Ybx1*^{-/-} embryos die late in embryogenesis, they develop normally up to E14.5, which allowed us to use fetal liver samples to study the effects of *Ybx1* deficiency in hematopoiesis. Our data demonstrated that loss of *Ybx1* does not alter normal hematopoiesis. However, we found that a conditional truncation mutation in *Ybx1* causes a defect in the competitive repopulating activity of HSCs, suggesting that this mutation may act in a dominant negative fashion. Finally, we investigated the impact of *Ybx1* and *Msy4* deficiency on “stress” hematopoiesis induced by expression of *MLL-AF9*; the loss of either protein has no phenotype, but loss of both caused a failure of *MLL-AF9* cells to serially replat *in vitro*. We have therefore demonstrated that *Ybx1* and *Msy4* have redundant functions in normal and leukemic hematopoietic cells, and are required for their ability to grow and survive long term.

MATERIALS AND METHODS

Mice

Knockout mouse models for *Ybx1* and *Msy4*, and mice expressing *PML-RARA* from the murine *Cstg* locus (*Ctsg-PML-RARA*) have previously been reported.^{3,4,7} All three strains have been backcrossed more than 10 generations into the C57BL6/J background. Mice with a floxed allele for *Ybx1* were generously provided by Anna Mandinova at Massachusetts General Hospital. Using homologous recombination mediated targeting, exons 5 and 6 were flanked with LoxP sites in C57BL6/J x SVJ129 embryonic stem cells. This strain was backcrossed for more than five generations into the C57BL6/J background for these studies. B6N.Cg-Tg(Vav1-cre)A2Kio/J mice were obtained from The Jackson Laboratory (Bar Harbor, ME) on a C57BL6/J background. *Vav1-Cre*-mediated recombination has been shown to occur in most hematopoietic cells, endothelial cells, and germ cells.⁸

Transplantation and Competitive Repopulation

The bone marrow cells from adult mice of various phenotypes were harvested, treated with ACK lysis buffer (0.15M NH₄Cl, 10mM KHCO₃, 0.1mM Na₂EDTA [Na₂-ethylenediaminetetraacetic acid]), counted, and injected retroorbitally into wildtype recipient mice 24 hours after receiving 1100 cGy of total body irradiation (Mark1 cesum¹³⁷ irradiator; J.L. Shepard).

Timed matings of *Ybx1*^{+/-} mice were used to generate fetal livers from E14.5 embryos, which were cryopreserved in 10% dimethylsulfoxide. Genomic DNA was prepared from the limbs of each embryo to determine their genotype. 1.0 x 10⁶ viable fetal liver cells were injected retroorbitally into wildtype C57BL6/J recipient mice 24 hours after receiving 1100 cGy of total body irradiation.

Competitive repopulation experiments were performed using retro-orbital injections of 1.5 x 10⁶ bone marrow cells into wildtype B6 Ly5.1 x Ly5.2 recipient

mice (CD45.1 x CD45.2) 24 hours after receiving 1100 cGy of total body irradiation. Two or three independent experiments were performed, and the data were analyzed individually. Peripheral blood was collected every 4-8 weeks, and bone marrow cells were harvested at 3-12 months for evaluation of engraftment using flow cytometry.

All mice were monitored for disease 3 times per week by abdominal palpation and observation. Peripheral blood was obtained in heparinized capillary tubes by retro-orbital phlebotomy after adequate methoxyflurane anesthesia (Vedco, Saint Joseph, MO). The Washington University Animal Studies Committee approved all animal experiments.

Flow cytometry and flow sorting

After lysis of red blood cells in ACK buffer, peripheral blood, bone marrow, or fetal liver cells were treated with anti-mouse CD16/32 (clone93; eBioscience, San Diego, CA) and stained with the indicated combinations of the following antibodies: B220, CD3e, Gr1, Ter119, Kit, Sca, CD34, CD16/32 (FCg), CD11b, FcεRI, CD45.1, and CD45.2 (eBioscience, San Diego, CA); or CD117 MicroBeads, mouse (Miltenyi Biotec, Bergisch Gladbach, Germany).

Analysis was performed using a FACScan (Beckman Coulter, Pasadena, CA), and data were analyzed using FlowJo (TreeStar, Ashland, OR).

Flow-sorting was performed on a Reflection high-speed cell sorter (iCyt, Champaign, IL) or on an autoMACS Pro Separator (Miltenyi Biotec, Bergisch Gladbach, Germany).

Expression array profiling

Total RNA was purified using TRIzol reagent (Life Technologies, Carlsbad, CA) and amplified using the whole transcript WT-Ovation RNA Amplification System

and biotin-labeled (NuGen Technologies, San Carlos, CA). Amplified RNA was then applied to the Mouse Exon 1.0ST array (Affymetrix, Santa Clara, CA) according to standard protocols from the Genome Technology Access Center at Washington University in St. Louis. (<https://gtac.wustl.edu/index.php>) Partek Genomics Suite (Partek, St. Louis, MO) was used for unsupervised hierarchical clustering and two-way ANOVA.

Hematopoietic progenitor assays

After red blood cell lysis in ACK buffer, bone marrow cells from individual mice or fetal liver cells from individual embryos were counted using a hemocytometer and plated (in triplicate) at a density of 10,000 cells per 1.1mL of methylcellulose-based media MethoCult M3534 containing interleukin-3 (IL-3), IL-6, and stem cell factor (SCF), or MethoCult M3334 containing Erythropoietin (Epo) (Stem Cell Technologies, British Columbia, Canada). Colonies with >30 cells were counted on day 7. Total cells were collected from the MethoCult media in Dulbecco modified Eagle medium with 2% fetal bovine serum (FBS), washed, and counted.^{9,10} Cells were replated at same density, or in 1:5 or 1:10 dilution to avoid high colony numbers (>500-1,000 colonies per plate). This process was repeated for 6 weeks, or until serial replating failed.

Retroviral production and transduction

MSCV-MLL-AF9-IRES-GFP¹¹ retroviral supernatants were generated and viral titers were determined by flow cytometry by GFP positivity in transduced 3T3 cells, as described previously.¹² MSCV-IRES-GFP versus MSCV-MLL-AF9-IRES-GFP retroviral constructs were transduced into bone marrow cells derived from 6-8 week old mice. The transduced marrow was then transplanted into lethally irradiated wildtype C57BL6/J mice obtained from The Jackson Laboratory, as previously reported.^{12,13} Briefly, after red blood cell lysis in ACK buffer, mononuclear bone marrow cells were cultured in media containing RPMI,

20% FBS, SCF, FLT3L, IL-3, and TPO (PeproTech, Rocky Hill, NJ 08553) for 24 hours. The cells were then spinfected on days -1 and 0. After a 2-hour rest period, 1×10^6 transduced bone marrow cells were then injected retroorbitally into wildtype C57BL6/J recipient mice 24 hours after receiving 1100 cGy of total body irradiation.

Western blotting

Primary antibodies included anti-Ybx1 (2397-1; Epitomics), anti-Actin (Clone C4 [MAB1501R]; Millipore), and anti-Msy4 generated by Lu *et al.*⁴ Bone marrow and spleen cells were first treated with ACK buffer to lyse red blood cells, and then the remaining nucleated marrow cells (or total organs) were lysed in urea buffer (7M urea, 2M thiourea, 4% chaps, 30mM Trizma) containing 1x Protease Inhibitor (Sigma-Aldrich, St. Louis, MO) and freeze-thawed in liquid nitrogen three times. The whole tissue or cell lysates were measured for their protein concentration using Precision Red Advanced Protein Assay (Cytoskeleton, Inc., Denver, CO) The lysates were electrophoresed in 10% sodium dodecyl sulfate-polyacrylamide (SDS) gels, and the proteins were transferred to a polyvinylidene difluoride membranes (GE Healthcare, Pittsburgh, PA). Western blotting was performed as previously described.^{3,4} The horseradish peroxidase signal was detected by an ECL detection system (Bio-Rad, Hercules, CA).

TAT-cre induction of floxing

After red blood cell lysis in ACK buffer, bone marrow cells from *ROSA-lox-STOP-lox-YFP mice* or mice with the *Ybx1* floxed allele were incubated with 100 units/ml TAT-cre (Excellgen, Rockville, MD) dissolved in media containing RPMI, 20% FBS, SCF, FLT3L, IL-3, and TPO for 2 hours. Cells were harvested by centrifugation, and resuspended in fresh media containing RPMI, 20% FBS, SCF, FLT3L, IL-3, and TPO for transplantation or viral transduction.

Quantitative PCR

The efficiency of *Ybx1* floxing was determined using genomic DNA derived from whole cell preparations. Quantitative PCR was performed using Taqman Gene Expression MasterMix (Life technologies, Carlsbad, CA) and a StepOnePlus Real Time PCR Systems (Life technologies) per the manufacturer's specifications. The Actb Taqman Gene Expression Assay (Primer Limited) containing primers for genomic *Actin b* and probes with VIC-MGB dye were purchased from Life Technologies. The Custom FloxedYbx1 Taqman Gene Expression Assay was designed to contain a forward primer (TAGGATGGTGGCTCTAATTGGA), a reverse primer (CGGATCCAAGCTTATGCATGAA), and a probe (TGACCACATGAACACAG) labeled with FAM-MGB dye, manufactured by Life Technologies. In the multiplexed platform, both the Actb and FloxedYbx1 reagents were added in the same reaction and detected by their distinct fluorescent signals, which allows for an internal control to be present in every sample. 0.5 ng of genomic DNA was used per 20 μ l reaction. Data analysis was performed using StepOnePlus v2.2.2 software and Microsoft Excel.

Statistical analysis

Exon-level summary data were generated using the RMA algorithm in Partek Genomics Suite (Partek, St. Louis, MO). Only core probe sets were used to limit the analysis within well-annotated exons. The ratio of average signal intensity for KLS samples and fetal liver engrafted bone marrow samples was calculated as fold change, and a one-way analysis of variance (ANOVA) was used to define genes with altered patterns of expression.

Statistical comparisons were made using Student's 2-tailed *t* test unless otherwise noted. $P \leq 0.05$ was considered to be significant.

RESULTS

Msy4 and Ybx1 are highly expressed in normal and leukemic hematopoietic cells

We assessed the expression patterns of all four mouse and human CSD genes in normal and leukemic hematopoietic cells. *Ybx1* is expressed throughout embryogenesis and in virtually all adult tissues.⁴ *Msy4* is expressed at high levels in mid-stage embryos, but its expression declines in late-stage embryos, and it was thought to be expressed only in the testis of adults⁴. *Ybx2* is expressed only in the germ cells and the testis of adult mice^{4,14} (**Figure 1-3A-B**). Finally, *Csde1* is expressed in all tissues of adult mice, and is developmentally regulated in testis¹⁵ (**Figure 1-3C-D**).

In a cohort of 200 Acute Myeloid Leukemia (AML) samples that included all FAB subtypes, *YBX1* and *MSY4* mRNA expression levels were found to be very high compared to mature promyelocytes and neutrophils isolated from healthy bone marrow donors. Surprisingly, *YBX1* and *MSY4* are *both* highly expressed in CD34⁺ myeloid progenitors from healthy bone marrow donors. This was unexpected, since previous studies did not detect expression of *Msy4* in Western blotting studies of whole bone marrow lysates from wildtype mice.⁴ *MSY2* expression was not detected in any AML samples, nor in normal healthy marrow samples. *UNR/CSDE1* is expressed at similar levels in AML samples and mature myeloid cells (**Figure 1-2A**). No correlation has been found between the expression levels of either *YBX1* or *MSY4* and patient outcomes.⁵

In mice, *Ybx1* mRNA is highly expressed in all the myeloid progenitor compartments of 6 week old animals, including SLAM (Lin⁻Kit⁺Sca⁺CD150⁺CD41⁻CD48⁻), KLS (Lin⁻Kit⁺Sca⁺), common myeloid progenitors (CMPs), granulocyte-macrophage progenitors (GMPs), megakaryocyte-erythroid progenitors (MEPs)

Promyelocytes, and Neutrophils. Expression levels were similar in *Ctsg-PML-RARA* mice and wildtype mice, and in 15 Acute Promyelocytic Leukemia (APL) tumors arising in the *Ctsg-PML-RARA* mice. *Msy4/Csda1* was likewise expressed in the hematopoietic stem/progenitor cell (HSPC) compartments noted above, and its expression was not significantly different in *Ctsg-PML-RARA* mice. However, *Msy4* expression was uniformly high in mouse APL samples. *Msy2* was not detected in any mouse APL samples, or in HSPC compartments or mature myeloid cells. *Csde1* expression was not different in APL tumors vs. healthy myeloid cells at any stage of differentiation¹⁶ (**Figure 1-2B**).

*Resting hematopoiesis is minimally altered in *Ybx1*^{+/-} and *Msy4*^{-/-} mice*

To determine the importance of *Ybx1* and *Msy4* for normal and leukemic hematopoiesis, we first assessed resting hematopoiesis in *Ybx1*^{+/-} and *Msy4*^{-/-} adult mice in the C57BL6/J background. Peripheral blood from *Ybx1*^{+/-} and *Msy4*^{-/-} mice and their wildtype littermates between 6 to 12 weeks of age (n=6) showed no significant difference in complete blood counts (CBCs), except that platelet counts were slightly but significantly higher in *Msy4*^{-/-} mice (average of 1954 K/ μ L vs. 1729 K/ μ L, $P < 0.05$). Next, we sacrificed *Ybx1*^{+/-}, *Msy4*^{-/-}, and wildtype littermates (n=2) and harvested bone marrow for immunophenotypic analysis of HSPC compartments. HSPC profiling was done by fluorescent staining of cell surface lineage markers that were detected by flow cytometry. Lin⁻, KLS, MEP, GMP, and CMP compartments were unaltered in *Ybx1*^{+/-} and *Msy4*^{-/-} mice compared to their wildtype littermates (**Figure 2-1**).

We also performed colony-forming assays in methylcellulose with different cytokines to quantify functional hematopoietic progenitors.^{9,10} Two different methylcellulose-based media were used: MethoCult M3534 (containing stem cell factor [SCF], interleukin [IL]-3, and IL-6) was used to assess CFU-GMs (colony forming unit-granulocyte, monocyte), which are derived from CMPs; MethoCult

M3334 (containing only erythropoietin) was used for the quantification of CFU-Es (colony forming units-erythroid), which are derived from MEPs. Bone marrow cells from each mouse were plated in triplicate in MethoCult for CFU-GM and CFU-E assessment. Colony numbers were not significantly altered in *Ybx1*^{+/-} or in *Msy4*^{-/-} derived bone marrow cells (**Figure 2-1**). Myeloid and erythroid development is therefore not detectably altered by *Ybx1* haploinsufficiency or *Msy4* deficiency.

Loss of Msy4 does not alter the long-term repopulating potential of HSCs

In a competitive repopulation assay, bone marrow cells from *Msy4*^{-/-} and *Msy4*^{+/+} (CD45.2) mice were mixed 1:1 with bone marrow from wildtype mice with a different surface marker Ly5.1 (CD45.1), and transplanted into lethally irradiated wildtype B6 Ly5.1 x Ly5.2 (CD45.1 x CD45.2) recipients (**Figure 2-2A**). We did not observe significant differences of the percentage of CD45.2⁺ cells in the peripheral blood at any time points ($P > 0.05$) (**Figure 2-2B**).

We analyzed the expression profiles of flow-sorted KLS cells (which are highly enriched in HSPCs) derived from *Msy4*^{-/-} and *Msy4*^{+/+} mice using the Mouse Exon1.0 ST array. The expression level of *Msy4* in wildtype KLS cells confirmed the finding that *Msy4* is expressed in HSPCs, as predicted from human studies^{5,6,16}. *Msy4*^{-/-} KLS cells did not express *Msy4*, as expected (**Figure 2-3A**). Unsupervised hierarchical clustering revealed no significant differences between the global expression patterns of *Msy4*^{-/-} and *Msy4*^{+/+} samples (**Figure 2-3B**). When we performed a supervised analysis (Two-way ANOVA), the only gene that was differentially expressed with statistical significance ($P < 0.05$, fold change >2 or <-2) was *Msy4* itself. Therefore, loss of *Msy4* does not lead to significant changes in mRNA expression patterns in KLS cells.

Hematopoiesis is minimally altered in $Ybx1^{-/-}$ fetal liver-derived cells

$Ybx1^{+/-}$ mice are phenotypically identical to their wildtype littermates³ and have normal basal hematopoiesis (**Figure 2-1A**); however, $Ybx1^{-/-}$ mice uniformly die at E18.5-E19.5. $Ybx1^{-/-}$ E14.5 embryos are phenotypically smaller than their $Ybx1^{+/+}$ littermates, and their total fetal liver cell numbers are significantly lower than that of wildtype littermates (**Figure 2-4A-B**). Fetal liver cells obtained from $Ybx1^{-/-}$ and $Ybx1^{+/+}$ E14.5 embryos were plated on MethoCult M3534 (containing IL-3, IL-6, and SCF) and had equivalent numbers of CFU-GMs ($P > 0.05$) (**Figure 2-4C**).

We injected 1×10^6 $Ybx1^{-/-}$ or $Ybx1^{+/+}$ E14.5 fetal liver cells into lethally irradiated wildtype C57BL6/J recipients. Both $Ybx1^{-/-}$ and $Ybx1^{+/+}$ E14.5 fetal liver cells engrafted in all recipient mice, although white blood cell counts were slightly but significantly lower in $Ybx1^{-/-}$ engrafted mice at 2 months post transplantation ($P < 0.05$). Immunophenotyping of HSPCs and CFUs from MethoCult plating of engrafted whole bone marrow cells from the fetal livers did not show any significant differences (**Figure 2-5A**).

Mouse Exon 1.0ST arrays were performed on the engrafted whole bone marrow cells derived from $Ybx1^{-/-}$ or $Ybx1^{+/+}$ fetal livers. The only gene that was differentially expressed with statistical significance ($P < 0.05$) by Two-way ANOVA was $Ybx1$ itself (**Figure 2-5B**). Unsupervised hierarchical clustering revealed that loss of $Ybx1$ does not lead to significant differences in global mRNA expression patterns (**Figure 2-5C**).

To quantify the long term engrafting potential of $Ybx1$ deficient fetal liver cells, we performed a competitive repopulation study. $Ybx1^{-/-}$ or $Ybx1^{+/+}$ fetal liver engrafted bone marrow cells (CD45.2) were mixed 1:1 with bone marrow from

wildtype B6 Ly 5.1 mice (CD45.1) and transplanted into lethally irradiated wildtype B6 Ly5.1 x Ly5.2 recipients (CD45.1 x CD45.2) (**Figure 2-5D**). Although both showed reduced numbers of engrafting cells over time (probably because the samples were derived from fetal livers and not adult bone marrow), there was no significant difference between the percentages of CD45.2⁺ cells in the peripheral blood of the two groups at any time point ($P > 0.05$) (**Figure 2-5E**).

The effect of a conditional Ybx1 allele on adult hematopoiesis

Ybx1^{lox/+} mice were generated by targeting LoxP sites to flank exons 5 and 6 of the *Ybx1* gene (Mandinova *et al.*, unpublished). This targeting strategy creates a deletion of exon 5 and 6, and a frame shift that is predicted to cause termination of translation within exon 7, leaving a truncated protein with an intact cold shock domain (**Figure 2-6A**). The predicted effect of the floxed allele is similar to the *Ybx1* allele described by Uchiumi *et al.*,¹⁷ which produced homozygous mice that were embryonic lethal; the phenotype of these mice was nearly identical to mice with a deletion of exon 3 that was generated in in our laboratory³.

Surprisingly, when *Ybx1*^{lox/+} mice were intercrossed, adult mice homozygous for the floxed allele were not detected, either in our laboratory, or the Mandinova Lab. However, when *Ybx1*^{lox/+} mice were intercrossed with our laboratory's *Ybx1*^{+/-} mice, *Ybx1*^{lox/-} mice were born at a ratio similar to that of *Ybx1*^{lox/+} mice (43% vs 36.4%) (**Table 2-1**). *Ybx1*^{lox/-} mice live to adulthood, which demonstrates that the floxed allele does not create haploinsufficiency (if it did, these mice should die at E18.5-19.5). Importantly, CBCs obtained from adult *Ybx1*^{lox/+} and *Ybx1*^{lox/-} mice were essentially normal, except that hemoglobin levels were slightly but significantly lower in *Ybx1*^{lox/+} mice (average of 13.18 g/dL vs. 14.03 g/dL, $P < 0.05$), and platelet counts were slightly higher in *Ybx1*^{lox/-} mice (average of 913.5 K/ μ L vs. 729.7 K/ μ L, $P < 0.05$) (**Figure 2-7A**).

Ybx1^{lox/+} mice were crossed with our laboratory's *Ybx1*^{+/-} mice, and also with *Vav1-Cre*^{+/-} mice obtained from The Jackson Laboratory. *Vav1-Cre* is expressed primarily in the hematopoietic cells of mice, with some 'leakage' in endothelial cells, and germ cells¹⁷. We performed Western blotting of tissue extracts from the bone marrow, spleen, intestine, liver, and kidney of *Ybx1*^{lox/+} x *Vav1-Cre*^{+/-}, *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-}, and *Vav1-Cre*^{+/-} littermates (**Figure 2-6B**). *Ybx1* protein levels appear to be slightly reduced in *Ybx1*^{lox/+} x *Vav1-Cre*^{+/-} bone marrow and spleen cells. Levels are further reduced (but not eliminated) in the bone marrow cells of *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-} mice, and minimally reduced in the spleen and intestines. No reduction was detected in the liver or kidneys. Residual *Ybx1* expression in bone marrow cells may be due to incomplete floxing by *Vav1-Cre*, or from non-floxed stromal cells that are also present in whole bone marrow samples. Western blots of sorted hematopoietic cells from these mice are in progress to resolve this issue. Importantly, *Msy4* protein levels were not altered in *Ybx1*^{lox/+} x *Vav1-Cre*^{+/-} or *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-} mice, as expected (data not shown).

CBCs obtained from *Ybx1*^{lox/+} x *Vav1-Cre*^{+/-} mice are normal. The red blood cell counts (average of 8.12 M/ μ L vs. 9.53 M/ μ L, $P < 0.01$) and hemoglobin levels (average of 12.65 g/dL vs. 13.8 g/dL, $P < 0.01$) in *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-} mice were slightly but significantly lower than wildtype C57BL6/J mice (**Figure 2-7A**).

To investigate the importance of *Ybx1* for the long-term repopulating potential of adult hematopoietic stem cells (HSCs), we performed a competitive repopulation by mixing bone marrow cells from *Ybx1*^{lox/+} x *Vav1-Cre*^{+/-}, *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-}, or *Vav1-Cre*^{+/-} littermates (CD45.2) with equal numbers of wildtype B6 Ly5.1 competitors, and injected these cells retroorbitally into lethally irradiated wildtype B6 Ly5.1 x Ly5.2 recipients (CD45.1 x CD45.2). Surprisingly, both *Ybx1*^{lox/+} x *Vav1-Cre*^{+/-} and *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-} recipients showed significantly lower

percentages of CD45.2⁺ cells engrafted at 4 weeks, compared to *Vav1-Cre*^{+/-} control recipients ($P < 0.001$). The percentage of CD45.2⁺ cells in both *Ybx1*^{lox/+} x *Vav1-Cre*^{+/-} and *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-} declined further from week 4 to week 8 ($P < 0.05$), and was stable at week 12. These data suggest that bone marrow cells with the *Ybx1* conditional allele in the *Vav1-cre* background (with or without a null *Ybx1* allele) have a striking defect in both engraftment and long-term repopulating potential. Since *Ybx1*^{+/-} bone marrow cells (with the exon 3 deletion) do not have alterations in their ability to engraft and repopulate long-term (data not shown), and since *Ybx1*^{-/-} fetal liver cells appear to engraft normally (**Figure 2-5E**), the *Ybx1* floxed allele does not appear to be a simple null allele that is equivalent to the exon 3 deletion of *Ybx1*.

Loss of Msy4 or Ybx1 does not affect leukemic hematopoiesis

To test whether the loss of *Msy4* alters leukemic hematopoiesis, we crossed *Msy4*^{-/-} mice with *Ctsg-PML-RARA* mice to obtain *Msy4*^{+/+} x *Ctsg-PML-RARA*^{+/-}, *Msy4*^{+/-} x *Ctsg-PML-RARA*^{+/-}, and *Msy4*^{-/-} x *Ctsg-PML-RARA*^{+/-} animals. In a serial replating experiment in MethoCult M3534 (containing IL-3, IL-6, and SCF), we examined the impact of *Msy4* deficiency on the aberrant replating potential of progenitors from *Ctsg-PML-RARA*^{+/-} bone marrow (**Figure 2-8A**). As expected, wildtype and *Msy4*^{-/-} progenitors did not replate after the second week. Bone marrow cells with the *Ctsg-PML-RARA* allele replated for 7 weeks, and the gene dosage of *Msy4* did not alter the replating potential ($P > 0.05$) (**Figure 2-8B**).

We next transduced bone marrow cells from *Msy4*^{-/-} mice, *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-} mice, and wildtype C57BL6/J mice with either MSCV-IRES-GFP (empty vector) or MSCV-MLL-AF9-IRES-GFP. We injected 1x10⁶ transduced cells into lethally irradiated C57BL6/J mice, and also performed serial replating in MethoCult M3534 (containing IL-3, IL-6, and SCF) (**Figure 2-9A**). Bone marrow cells from both the *Msy4*^{-/-} mice and the *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-} mice did not exhibit an

altered replating ability with MSCV-MLL-AF9-IRES-GFP *in vitro*, nor was leukemia-free survival altered after transplantation into adult mice (**Figure 2-9B-C**).

Hematopoietic phenotypes in bone marrow cells deficient for both Ybx1 and Msy4

To reduce the complexity of breeding experiments, we decided to induce floxing in bone marrow cells *ex vivo* by incubating the cells with purified TAT-Cre protein derived from *E.coli*, which can directly traverse cell membranes, and efficiently induce floxing in living cells^{18,19}. To test the efficiency of this process, we first treated both unfractionated and flow-sorted Kit⁺ bone marrow cells (purified on an autoMACS Pro Separator) from *ROSA-lox-STOP-lox-YFP* mice (generously provided by Fehniger Lab at Washington University in St. Louis). After treating the cells with varying doses of TAT-Cre for varying times *in vitro*, we measured YFP expression (which is induced by floxing). We observed no alterations in cell viability when cells were treated with TAT-cre for 2-4 hours, with >90% viable cells present at 72 hours, which was not different from non-treated controls (data not shown). YFP expression was dependent on the concentration of TAT-cre used, but was not sensitive to the length of treatment, suggesting that the entry of TAT-Cre into cells is relatively rapid (**Table 2-2**). We chose an optimal dose of 100 units/mL TAT-cre for 2 hours as the standard for all subsequent experiments (**Figure 2-10**).

Ybx1^{lox/+} animals were bred into our laboratory's *Ybx1*^{+/-} and *Msy4*^{-/-} mice to produce *Ybx1*^{lox/-} and *Ybx1*^{lox/-} x *Msy4*^{-/-} mice. CBCs obtained from these adult mice are normal, except that white blood cell, neutrophil, lymphocyte, and platelet counts were slightly but significantly higher in *Ybx1*^{lox/-} x *Msy4*^{-/-} mice ($P < 0.05$) (**Figure 2-7A**). We treated bone marrow cells from *Ybx1*^{lox/-} and *Ybx1*^{lox/-} x *Msy4*^{-/-} mice with TAT-cre to induce floxing in the *Ybx1* allele. After 48 hours,

TAT-Cre treated bone marrow cells were spinfected with either MSCV-IRES-GFP (empty vector) or MSCV-MLL-AF9-IRES-GFP, followed by serial replating in MethoCult M3534 (containing IL-3, IL-6, and SCF). The *Ybx1*-floxed population was quantified using a multiplexed qPCR analysis, which uses different fluorescent probes to measure amplification of the floxed *Ybx1* allele. *Actin b* served as an internal control (**Figure 2-11A**). Based on the Western blotting results from *Ybx1^{lox/-} x Vav1-Cre^{+/-}* animals, production of the floxed allele in *Ybx1^{lox/-} x Msy4^{-/-}* cells should cause a near complete loss of expression of *Ybx1* in the *Msy4* deficient background.

TAT-cre treated *Ybx1^{lox/-}* and *Ybx1^{lox/-} x Msy4^{-/-}* cells were transduced with MSCV-MLL-AF9-IRES-GFP to determine whether bone marrow cells experiencing a proliferative stress require the CSD proteins for survival. Hematopoietic cells expressing *MLL-AF9* have a strong serial replating phenotype, caused by increased progenitor self-renewal. The *Ybx1*-floxed population did not change in abundance after serial replating of the samples derived from *Ybx1^{lox/-}* mice, suggesting that the near complete loss of *Ybx1* does not alter the ability for these cells to self-renew and proliferate in this *ex vivo* replating assay. However, the *Ybx1*-floxed population from *Ybx1^{lox/-} x Msy4^{-/-}* marrow slowly declined in abundance after 3 weeks of replating (**Figure 2-11B**). *Ybx1* and *Msy4* are therefore both required to prevent the loss of hematopoietic cells that are rapidly proliferating due to *MLL-AF9* expression.

DISCUSSION

Based on the studies of knockout mice, we have previously suggested that the cold shock domain proteins Ybx1 and Msy4 play a critical role in rapidly proliferating tissues, where they suppress senescence. *Ybx1* deficient mice exhibit late embryonic lethality, while *Ybx1 x Msy4* double knockout embryos die at E8.5. The embryonic expression of *Msy4* compensates for *Ybx1* loss during mid-stage embryogenesis, but because *Msy4* expression declines in late-stage embryogenesis, the mouse dies because it cannot sustain rapid cell growth. Both YBX1 and MSY4 are highly expressed in diverse human cancers, suggesting that high level expression of these proteins may be a critical adaptation that cancer cells require to sustain their long term proliferative potential. In this study, we evaluated the roles of Ybx1 and Msy4, the two major cold shock domain proteins, in normal and leukemic hematopoietic cells. Although Ybx1 had previously been known to be expressed in all adult tissues (including bone marrow cells) it was not previously known that Msy4 is highly expressed in early hematopoietic progenitors². We observed no hematopoietic phenotype in fetal liver or bone marrow cells lacking either protein individually. However, the double knockout revealed the redundancy of these two proteins in hematopoietic cells under proliferative stress provided by a leukemia-initiating oncogene: both proteins were required for MLL-AF9 expressing cells to serially replat.

We first evaluated the consequences of Msy4 and Ybx1 deficiency individually. Since *Msy4*^{-/-} mice are viable, we simply evaluated hematopoiesis in adult animals. These animals have normal peripheral blood counts, and normal numbers of HSPCs and CFUs in their bone marrow. A competitive repopulation assay revealed a normal number of long term repopulating cells in *Msy4* deficient marrow. Since *Msy4* is primarily expressed in early hematopoietic cells, we used expression array profiling to evaluate the consequences of *Msy4* deficiency in KLS cells (which are highly enriched in hematopoietic stem and progenitor cells).

Remarkably, there were no differences in gene expression in *Msy4* deficient KLS cells—except for the loss of *Msy4* itself.

In two independent leukemia models characterized by aberrant self-renewal of HSPCs (*Ctsg-PML-RARA* mice and MLL-AF9 expressing bone marrow cells), we demonstrated that the loss of *Msy4* did not affect the abnormal serial replating ability (i.e. self-renewal) of progenitors. Similarly, *Msy4* deficiency did not alter the ability of MLL-AF9 to cause a rapidly fatal leukemia in mice. Therefore, *Msy4* deficiency did not produce a measureable hematopoietic phenotype, either in basal or stressed systems.

Since *Ybx1* deficiency causes perinatal lethality, we used fetal liver cells from E14.5 *Ybx1*^{-/-} embryos to initially evaluate their hematopoietic potential. Although *Ybx1*^{-/-} fetal livers are smaller than their wildtype counterparts, they contained the same relative proportion of CFUs (in terms of colonies counted per 10,000 fetal liver cells), and exhibited the same engraftment potential when transferred to adult mice. The long term repopulating potential of *Ybx1* deficient fetal liver cells was not different from their wildtype littermates. The engrafted bone marrow cells derived from *Ybx1*^{-/-} E14.5 fetal liver transplants had global mRNA expression profiles that were essentially identical to that of their wildtype counterparts—except for the loss of *Ybx1* itself. Collectively, these results showed that homozygosity for the null allele of *Ybx1*, provided by the deletion of exon 3, does not detectably alter the hematopoietic potential of fetal liver cells derived from these mice.

We obtained mice with a conditional, floxed *Ybx1* allele (*Ybx1*^{lox/+}), to study the loss of *Ybx1* in adult hematopoietic cells. Mice heterozygous for this allele are phenotypically normal and have normal blood counts. However, for unexplained reasons, heterozygous breeding pairs do not produce homozygous (*Ybx1*^{lox/lox})

adult animals. This result suggests that this allele does not represent a simple null mutation. Indeed, since *Ybx1*^{lox/-} mice are viable and born in the expected ratios, it is clear that the two *Ybx1* mutations are not equivalent. The data strongly suggests that the floxed allele has a phenotype that is independent of its effect on *Ybx1* itself.

By breeding *Ybx1*^{lox/+} and *Ybx1*^{lox/-} mice with *Vav1-cre*^{+/-} mice, we showed that the floxing of the *Ybx1* allele caused a drastic reduction of protein levels within the bone marrow compartment. Surprisingly, however, bone marrow cells from *both* *Ybx1*^{lox/+} x *Vav1-cre*^{+/-} and *Ybx1*^{lox/-} x *Vav1-cre*^{+/-} mice demonstrated reduced engraftment and long-term repopulating potentials, again suggesting that the floxed *Ybx1* allele must have a dominant negative phenotype in this specific system (unlike the null allele with the deletion of exon 3).

When exons 5 and 6 of the floxed *Ybx1* allele are removed by Cre-mediated recombination, a truncated protein that retains the cold shock domain is predicted (**Figure 2-6A**). The Uchiumi *Ybx1* mutation targets the same exons (5 and 6), and ES cells heterozygous for this mutation have a phenotype, displaying increased sensitivity to cisplatin and mitomycin C, but not to etoposide, X-irradiation, or UV irradiation²⁰. In contrast, mice and/or embryonic fibroblasts that are heterozygous for the *Ybx1* mutation targeting exon 3 (a true null allele) have no measurable phenotypes³. To determine whether the floxed allele could produce the truncated protein that is predicted, we cloned the predicted open reading frame of the floxed *Ybx1* allele into the pcDNA3.1 vector with a 3XFlag tag at the 3' end of the cDNA. After transient transfection into 293T cells, we were able to detect a truncated *Ybx1* protein of the predicted length with antibodies that detected either the N-terminal domain of *Ybx1*, or the Flag tag (**Figure 2-12**). However, we have not been able to detect this product in lysates made from the bone marrow cells of *Ybx1*^{lox/+} x *Vav1-Cre*^{+/-} or *Ybx1*^{lox/-} x *Vav1-*

Cre^{+/-} mice (data not shown). Further studies of the effects of this truncated Ybx1 protein will be required to clarify the hematopoietic phenotype of the conditional knockout mice.

Since mice deficient for both *Ybx1* and *Msy4* die early in embryogenesis, fetal livers were not available for analysis of hematopoietic phenotypes. We therefore bred the conditional, floxed *Ybx1* allele (*Ybx1*^{lox/+}) to our *Ybx1*^{+/-} mice and *Msy4*^{-/-} mice to produce *Ybx1*^{lox/-} and *Ybx1*^{lox/-} x *Msy4*^{-/-} mice; both were viable and had normal CBCs. To eliminate the need for breeding these triply transgenic mice to various Cre-expressing mice, we applied TAT-Cre protein *ex vivo* to induce floxing. This system also allows us to examine the consequences of *Ybx1* floxing exclusively in hematopoietic cells, eliminating the potentially confounding effects of floxing in other tissues, such as bone marrow stromal cells, etc. The floxing of bone marrow cells from *Ybx1*^{lox/-} x *Msy4*^{-/-} creates a population of cells that have neither protein, which can be quantified and tracked with a qPCR assay specific for the *Ybx1* floxed allele.

We transduced TAT-cre treated bone marrow samples from both *Ybx1*^{lox/-} and *Ybx1*^{lox/-} x *Msy4*^{-/-} mice with the MSCV-MLL-AF9 retrovirus, and serially replated these cells in MethoCult M3534 (containing IL-3, IL-6, and SCF). The *Ybx1*-floxed population remained stable after many rounds of plating, probably because these cells were rescued by *Msy4*. However, the *Ybx1*-floxed population in *Ybx1*^{lox/-} x *Msy4*^{-/-} samples slowly disappeared with serial replating, suggesting that expression of either *Ybx1* or *Msy4* is required to maintain the high rates of self-renewal and proliferation induced by MLL-AF9.

Collectively, these data suggest that highly proliferative cells, such as hematopoietic progenitors and leukemia cells (as well as other primary tumors, and cancer cell lines) may require high levels of *YBX1* because it suppresses

senescence, allowing for the long term expansion of these cells. The expression of MSY4, which clearly has functions similar to and overlapping with YBX1, may provide a “fail-safe” system to back up the essential activity of YBX1. Our data clearly demonstrate that the loss of both proteins leads to a reduced ability for leukemic cells to proliferate *in vitro*, and also suggests that both proteins may be required to maintain normal hematopoiesis. Since both of these proteins are highly expressed in all AML samples (and many other cancers), inhibition of their combined functions (perhaps by inhibiting the shared cold shock domains) may provide a novel approach for inducing senescence and death in transformed cells.

REFERENCES

1. Eliseeva, I. & Kim, E. Y-box-binding protein 1 (YB-1) and its functions. *Biochemistry* **76**, 1402–1433 (2011).
2. Matsumoto, K. & Bay, B.-H. Significance of the Y-box proteins in human cancers. *J. Mol. Genet. Med.* **1**, 11–7 (2005).
3. Lu, Z. H., Books, J. T. & Ley, T. J. YB-1 Is Important for Late-Stage Embryonic Development , Optimal Cellular Stress Responses , and the Prevention of Premature Senescence. *Mol. Cell. Biol.* **25**, 4625–4637 (2005).
4. Lu, Z. H., Books, J. T. & Ley, T. J. Cold shock domain family members YB-1 and MSY4 share essential functions during murine embryogenesis. *Mol. Cell. Biol.* **26**, 8410–7 (2006).
5. Network, T. C. G. A. R. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–74 (2013).
6. Payton JE, Grieselhuber NR, Chang LW, Murakami M, Geiss GK, Link DC, Nagarajan R, Watson MA, L. T. High throughput digital quantification of mRNA abundance in primary human acute myeloid leukemia samples. *J. Clin. Invest.* **119**, (2009).
7. Westervelt, P. *et al.* High-penetrance mouse model of acute promyelocytic leukemia with very low levels of PML-RARalpha expression. *Blood* **102**, 1857–65 (2003).
8. Georgiades, P. *et al.* VavCre transgenic mice: a tool for mutagenesis in hematopoietic and endothelial lineages. *Genesis* **34**, 251–6 (2002).
9. Welch, J. S., Klco, J. M., Varghese, N., Nagarajan, R. & Ley, T. J. Rara haploinsufficiency modestly influences the phenotype of acute promyelocytic leukemia in mice. *Blood* **117**, 2460–8 (2011).
10. Walter, M. J. *et al.* Reduced PU.1 expression causes myeloid progenitor expansion and increased leukemia penetrance in mice expressing PML-RARalpha. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12513–8 (2005).
11. Sykes, S. M. *et al.* AKT/FOXO signaling enforces reversible differentiation blockade in myeloid leukemias. *Cell* **146**, 697–708 (2011).

12. Luo, H. *et al.* c-Myc rapidly induces acute myeloid leukemia in mice without evidence of lymphoma-associated antiapoptotic mutations. *Blood* **106**, 2452–61 (2005).
13. Wartman LD, Larson DE, Xiang Z, Ding L, Chen K, Lin L, Cahan P, Klco JM, Welch JS, Li C, Payton JE, Uy GL, Varghese N, Ries RE, Hock M, Koboldt DC, McLellan MD, Schmidt H, Fulton RS, Abbott RM, Cook L, McGrath SD, Fan X, Dukes AF, Vickery T, Kalicki J, L, L. T. Sequencing a mouse acute promyelocytic leukemia genome reveals genetic events relevant for disease progression. *J. Clin. Invest.* **121**, 1445–1455 (2011).
14. Gu, W. *et al.* Mammalian male and female germ cells express a germ cell-specific Y-Box protein, MSY2. *Biol. Reprod.* **59**, 1266–74 (1998).
15. Jeffers, M., Paciucci, R. & Pellicer, a. Characterization of unr; a gene closely linked to N-ras. *Nucleic Acids Res.* **18**, 4891–9 (1990).
16. Wartman, L. D. *et al.* Expression and function of PML-RARA in the hematopoietic progenitor cells of Ctsg-PML-RARA mice. *PLoS One* **7**, e46529 (2012).
17. Uchiumi, T. *et al.* YB-1 is important for an early stage embryonic development: neural tube formation and cell proliferation. *J. Biol. Chem.* **281**, 40440–9 (2006).
18. Peitz, M., Pfannkuche, K., Rajewsky, K. & Edenhofer, F. Ability of the hydrophobic FGF and basic TAT peptides to promote cellular uptake of recombinant Cre recombinase: a tool for efficient genetic engineering of mammalian genomes. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 4489–94 (2002).
19. Capasso, P., Aliprandi, M., Ossolengo, G., Edenhofer, F. & de Marco, A. Monodispersity of recombinant Cre recombinase correlates with its effectiveness in vivo. *BMC Biotechnol.* **9**, 80 (2009).
20. Shibahara, K. *et al.* Targeted disruption of one allele of the Y-box binding protein-1 (YB-1) gene in mouse embryonic stem cells and increased sensitivity to cisplatin and mitomycin C. *Cancer Sci.* **95**, 348–53 (2004).

FIGURE LEGENDS.

Figure 2-1. Resting hematopoiesis in *Ybx1*^{+/-} and *Msy4*^{-/-} mice

- A. Comparisons of hematopoietic values between *Ybx1* haploinsufficient mice and wildtype littermates. Complete blood counts were performed, including white blood cells (WBC), neutrophils (NE), lymphocytes (LY), monocytes (MO), hemoglobin (Hb), and platelet (PLT) counts. Values for hematopoietic stem and progenitor cell populations were measured, (including Lin⁻, KLS, GMP, CMP, and MEP compartments) by flow cytometry. A comparison of the numbers of CFU-GM and CFU-E present in the bone marrow cells of adult mice was also performed.
- B. Comparisons are shown for peripheral blood CBCs, hematopoietic stem and progenitor cell populations, and numbers of CFU-GM and CFU-E for adult *Msy4*^{-/-} mice and their wildtype littermates.

Figure 2-2. Competitive repopulation of *Msy4*^{-/-} mice

- A. Experimental schema.
- B. The peripheral blood was analyzed at weeks 4, 6, 12, 19, 30, and 44. Donor contributions were analyzed for the percentage of CD45.2⁺ cells (experimental donors) over the total of CD45.1⁺ (competitors) and CD45.2⁺ cells. Values for individual recipients are displayed. No significant differences were found at any sampling time between *Msy4*^{-/-} and *Msy4*^{+/+} donor mice.

Figure 2-3. Expression array data comparing KLS cells from *Msy4*^{-/-} and *Msy4*^{+/+} mice

- A. Expression profiles of KLS cells from *Msy4*^{-/-} mice or their wildtype littermates. Individual data points represent results from each mouse. A Student's 2-tailed *t* test was used to compare values from all the *Msy4*^{-/-} vs. *Msy4*^{+/+} mice.
- B. Unsupervised hierarchical clustering of all samples, which did not organize by genotype.

Figure 2-4. Characterization of *Ybx1*^{-/-} E14.5 fetal livers

- A. *Ybx1*^{-/-} E14.5 embryos are smaller in size compared to *Ybx1*^{+/+} littermates.
- B. Total cell numbers from *Ybx1*^{-/-} E14.5 fetal livers are significantly less than that of *Ybx1*^{+/+} littermates.
- C. Numbers of CFU-GMs from 10,000 E14.5 fetal liver cells plated in MethoCult M3534 (containing SCF, IL-3, and IL-6) are not significantly different between *Ybx1*^{-/-} and *Ybx1*^{+/+} littermates.

Figure 2-5. Hematopoiesis in mice engrafted with *Ybx1*^{-/-} E14.5 fetal liver cells

- A. Comparison of peripheral blood CBCs, hematopoietic stem and progenitor cell populations, and numbers of CFU-GM and CFU-E between mice engrafted with *Ybx1*^{-/-} and *Ybx1*^{+/+} E14.5 fetal liver cells.
- B. Expression profiles of *Ybx1* in whole bone marrow from mice engrafted with *Ybx1*^{-/-} or *Ybx1*^{+/+} E14.5 fetal liver cells. Individual data points represent results from each tested mouse.
- C. Unsupervised hierarchical clustering of all samples, which do not organize by genotypes.
- D. Experimental schema for competitive repopulation.
- E. Peripheral blood from the competitive repopulation recipients were analyzed at week 4, 8, 12, 20, and 44. Donor contributions were analyzed for the percentage of CD45.2⁺ cells (experimental donors) over the total of CD45.1⁺ (competitors) and CD45.2⁺ cells. Values for individual recipients are displayed. No significant differences were found at any sampling time between bone marrow from mice engrafted with *Ybx1*^{-/-} or *Ybx1*^{+/+} E14.5 fetal liver cells by using Student's 2-tailed *t* test.

Figure 2-6. Characterization of *Ybx1*^{lox/+} mice

- A. The mutational strategies and predicted consequences of the *Ybx1* mutations used in this study are shown.
- B. Tissue survey of *Ybx1* protein expression in *Vav1-Cre*^{+/-}, *Ybx1*^{lox/+} x *Vav1-Cre*^{+/-}, and *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-} mice. Western blot analyses of whole tissue lysates from 6-8 week old mice are shown. Top panel, anti C-terminal *Ybx1* antibody. Bottom panel, anti-Actin antibody, as a loading control.

Figure 2-7. Effects of the *Ybx1* floxed allele on adult hematopoiesis

- A. CBCs of *Ybx1*^{lox/+} mice with various genetic backgrounds. Comparison of peripheral blood CBCs for *Ybx1*^{lox/+}, *Ybx1*^{lox/-}, *Ybx1*^{lox/-} x *Msy4*^{-/-}, *Ybx1*^{lox/+} x *Vav1-Cre*^{+/-}, and *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-} mice. N = 2-6 for each genotype. All mice were 6-12 weeks of age, except for one 4-week-old *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-} mouse.
- B. Competitive repopulation of *Ybx1*^{lox/+} x *Vav1-Cre*^{+/-}, *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-}, and *Vav1-Cre*^{+/-} mice. Peripheral blood cells were analyzed every 4 weeks for 12 weeks. Donor contributions were analyzed for the percentage of CD45.2⁺ cells (experimental donors) over the total of CD45.1⁺ (competitors) and CD45.2⁺ cells. Values for individual recipient mice are displayed.

Figure 2-8. Effects of *Msy4* deficiency on serial replating by *Ctsg-PML-RARA* bone marrow cells

- A. Experimental schema.
- B. Colony formation and serial replating. Bone marrow cells from littermate-matched 6-12 week old mice (2-3 mice from each genotype) were plated in triplicate in MethoCult M3534 (containing SCF, IL-3, and IL-6). After 7 days, colonies quantified on each plate. Cells were replated in triplicate and 7 days later, colonies were counted and replated. Replating continued for 6 weeks, or until colony formation failed.

Figure 2-9. Effects of *Msy4* or *Ybx1* deficiency on a retroviral MLL-AF9 leukemia model

- A. Experimental schema.
- B. Tumor watch. The indicated cohorts of mice transplanted with *Msy4*^{-/-}, *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-}, and wildtype bone marrow cells spininfected with either MSCV-IRES-GFP (empty vector) or MSCV-MLL-AF9-IRES-GFP were prospectively established in a tumor watch. Mice were followed for 100 days, and all moribund animals were sacrificed and examined. Fatal leukemia occurred in mice transplanted with bone marrow cells transduced with MSCV-MLL-AF9-IRES-GFP, but not with MSCV-IRES-GFP. There were no significant differences in the leukemia penetrance among mice transplanted with MSCV-MLL-AF9-IRES-GFP transduced *Msy4*^{-/-}, *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-}, or wild type bone marrow cells. (*P* > 0.05)
- C. Colony formation and serial replating of *Msy4*^{-/-}, *Ybx1*^{lox/-} x *Vav1-Cre*^{+/-}, and wildtype bone marrow cells spininfected with either MSCV-IRES-GFP (empty vector) or MSCV-MLL-AF9-IRES-GFP. Transduced cells were plated in in MethoCult M3534 (containing SCF, IL-3, and IL-6). Replating continued for 6 weeks, or until colony formation failed.

Figure 2-10. Optimal induction of floxing by TAT-cre

Flow cytometric analysis of *ROSA-lox-STOP-lox-YFP* bone marrow cells 48 hours after treatment with 0 or 100 units/mL TAT-cre for 2 hours. YFP expression is dependent on floxing. Only TAT-cre treated cells displayed YFP expression.

Figure 2-11. Effects of *Ybx1* and *Msy4* deficiency on aberrant replating of bone marrow cells expressing MLL-AF9

- A. Multiplexed platform measuring *Ybx1*-floxed (FloxedYb1) and *actin b* (*Actb*) signals in the same sample. Standard curves can be generated by $\Delta\Delta C_T$ value of controls with a set percentage of floxed *Ybx1* alleles.
- B. Colony formation and serial replating of *Ybx1*^{lox/-} and *Ybx1*^{lox/-} x *Msy4*^{-/-} bone marrow cells treated with TAT-cre, and then spininfected with either MSCV-IRES-GFP (empty vector) or MSCV-MLL-AF9-IRES-GFP.

- Transduced cells were plated in MethoCult M3534 (containing SCF, IL-3, and IL-6). Replating continued for 4 weeks, or until colony formation has failed. Genomic DNA was harvested from the cells each time they were replated, and evaluated with the qPCR assay for the fate of the floxed population.
- C. Percentage of *Ybx1*-floxed cells (*Ybx1*^{-/-}) present after each weekly replating event.

Figure 2-12. Expression of the truncated Ybx1 protein predicted from the floxing of the *Ybx1* conditional allele

293T cells were transfected with pcDNA3.1-Ybx1Trunc (encoding the putative Ybx1 product resulting from deletion of exon 5 and 6), or pcDNA3.1-Ybx1TruncFlag with 3xFlag sequence at the 3' end of Ybx1Trunc. 48 hours post transfection, total cell lysates were prepared in RIPA buffer containing 1x Protease inhibitor. Lysates were loaded on SDS-PAGE gels and transferred to polyvinylidene difluoride membranes. Western blotting was performed using an antibody against the N-terminal domain of Ybx1 (Ab12148, Abcam), and an anti-Flag antibody (Clone M2, Sigma-Aldrich). Both detected the truncated Ybx1 mutant. An anti-actin antibody was used for loading control. An *Ybx1*^{-/-} MEF cell extract was used to define the Ybx1-specific protein band.

Figure 2-1. Resting hematopoiesis in *Ybx1*^{+/-} and *Msy4*^{-/-} mice

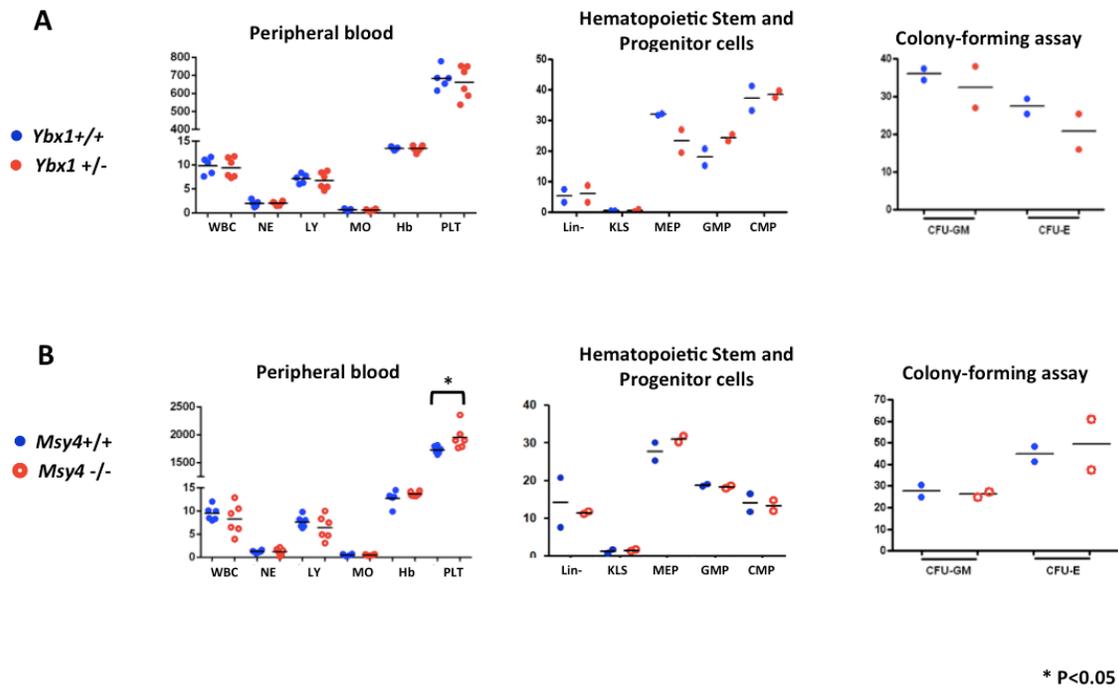
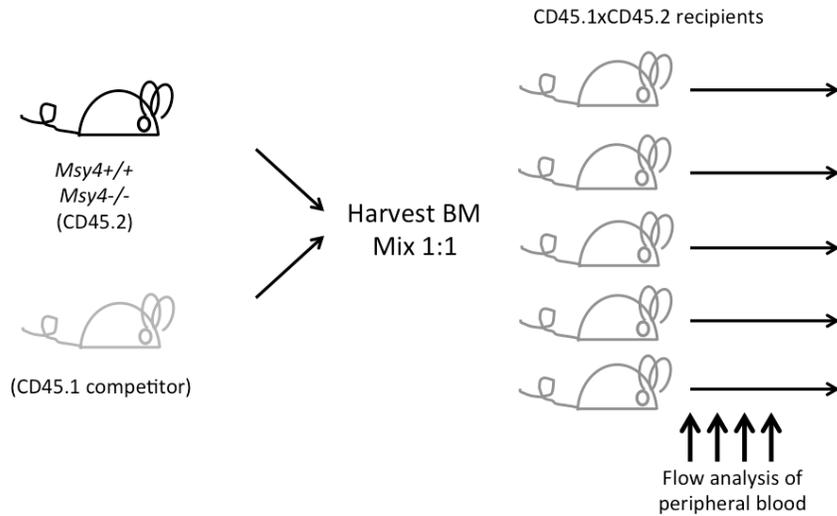


Figure 2-2. Competitive repopulation of *Msy4*^{-/-} mice

A

Experimental Schema



B

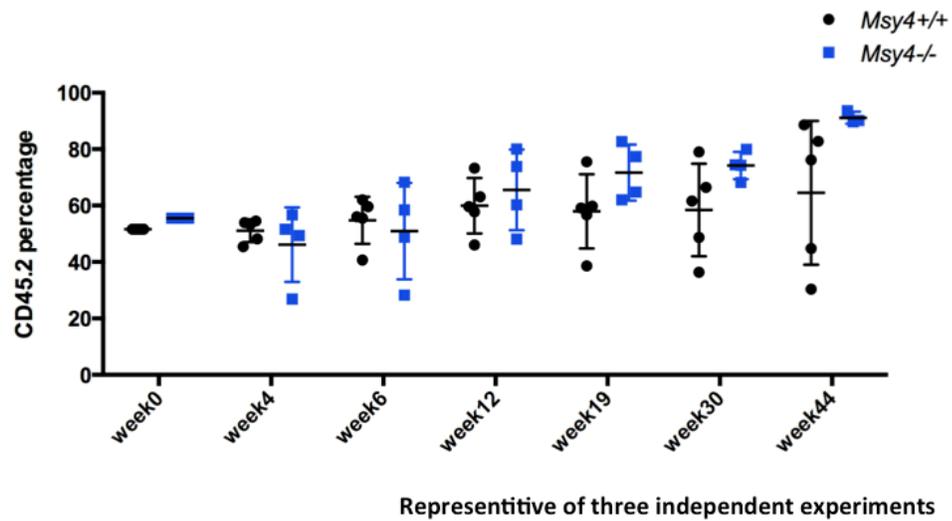


Figure 2-4. Characterization of *Ybx1*^{-/-} E14.5 fetal livers

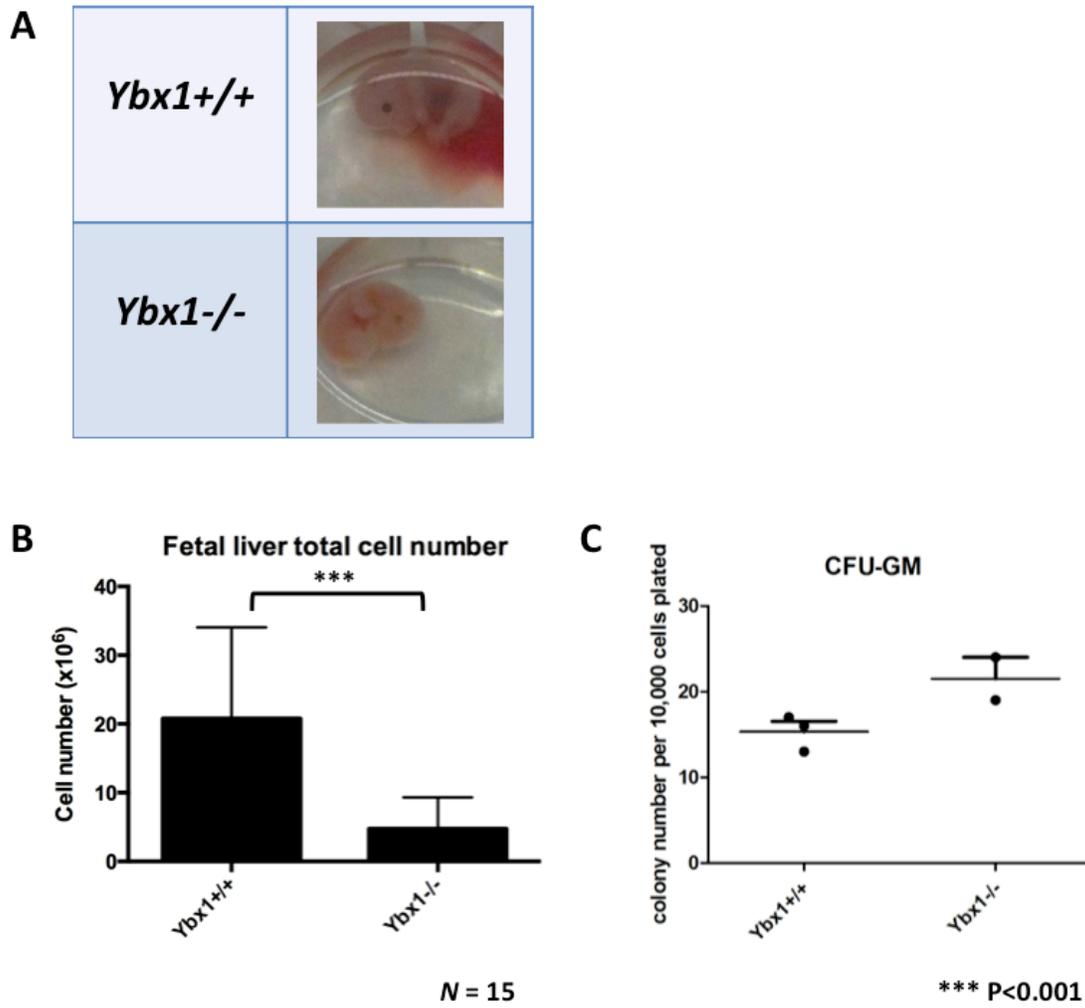


Figure 2-5. Hematopoiesis in mice engrafted with *Ybx1*^{-/-} E14.5 fetal liver cells

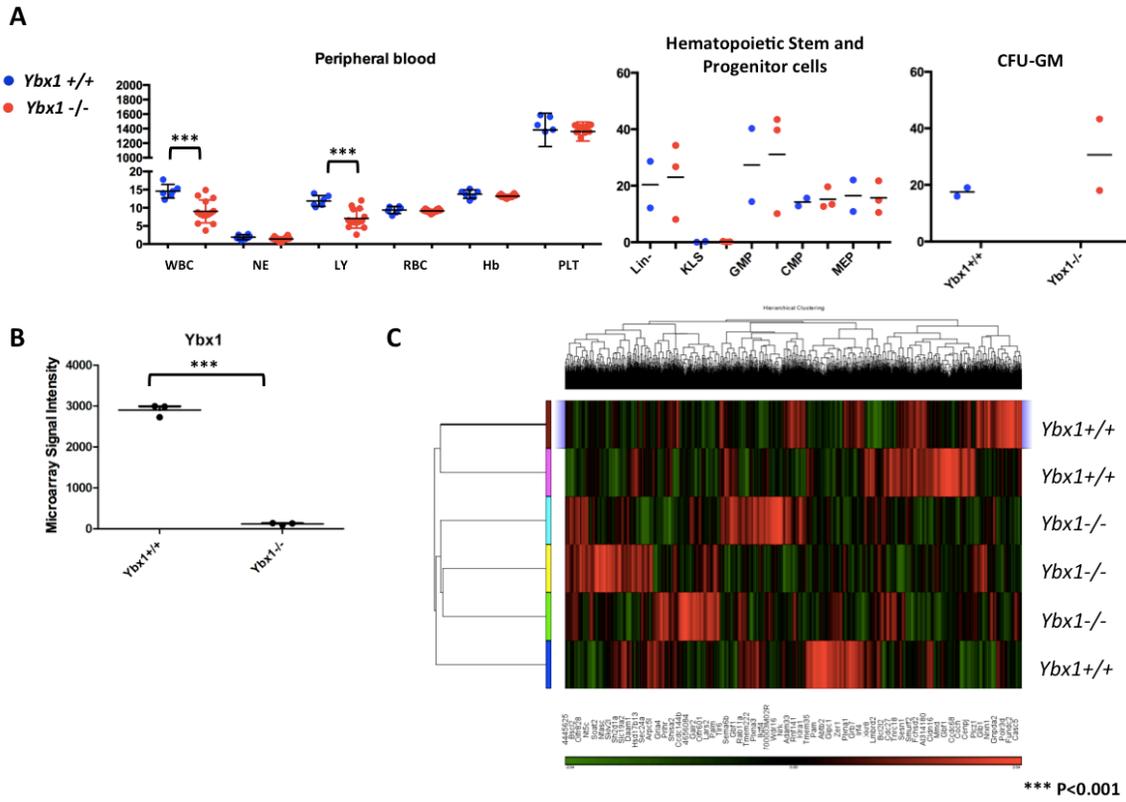
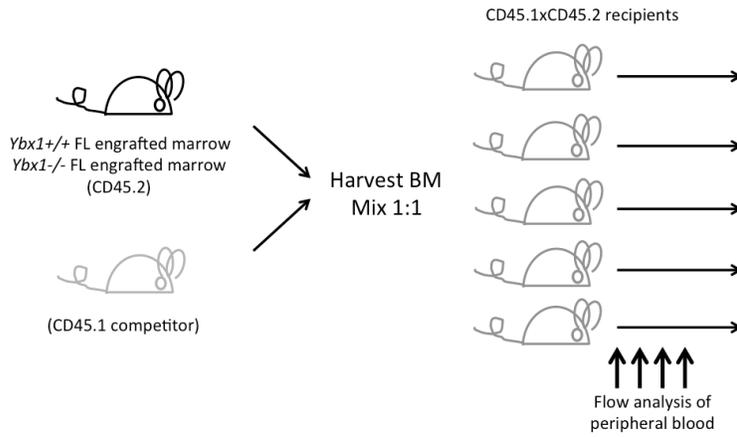


Figure 2-5. Hematopoiesis in mice engrafted with *Ybx1*^{-/-} E14.5 fetal liver cells (Continued)

D

Experimental Schema



E

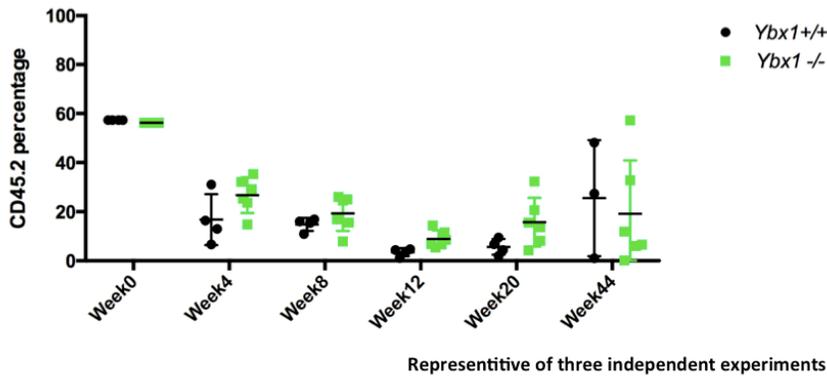


Figure 2-6. Characterization of Ybx1^{lox/+} mice

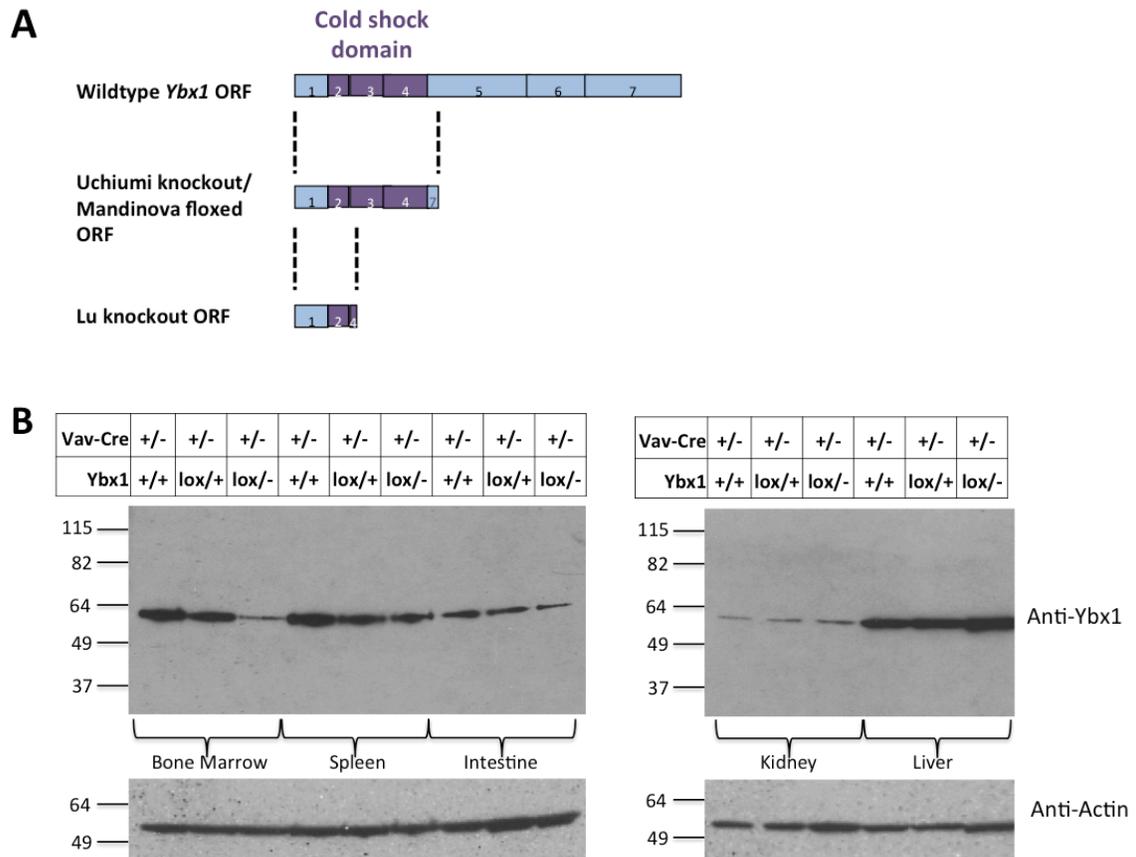
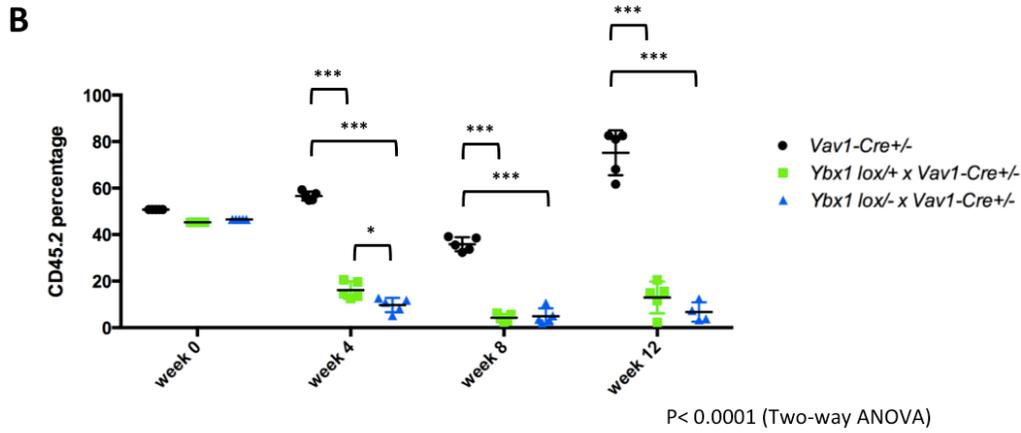
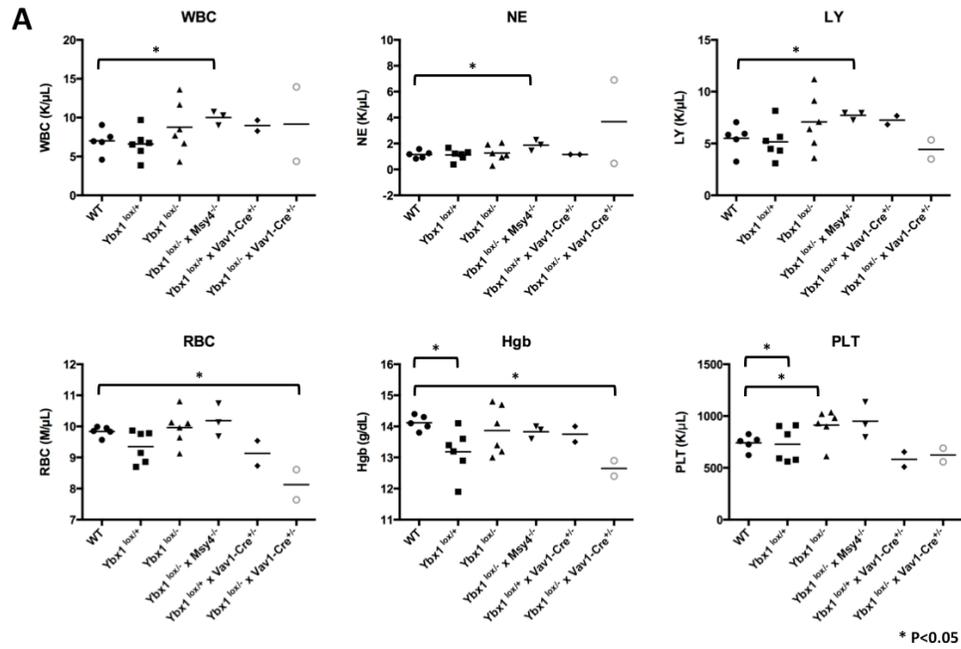


Figure 2-7. Effects of the *Ybx1* floxed allele on adult hematopoiesis



* p<0.05
*** P<0.001

Representative of two independent experiments

Figure 2-8. Effects of *Msy4* deficiency on serial replating by *Ctsg-PML-RARA* bone marrow cells

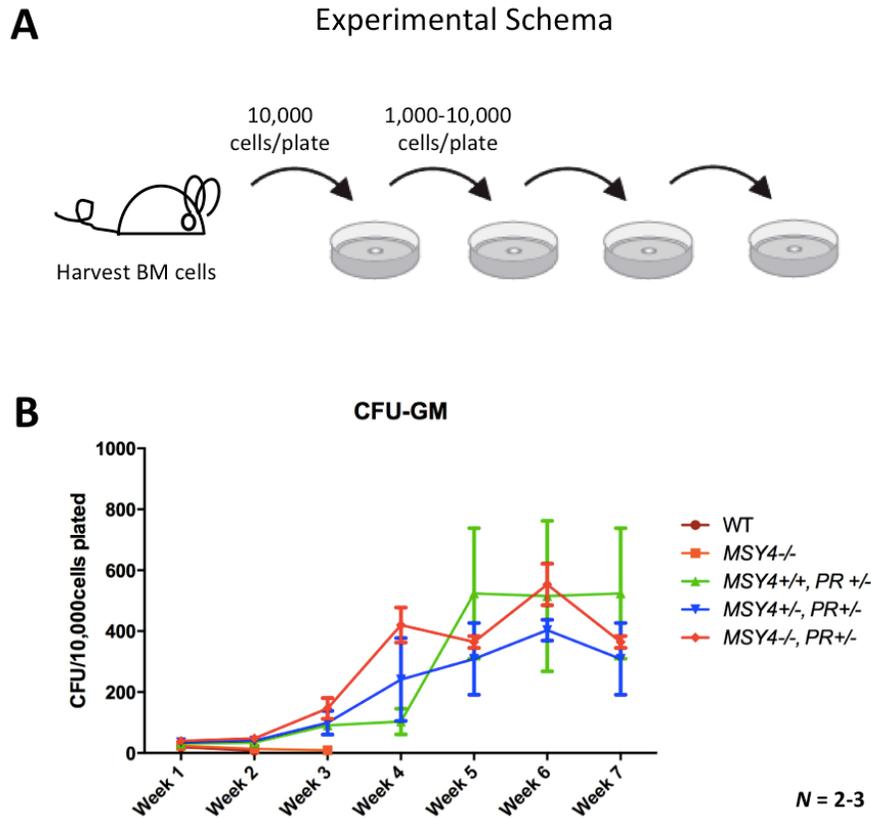


Figure 2-9. Effects of *Msy4* or *Ybx1* deficiency on a retroviral MLL-AF9 leukemia model

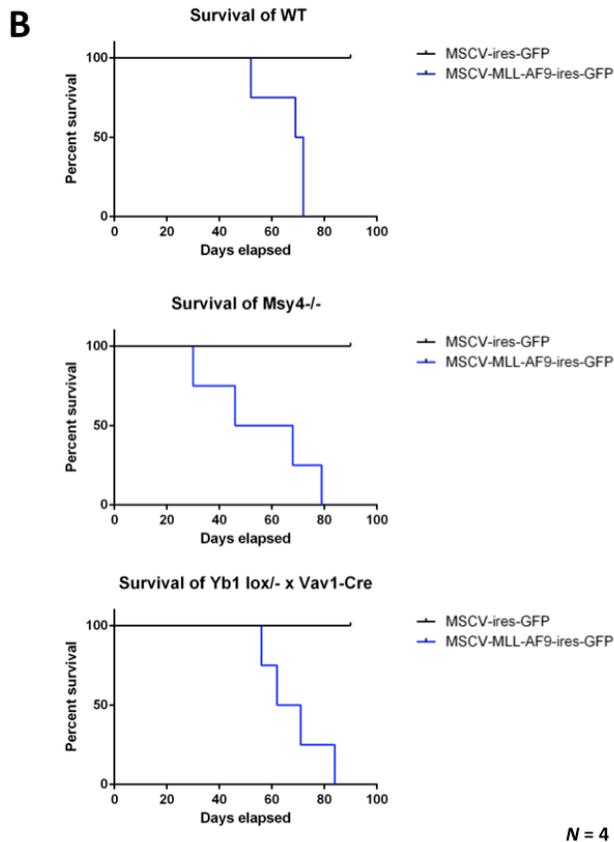
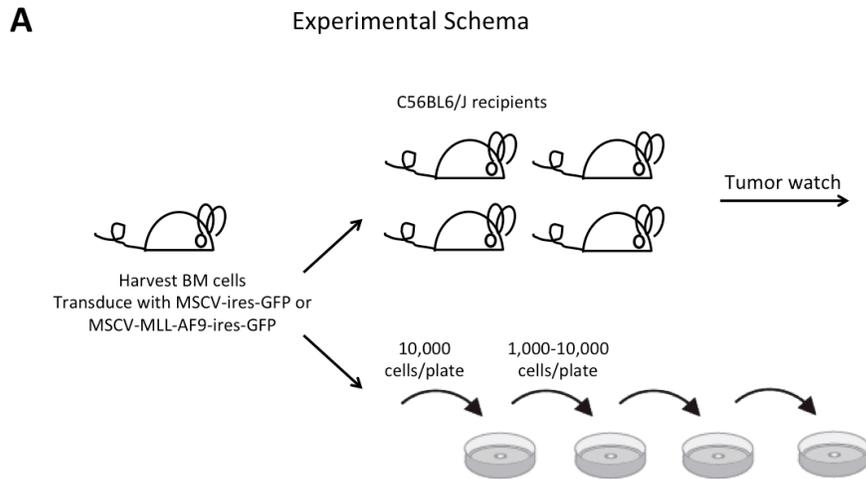


Figure 2-9. Effects of *Msy4* or *Ybx1* deficiency on a retroviral MLL-AF9 leukemia model (Continued)

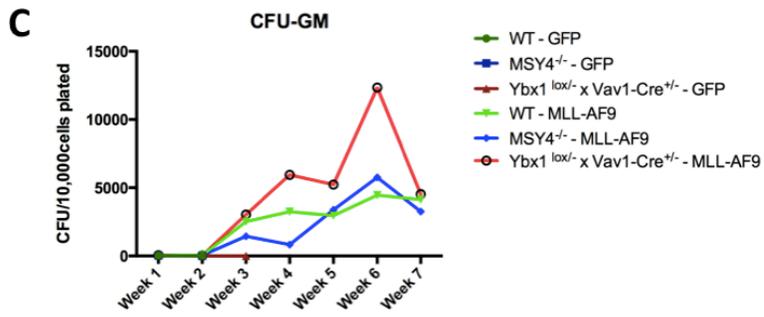


Figure 2-10. Optimal induction of floxing by TAT-cre

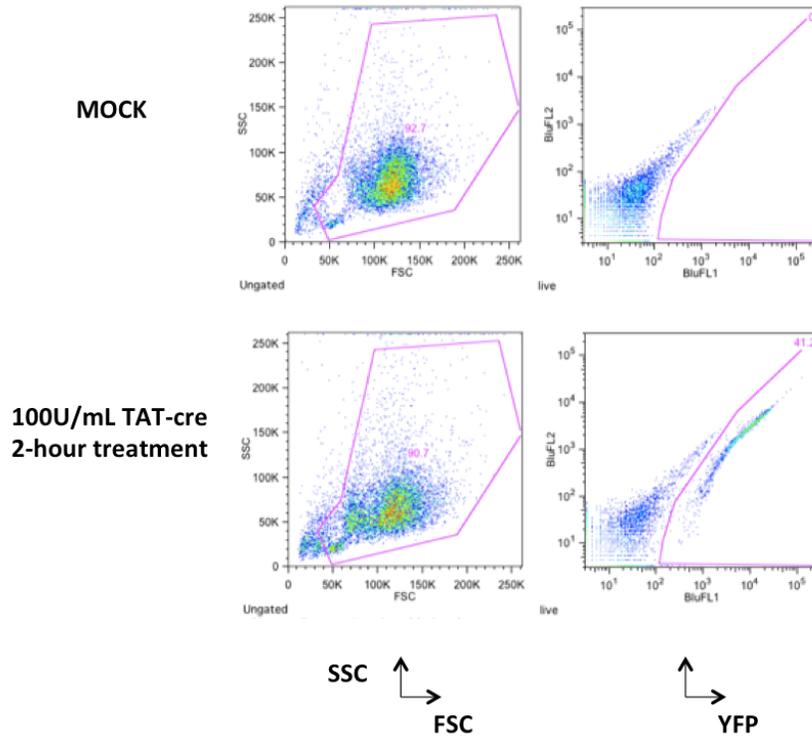


Figure 2-11. Effects of *Ybx1* and *Msy4* deficiency on aberrant replating of bone marrow cells expressing MLL-AF9

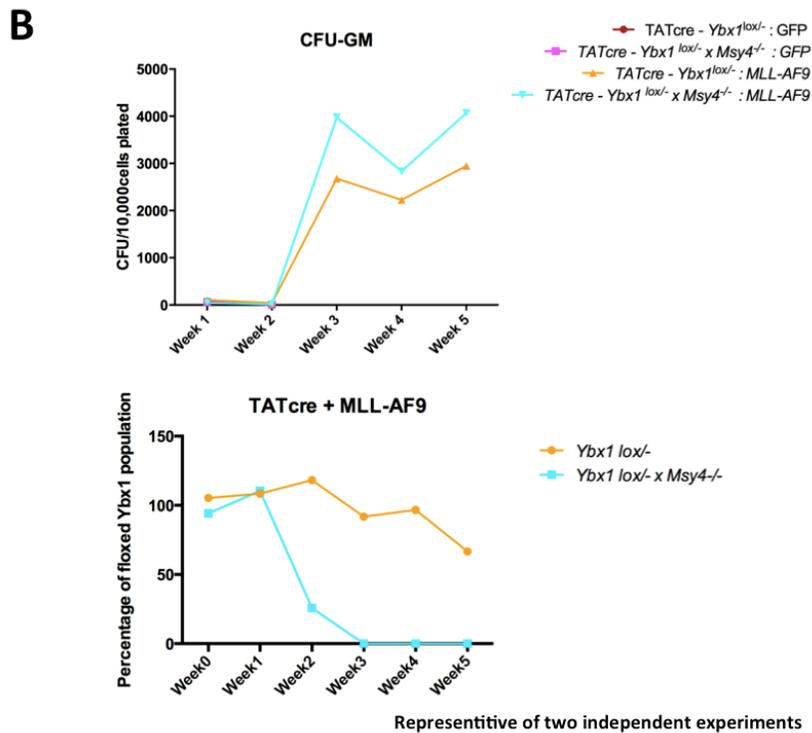
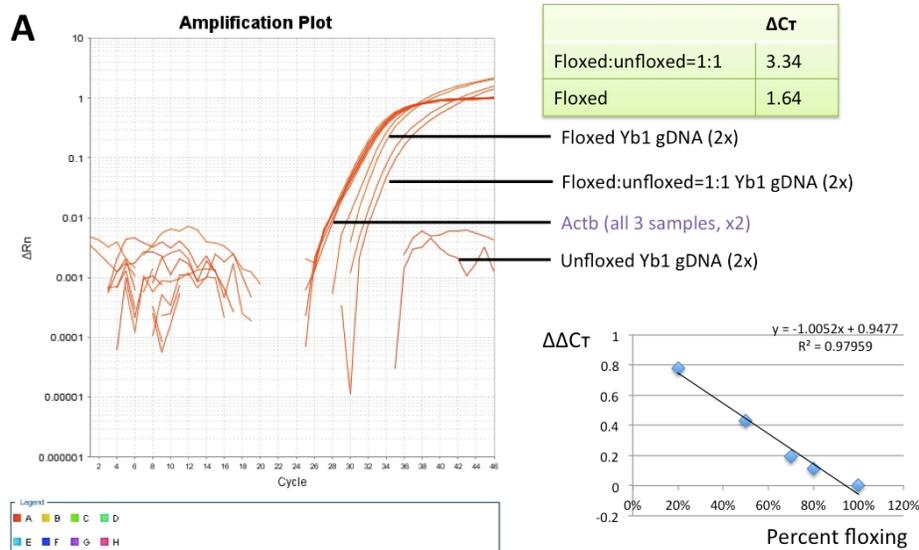


Figure 2-12. Expression of the truncated Ybx1 protein predicted from the floxing of *Ybx1* conditional allele

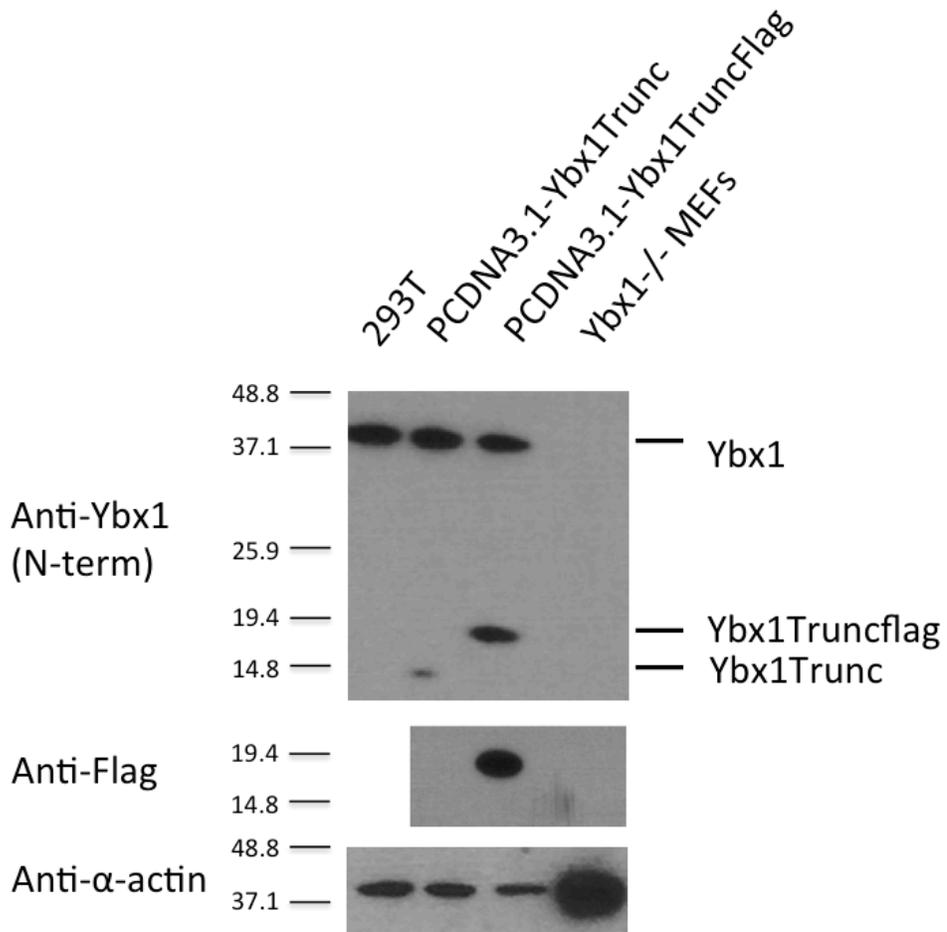


Table 2-1. Genotype distribution from $Ybx1^{lox/+}$ crosses

$Ybx1^{lox/+}$ x $Ybx1^{lox/+}$

Genotype	Predicted % frequency of total	No. of pups (% frequency of total ***)
$Ybx1^{+/+}$	25%	0 (0)
$Ybx1^{lox/+}$	50%	26 (100)
$Ybx1^{lox/lox}$	25%	0 (0)

*** $P < 0.001$

$Ybx1^{lox/+}$ x $Ybx1^{+/-}$

Genotype	Predicted % frequency of total	No. of pups (% frequency of total ***)
$Ybx1^{+/+}$	25%	10 (9.3)
$Ybx1^{+/-}$	25%	12 (11.2)
$Ybx1^{lox/+}$	25%	46 (43.0)
$Ybx1^{lox/-}$	25%	39 (36.4)

*** $P < 0.001$

Statistical comparisons were made using Chi-square test

Table 2-2. TAT-cre testing on *ROSA-lox-STOP-lox-YFP* bone marrow cells. Floxing efficiency is shown in YFP percentage by flow cytometry

WBM	50 units/mL	75 units/mL	100 units/mL
2hr	15.30%	27.30%	41.50%
3hr	17.00%	34.10%	44.40%
4hr	8.09%	28.10%	41.20%

Kit+	50 units/mL	100 units/mL
2hr	21.80%	42.50%
3hr	25.90%	44.40%
4hr	4.89%	45.90%

Chapter 3

Genetic and Functional Heterogeneity of Induced Pluripotent Stem Cells

Derived from Adult Skin Fibroblasts

ABSTRACT

Induced pluripotent stem cells (iPSCs) have tremendous potential as a tool in disease modeling, drug testing, and other applications. However, functional heterogeneity has been frequently observed among iPSC clones, leading to concerns about the validity of studies using single iPSC clones to functionally characterize a given genetic alteration. Published studies characterizing genetic, epigenetic, and transcriptional heterogeneity among iPSCs have thus far been small in scale (2 to 5 clones produced from the same parental cells), or have not addressed functional heterogeneity. Here, we report a study of 24 mouse iPSC (miPSC) clones derived from skin fibroblasts obtained from two different sites of the same 8-week-old C57BL6/J male mouse. We first assessed the ability of each clone to differentiate into hematopoietic progenitor cells *in vitro*, and found a wide range of potentials. To determine whether the hematopoietic potential of the miPSCs was associated with specific genetic alterations, we performed exome sequencing on all 24 clones and the two parental fibroblast pools from which they were derived. The exomes of the two parental fibroblast pools were essentially identical, as expected. When we compared each clone to its parental fibroblasts, we found an average of 28 variants per clone, and of which an average of 26 were unique for each clone. No specific association was found between the mutational spectrum and the hematopoietic potential of each miPSC clone. Finally, we selected three miPSC clones with the greatest and three with the least potential to produce hematopoietic progenitors, and assessed differences in gene expression using an array-based platform. Although the expression profiles of these six samples were nearly identical, 96 genes were differentially expressed between the clones with normal vs. low hematopoietic potential. The transcription factors Wilms tumor 1 (*Wt1*) and Lymphoid enhancer binding factor 1 (*Lef1*) were expressed at significantly lower levels in all three clones with low potential to make hematopoietic progenitors. The roles of these genes for defining the hematopoietic potential of mouse ESCs and iPSCs are currently under investigation.

INTRODUCTION

Pluripotent stem cells, such as embryonic stem cells (ESCs), are defined by their ability to self-renew and differentiate into any somatic cell type. In 2006, the Yamanaka group successfully reprogrammed mouse somatic cells into pluripotent stem cells, referred to as induced pluripotent stem cells (iPSCs), by introducing a combination of four transcription factors: *Oct3/4*, *Sox2*, *c-Myc*, and *Klf4*.¹ One year later, both the Yamanaka group and the Thomson group successfully reprogrammed human somatic cells to iPSCs.^{2,3} Like ESCs, iPSCs demonstrate unlimited self-renewal in culture, express markers associated with pluripotency (such as alkaline phosphatase and SSEA-1), and can generate teratomas comprised of all 3 germs layers (ectoderm, mesoderm and endoderm) in immunodeficient mice.¹

iPSCs reprogrammed from patient cells can be valuable reagents for studying the pathobiology of specific diseases.⁴ However, concerns over the use of iPSCs in translational studies have been raised. For example, reprogramming may select for cells within the pool of parental cells that are the most “fit” for reprogramming. Previous studies have shown an association with human iPSC (hiPSC) reprogramming and mutations known to be related to cancer^{5,6}; however, cells containing these mutations may pre-exist at low frequencies in the parental cell lines.^{5,7} Indeed, using whole genome sequencing to characterize multiple derived mouse iPSC (miPSC) clones from three independent reprogramming experiments, and the parental MEFs (mouse embryo fibroblasts) from which they were derived, Young *et al.* found that all 4 iPSC clones from one experiment shared 157 genetic variants, which could also be detected in <1 in 500 cells in the parental cell pool; in the other two reprogramming experiments, all genetic variants were unique to each clone.⁸ These data suggested that reprogramming and its associated cloning “captures” the mutational history of the reprogrammed cell, and that some cells within a given MEF population may have

a higher fitness for reprogramming due to specific background mutations. Further, iPSC clones have been shown to retain an epigenetic memory of the donor cell types, which may influence their ability to differentiate into different lineages.⁹⁻¹¹ In two independent studies, Kim *et al.* showed that as a result of such epigenetic differences, miPSC clones derived from mouse peripheral blood cells had a greater potential to differentiate into hematopoietic cells than fibroblast-derived, neural progenitor-derived, or smooth muscle cell-derived miPSC clones.^{10,11} Finally, iPSC clones derived from the same parental cells can show variable potentials to differentiate into a specific lineage, such as neurons^{9,10}, hematopoietic progenitors^{10,14}, or hepatocytes¹⁵. This is perhaps not surprising, considering that each clone is derived from an individual cell within a heterogeneous population of parental cells.⁸

While some studies have characterized the genetic, epigenetic and transcriptomic heterogeneity among different iPSC clones^{16,17} (**Table 3-1**), only a handful have evaluated functional heterogeneity, and these have been limited to only a few iPSC clones each. For example, using 3 iPSC lines derived from the same parental fibroblasts, Mills *et al.* found an association between the iPSC proliferative rate and hematopoietic potential, as well as a distinct expression profile and copy number variants (CNVs).¹⁴ Further, Bock *et al.* described a “deviation scoreboard” (derived from genome-wide maps of DNA methylation by reduced representation of bisulfite sequencing, and also gene expression patterns) to predict the neural lineage potentials of hiPSC lines¹⁸; this data was validated in an independent study using the same hiPSC lines to differentiate into motor neurons.¹² In contrast, Kajiwara *et al.* found that gene expression or DNA methylation patterns could not predict the propensity for hepatic differentiation from different hiPSC clones.¹⁵

Although limited in scope, these studies clearly demonstrated that functional and genetic heterogeneity exists within iPSC clones derived from the same source, and that the current standards for pluripotency testing do not fully define the differentiation potential of iPSC clones. This heterogeneity is very important for studies of patient-derived iPSCs, since most studies employing these cells have analyzed only a small number of iPSC clones from each donor source.

In this study, we characterized the genetic and functional heterogeneity of 24 miPSC clones derived from skin fibroblasts taken from two different sites (right and left axilla) of the same adult C57BL6/J mouse. Reprogramming was achieved with an integrating polycistronic lentivirus (containing cDNAs encoding *Oct3/4*, *Sox2*, and *Klf4*, as well as an IRES-*GFP* cassette), so that the genetic identity of each iPSC clone could be confirmed by mapping lentiviral integration sites. Twelve clones from each fibroblast pool (24 total) were generated; all were shown to express standard pluripotency markers. The 24 miPSC clones exhibited widely variable abilities to generate hematopoietic progenitors *in vitro* (compared to wildtype mouse embryonic stem cells); some clones consistently demonstrated little or no ability to generate hematopoietic progenitors. We performed exome sequencing on all clones, as well as the two parental fibroblasts from which they were derived. Each clone had a unique pattern of lentiviral integration sites that assured the identity of each clone. We found an average of 28 (range 5-56) total exomic mutations per clone, of which 26 (on average) were unique for that clone. Although a subset of clones contained a small number of common mutations, no common mutations were found among the clones with poor hematopoietic potential. Finally, expression array analyses of six hematopoietic “outlier” clones (three with high and three with low hematopoietic potential) revealed lower expression of the transcription factors Wilms tumor 1 homolog (*Wt1*) and Lymphoid enhancer binding factor 1 (*Lef1*) in all three clones with poor hematopoietic potential. These findings will serve as a

foundation to investigate the reasons why some iPSC clones have a limited ability to differentiate into hematopoietic progenitor cells.

MATERIALS AND METHODS

Production of miPSC clones

Skin fibroblasts from the right and left axillae (Ax1 and Ax2) of a single 8 week old adult C57BL6/J male mouse were prepared, and iPSC clones were generated as previously described.¹⁹ Briefly, 2.5×10^5 fibroblasts were seeded on 6-well plates. The next day, the cells were transduced with the OSK-GFP lentivirus (kindly provided by Dr. Tim Townes) at an MOI of 1:3. After 24 hours of incubation with the virus, the cells were trypsinized and transferred to a 100-mm petri dish with a feeder mouse embryonic fibroblast (MEF) layer and mouse embryonic stem cell (ESC) media containing recombinant LIF. Cells were grown for 2-3 weeks with daily media changes before individual clones were picked and expanded on MEF feeder layers.

Pluripotency characterization

After passaging for at least 4 weeks, 24 clones were analyzed by flow cytometry for GFP expression, which indicated that the lentivirus is stably integrated. All clones were tested for expression of the ESC marker Oct3/4, using wildtype mESC and MEFs as controls. For intracellular Oct3/4 staining (eBioscience, San Diego, CA), cells were fixed with 4% paraformaldehyde and permeabilized with 1% saponin. Pluripotency staining for other markers, including Nanog, SSEA-1, and Alkaline Phosphatase for all 24 clones, as well as teratoma formation in NSG mice for the 6 hematopoietic outlier clones, are in progress.

In vitro hematopoietic differentiation from miPSCs

The miPSCs hematopoietic differentiation assay is modified from an hiPSC hematopoietic differentiation protocol.²⁰ Briefly, 1×10^5 single miPSCs or mESCs were seeded in gel-coated 100-mm petri dish with OP9 stromal cells overgrown for 8-10 days in differentiation media containing 10% fetal bovine serum (FBS),

100 μ M monothioglycerol (Sigma-Aldrich, St. Louis, MO), and 50 μ g/ml ascorbic acid (Sigma-Aldrich, St. Louis, MO). Media was changed daily for 7 days, at which time all the cells in the dish, including OP9s, were collected. Up to 1×10^7 unsorted cells were stained with the following monoclonal antibodies: B220, CD3e, Gr1, Ter119, Kit, Sca, CD34, and CD16/32 (FC γ) (eBioscience, San Diego, CA) and analyzed by flow cytometry. 1×10^5 unsorted cells were plated into 1.1 ml of methylcellulose media containing Epo, SCF, IL-3, and IL-6 (MethoCult GF M3434; Stem Cell Technologies, British Columbia, Canada) in 60-mm petri-dishes in triplicate. Colony numbers were counted after 7-8 days of culture. After dissolving the MethoCult in warm media, cells were stained with the myeloid and erythroid lineage markers CD34, CD11b, Kit, Gr-1, and Ter119 (eBioscience, San Diego, CA) and analyzed by flow cytometry. 1×10^5 unsorted cells were stained with Wright-Giemsa stain (Sigma-Aldrich, St. Louis, MO) for morphologic examination, both after 7 days of OP9 culture and after another 7 days in MethoCult. Multiple lots of OP9 cells from ATCC and multiple lots and brands of FBS were systematically tested, and hematopoietic differentiation efficiency was found to be dependent on neither. (**Table 3-2**)

Illumina library construction and exome sequencing

Genomic DNA from all 24 miPSC clones and the two parental fibroblast lines were fragmented using a Covaris LE220 DNA Sonicator (Covaris, Woburn, MA) within a size range between 100-400bp using the following settings: volume = 50 μ L, temperature = 4 $^{\circ}$ C, duty cycle = 20, intensity = 5, cycle burst = 500, time = 120 seconds. The fragmented samples were transferred from the Covaris plate and dispensed into a 96 well BioRad Cycle plate by the CyBio-SELMA instrument. Small insert dual indexed Illumina paired end libraries were constructed with the KAPA HTP sample prep kit according to the manufacturer's recommendations (KAPA Biosystems, Woburn, MA) on the SciClone instrument according to the manufacturer's recommendations (Perkin Elmer, Waltham, MA). Dual indexed adaptors were incorporated during ligation; the same 8bp index

sequence is embedded within both arms of the library adaptor. Libraries were enriched with a single PCR reaction for 8 cycles. The final size selection of the library was achieved by a single AMPure XP paramagnetic beads (Agencourt, Beckman Coulter Genomics, Beverly, MA) cleanup targeting a final library size of 300-500bp. The libraries undergo a qualitative (final size distribution) and quantitative assay using the HT DNA Hi Sens Dual Protocol Assay with the HT DNA 1K/12K chip on the LabChip GX instrument (Perkin Elmer, Waltham, MA). Twenty six libraries (from the 24 iPSC clones and the two parental fibroblast pools), at 192ng per library, were pooled pre-capture on the Ep5075 platform, captured (see ***Exome capture, and validation capture***), and sequenced on an Illumina HiSeq 2000 using 100 bp paired-end reads. Exome sequencing coverage for the 24 iPSC clones and the fibroblast preparations from which they were derived are included in **Table 3-3**.

Variant detection pipeline

Sequence data was aligned to mouse reference sequence mm9 (with the OSK vector sequence added) using bwa version 0.5.9²¹ (params: -t 4 -q 5:). Bam files were deduplicated using picard version 1.46.

Single Nucleotide Variants (SNVs) were detected using the union of three callers: 1) samtools version r963²² (params: -A -B) intersected with Somatic Sniper version 1.0.2²³ (params: -F vcf -q 1 -Q 15) and processed through false-positive filter v1 (params: --bam-readcount-version 0.4 --bam-readcount-min-base-quality 15 --min-mapping-quality 40 --min-somatic-score 40) 2) VarScan version 2.2.6²⁴ filtered by varscan-high-confidence filter version v1 and processed through false-positive filter v1 (params: --bam-readcount-version 0.4 --bam-readcount-min-base-quality 15 --min-mapping-quality 40 --min-somatic-score 40), and 3) Strelka version 0.4.6.2²⁵ (params: isSkipDepthFilters = 1).

Indels were detected using the union of 4 callers: 1) GATK somatic-indel version 5336²⁶ filtered by false-indel version v1 (params: --bam-readcount-version 0.4 --

bam-readcount-min-base-quality 15), 2) pindel version 0.5²⁷ filtered with pindel false-positive and vaf filters (params: --variant-freq-cutoff=0.08), 3) VarScan version 2.2.6²⁴ [filtered by varscan-high-confidence-indel version v1 then false-indel version v1 (params: --bam-readcount-version 0.4 --bam-readcount-min-base-quality 15), and 3) Strelka version 0.4.6.2²⁵ (params: isSkipDepthFilters = 1).

Viral integration sites were detected using Breakdancer version 1.4.1²⁸

Exome capture, and validation capture

Two library pools were made for exome capture, each containing all 26 libraries and a total input of ~5ug into capture. One pool was captured using the Agilent SureSelect Mouse All Exon Library Kit according to manufacturer's recommendations with these exceptions:

- 1) 5 µg Mouse Cot DNA and 1mM library adapter blockers were added to the hybridization reaction.
- 2) Each sample was amplified in the PCR using 20µl of enriched ssDNA library fragments, KAPA HotStart Polymerase, and 200nM each forward primer and reverse primer.

The other pool was captured using the Nimblegen SeqCap EZ Library reagent with the same exceptions. Both products have a probe space of ~50Mb. The final concentration of each capture pool was verified through qPCR utilizing the KAPA Library Quantification Kit - Illumina/LightCycler® 480 kit according to the manufacturer's protocol (Kapa Biosystems, Woburn, MA) to produce cluster counts appropriate for the Illumina HiSeq2000 platform. Each capture pool was loaded across 5 lanes of the HiSeq2000 version 3 flow cell according to the manufacturer's recommendations (Illumina, San Diego, CA). 2 X 101bp read pairs were generated for each sample, yielding approximately 6-7Gb of data per sample.

For the validation array, genomic DNA of all 24 miPSC clones was isolated from sorted GFP positive iPSCs to minimize MEF contamination. The custom capture reagent (NimbleGen) contained all predicted somatic mutations from all 24 iPSC clones. Capture was performed as described above for the NimbleGen exome reagent. The capture validation array coverage for the 24 iPSC clones, and the fibroblast preparations from which they were derived are included in **Table 3-4**.

Expression profiling

Expression arrays were performed as previously described.²⁹ Briefly, RNA from six miPSC clones as well as wildtype mESC B6/GFP (2 replicates) were prepared from sorted GFP positive cells to minimize MEF contamination; RNA was purified using the TRIzol reagent (Life Technologies, Carlsbad, CA), processed with the WT-Ovation RNA Amplification System (NuGen Technologies, San Carlos, CA) and analyzed using the Mouse Exon 1.0ST array (Affymetrix, Santa Clara, CA) according to standard protocols from the Genome Technology Access Center at Washington University in St. Louis (<https://gtac.wustl.edu/index.php>). Partek Genomics Suite (Partek, St. Louis, MO) was used for unsupervised hierarchical clustering of miPSC and mESC global RNA levels and two-way ANOVA. A *P* value of ≤ 0.05 was considered to be significant.

Results

Functional heterogeneity among miPSC clones derived from the same parental fibroblasts

To investigate functional heterogeneity among iPSC clones derived from the same parental cells, we generated 12 miPSC clones from two independent pools of fibroblasts from the same adult C57BL6/J mouse. Reprogramming was performed using an established polycistronic lentivirus containing cDNAs encoding *OCT3/4*, *SOX2*, *KLF4* (OSK)¹⁹, and an IRES-*GFP* cassette to mark stably transduced cells. All clones were GFP positive and expressed Oct3/4 (**Table 3-5**). Six of the 24 iPSC clones were evaluated for pluripotency by injecting them into immunodeficient mice, with assessment for cystic teratoma formation currently pending. Three of the tested clones (Ax1-35, Ax2-26 and Ax2-39) were from lines with robust hematopoietic potential, and three were from lines (Ax1-18, Ax2-34 and Ax2-48) with little or no ability to generate hematopoietic progenitors (see below).

To induce the production of murine hematopoietic stem/progenitor cells (HSPCs) from ESCs and iPSCs, we modified a protocol for hiPSC hematopoietic differentiation, as described in Methods and Materials. After co-culture on OP9 stromal cells for one week, control wild type mESC lines derived from C57BL/6 mice (B6/BLU or B6/GFP) consistently differentiated into hematopoietic progenitors, as determined by morphologic examination (**Figure 3-1A**), and the identification of cells with the immunophenotypic characteristics of KLS cells (Lin⁻Kit⁺Sca⁺), common myeloid progenitors (CMPs), granulocyte-macrophage progenitors (GMPs), and megakaryocyte-erythroid progenitors (MEPs) (**Figure 3-1B**). After another week of culture in MethoCult with hematopoietic cytokines (SCF, IL-3, IL-6, and Epo), colony forming units (CFUs) were enumerated as an independent measure of hematopoietic progenitor production. Erythrocytes and mast cells were readily identified by morphologic examination (**Figure 3-1C**) and

cells expressing CD34, Kit, Ter119, and CD11b were quantified by flow cytometry (**Figure 3-1D**).

In two independent experiments, the 24 miPSC clones exhibited variable but reproducible potentials in their ability to produce functional hematopoietic progenitor cells, as measured by colony formation in methylcellulose (**Figure 3-2A**). Although most iPSC clones produced colonies as efficiently as wild type mouse ESC lines, a few consistently displayed a reduced potential. One clone (Ax1-18) was incapable of forming colonies. Immunophenotyping of the cells from the MethoCult cultures from several clones with a low level of hematopoietic colony production (Ax2-48, Ax1-10, and Ax2-34) revealed an unaltered percentage of CD11b⁺, CD34⁺, Kit⁺, and Ter119⁺ cells (**Figure 3-2C**), indicating that the few progenitor cells that differentiated from these clones were phenotypically normal.

Immunophenotypic analysis of the miPSC-derived progenitors revealed variable numbers of KLS cells, GMPs, CMPs, and MEPs among the 24 clones. This variability did not directly correlate with the colony performing ability, except for one clone (Ax1-18). After 7 days of co-culture on OP9 cells, Ax1-18 produced very few cells with the immunophenotypic characteristics of KLS cells, GMPs, CMPs, or MEPs; this clone was incapable of forming hematopoietic colonies on Methocult media (**Figure 3-2B**).

Genetic heterogeneity among miPSC clones derived from the same parental fibroblasts

We performed exome sequencing on all 24 miPSC clones (and the two parental fibroblast pools from which they were derived) to determine the genetic relationships of the miPSC clones to each other, and to their parental fibroblasts.

Exome sequencing coverage is shown in **Figure 3-3A** and **Table 3-3**. Using genomic DNA collected from miPSCs purified by flow cytometry (GFP-positive), all variants were validated using a liquid phase custom capture array and deep digital sequencing, which also allowed us to accurately determine the variant allele frequency (VAF) and clonal architecture of each iPSC clone. The validation array coverage for the 24 iPSC clones and parental fibroblasts are shown in **Figure 3-3B** and **Table 3-4**; average coverage was 604x. As expected, few differences were detected between the two parental fibroblast pools from the same mouse; most had VAFs of <20%, suggesting that they arose during expansion in tissue culture. In contrast, when comparing the miPSCs to their parental fibroblasts, a total of 606 miPSC variants (**Table 3-6**; **3-7**) were detected in the 24 clones (mean 28, range 5-56); 27 of these 606 variants were shared among different clones (**Table 3-7**). No correlation was observed between the number of mutations and the hematopoietic differentiation potential of the miPSC clones ($r^2=0.0006065$) (**Figure 3-3C**).

Among the common mutations, five Ax2 clones (Ax2-11, Ax2-16, Ax2-24, Ax2-26, and Ax2-39) shared a missense mutation in *Hjulp* (Holliday junction recognition protein; VAF ~20%), while five Ax1 clones (Ax1-3, Ax1-5, Ax1-8, Ax1-10, and Ax1-18) shared a missense mutation in *Dux* (Double homeobox; VAF ~20%). Both of these genes have been implicated in cancer (**Table 3-7**). However, since nearly all of these shared mutations have VAFs in the iPSCs that are in the 5-20% range, and since most of these mutations were also detected at low levels in the parental fibroblast pools, the data are not consistent with the idea that the clones are derived from the same parental cells (where the shared mutations in the iPSC clones would be expected to have VAFs of ~50%). The origin and significance of these shared mutations is therefore unclear. However, the large, identical indel present in the *Ppig* gene of clones Ax1-2 and Ax1-35 is diagnostic of a shared parental origin: these two clones arose from the same small population of fibroblasts, and they clearly acquired additional mutations

later (10 private mutations exist in Ax1-2, and 35 in Ax1-35, assuring their unique identities; each has distinct lentiviral integration sites [**Table 3-8**]).

Among the private mutations (excluding silent mutations), there are mutations in several genes that have been implicated in cancer development, such as *Flt1* in Ax1-10 (VAF ~50%), *Bcl-6* in Ax-16 (VAF ~40%), and *Ptpn9* in Ax-23 (VAF ~50%); there are also mutations in genes with important roles in ESC development and hematopoiesis, such as *Dnmt1* in Ax1-23 (VAF ~50%), *Dot1l* in Ax1-8 (VAF ~50%), and *Gata-4* in Ax2-30 (VAF ~50%) (**Table 3-6**). Since the VAFs of these mutations are all in the range of 50%, they are probably heterozygous mutations present in all the cells of the iPSC clone. They may represent random but relevant mutations in the parental cells that contributed to reprogramming 'fitness', or they may have occurred at the time of reprogramming, and facilitated the expansion of cells^{5,7,8}.

An average of 3 OSK lentiviral integration sites (range 1-8) were identified in all 24 miPSC clones; each had a unique set of integration events, establishing clonal identity (**Table 3-8**). No insertion events were identified in genes known to be important in ESC/iPSC function or hematopoietic development. No integration sites were shared by all four miPSC clones with poor hematopoietic potential. Several integration "hotspots" was identified: Chr2: 98502394-98507281 (14 clones from both Ax1 and Ax2), Chr9: 3000297-3034834 (15 clones from both Ax1 and Ax2), and ChrX: 100516732- 100525464 (5 clones from both Ax1 and Ax2). Of these, only the hotspot on Chromosome 2 has previously been reported³⁰. Integrations at these "hotspots" were identified in clones with both good and poor hematopoietic potential (**Table 3-9**). Breakpoint assemblies revealed two integration sites that were shared by three clones, breakpoint (ChrX:100525589; OSK:6660) was shared between Ax2-30 and Ax2-48, and breakpoint (ChrX:100516717 OSK:3115) was shared among Ax2-30, Ax2-48,

and Ax2-11. Finally, we evaluated the expression of “nearest neighbor” genes for each integration site in the six outlier clones with good and poor hematopoietic potential. Only one gene, *Zfp280d*, displayed altered expression; it lies 149Kb downstream from a lentiviral integration site in clone Ax2-26, and was expressed at ~50% the level found in the other 5 clones. The integration sites in the three clones with poor hematopoietic potential were not associated with altered expression of any nearest neighbor genes (**Table 3-10**). These data suggest that lentiviral integration sites are unlikely to explain the functional heterogeneity observed among the 24 clones.

Most of the iPSC clones have a distinct group of variants with VAF clusters of approximately 50%, suggesting that these represent heterozygous mutations present in nearly all of the cells in the clone (**Figure 3-3D**). These data suggest that these variants were most likely present in the cell that was reprogrammed, and represent private mutations in each parental cell that were “captured” by cloning⁸. Analysis of the sequencing data from the parental fibroblasts did identify a variant in *Dcbld1* with a VAF of 4.69% in Ax1, and 3.11% in Ax2 fibroblasts, suggesting that a small fraction of cells within each independent skin sample (about 7-8% of cells) contained this mutation; it probably represents mosaicism within the skin of this animal. Indeed, 4 independent iPSC clones (2 from Ax1, and 2 from Ax2) had the same exact variant, but with VAFs of ~50%. These data clearly demonstrate how preexisting mutations in parental cells are captured by reprogramming and cloning. To determine whether this mutation improved fitness for reprogramming, we analyzed a total of 96 iPSC clones from skin fibroblasts of this mouse, and found the variant in a total of 9 clones (9.4%). From the VAF in the primary skin samples, we estimate that 7.8 % of the skin cells contained the variant, which is not statistically different from the number of iPSCs that contained the mutation (9.4%, $P = 0.78$). Many of the miPSCs also contained several variants at a frequency lower than 50%, suggesting that these mutations were most likely acquired after reprogramming, during expansion of the cells in

tissue culture. Some of the clones, such as Ax1-5 and Ax2-6, clearly have prominent VAF clusters of 25% and 10%, respectively, which probably represent subclones that arose shortly after the reprogramming event. Since these subclones represent only a portion of the cells in the clone, they may not provide a significant growth advantage, since they had not become dominant at the time of sampling.

Expression profiling of selected miPSC clones with different hematopoietic potential

To identify differences in gene expression in clones with different functional properties, we selected three miPSC clones (Ax1-35, Ax2-26 and Ax2-39) with hematopoietic potential similar to that of wildtype mESC from the B6/BLU line, and three clones with a consistently limited potential to form hematopoietic progenitors in the colony assay (Ax1-18, Ax2-34 and Ax2-48). RNA was prepared from each miPSC clone after sorting for GFP positive cells (to remove contaminating MEFs), and then analyzed on the Mouse Exon1.0 ST array (Affymetrix). Wildtype B6/GFP mESCs were also analyzed as an ESC control. All clones were maintained on MEFs in ESC media to maintain pluripotency prior to the collection of cells.

Unsupervised hierarchical clustering revealed no significant differences in the global expression patterns among the six miPSC clones and B6/GFP mESCs (**Figure 3-4A**). However, when we performed a supervised analysis (two-way ANOVA) comparing the three clones with good and poor hematopoietic progenitor potentials, we found 96 genes that were expressed at significantly different levels between the two groups: 42 genes were downregulated and 54 genes were upregulated in the 3 clones with poor hematopoietic potential ($P < 0.05$, fold change >2 or <-2) (**Figure 3-4B**). We next examined the annotation and known functions of each named gene. Two transcription factors known to be

important for embryonic and/or hematopoietic development (Wilms tumor 1 homolog [*Wt1*] and Lymphoid enhancer binding factor 1 [*Lef1*]) were expressed at significantly lower levels in the three clones with poor hematopoietic potential. However, the expression of *Wt1* in these clones was similar to that of B6/GFP ESCs, which represent the “gold standard” for hematopoietic potential ($P > 0.05$). Importantly, the pluripotency genes *Pou5f1* (encoding Oct3/4), *Nanog*, and *Sox2* were expressed at similar levels in all six iPSC clones, and in B6/GFP ES cells (**Figure 3-4C**).

Discussion

In this study, we evaluated functional differences among the hematopoietic potentials of 24 miPSC clones derived from two independent preparations of skin fibroblasts from the same 8-week-old C57BL6/J mouse. Among the 24 clones, we observed varying abilities to produce hematopoietic progenitor cells that could produce colonies in methylcellulose plating experiments. We defined the mutational landscapes in the exomes of all 24 clones in an attempt to define genetic mechanisms that might be relevant for phenotypic variation, but did not identify specific mutations that explained the phenotypes. Finally, we compared the expression profiles of clones with extreme outlier phenotypes for hematopoiesis *in vitro*; this study yielded a small set of candidate genes (including *Wt1* and *Lef1*) that could be relevant for hematopoietic differentiation in mouse iPSCs. These genes are currently being assessed in functional assays.

Our observations regarding clonal heterogeneity are consistent with previous studies by Kim *et al.*, using 8 miPSC clones derived from mouse peripheral blood cells¹⁰, and Mills *et al.*, using 3 hiPSC clones derived from human fibroblasts¹⁴. Other studies have also reported clonal differences in the abilities of iPSCs to differentiate into neurons^{9,10}, hepatocytes¹⁵, or embryoid bodies (EBs)¹⁸. Among these reports, Boulting *et al.* found one iPSC line (out of three derived from the same parental line) that failed to differentiate into motor neurons, for reasons that were unclear¹². Several of these studies also used molecular tools such as SNP karyotyping, pyrosequencing, bisulfite sequencing, and microarrays to characterize iPSCs, but no clear cut mechanism was defined to explain the phenotypic variation among clones in any of these reports.

In an attempt to determine whether mutations specific to individual iPSC clones might contribute to phenotypic heterogeneity, we performed whole exome sequencing of the 24 miPSC clones used in this study, along with the two

parental fibroblast pools from which they were derived. After identifying potential mutations in each clone, we validated them all using a custom capture array that included all variants for all clones, followed by deep digital sequencing. Most of the variants found in most of the clones had VAFs that clustered at ~50%, suggesting that all of the cells in the clone contained these heterozygous mutations. Studies from our laboratory⁸ previously suggested that mutations present in all the cells of an iPSC clone were probably present in the individual fibroblasts that were reprogrammed and cloned, which in essence “captures” the mutational history of each cell. The data obtained in this study (which is the largest collection of exome sequencing data of iPSCs from a single source), supports the hypothesis that most genetic variation in iPSCs comes from the parental cells from which the iPSCs were derived^{5,16,17,31,32}.

The average number of mutations found in each iPSC clone in this study was 28 (with a range of 5-56). In contrast, Young *et al.* found an average of 11 mutations in the coding regions of each miPSC clone sequenced in that study.⁸ There are at least two potential explanations for this discrepancy: first, the number of mutations may be higher in this study because the fibroblasts were derived from an 8 week old mouse, instead of the mouse embryo fibroblasts (from E13.5 embryos) used in the Young *et al.* study: aging is associated with an increased mutational burden in fibroblasts.³⁵ Secondly, the exome sequencing performed in this study had greater coverage than the whole genome sequencing study (~50x vs. ~20x). Deeper coverage allows for a higher likelihood of identifying variants present in a fraction of the cells in the sample, which were identified in all of the iPSC clones in this study.

We found no association between the hematopoietic potential of each clone, and the number or type of mutations that were identified in each. In the clones with limited ability to form hematopoietic progenitors, we did not find recurring

mutations in genes that are known to affect hematopoietic lineage determination (e.g. *HoxB4*, *Gata2*, *Pu.1*, *Cebpa*, etc.). Likewise, we did not detect any common mutations (or mutations in common pathways) among these clones, and no commonality among the lentivirus integration sites among these clones. However, since our analysis was restricted to the exomes, it is possible that these clones could have common mutations in non-coding regions that could somehow alter the developmental fate of iPSCs. Since mutations of this kind could potentially cause alterations in gene expression, we decided to perform expression profiling on extreme outlier clones.

We compared the expression profiles of 3 iPSC clones with little or no ability to form hematopoietic colonies to that of 3 clones with robust differentiation potential, and 2 replicates from a mouse B6 ES cell line. The expression profiles of all 8 samples were remarkably similar, but we were able to identify a very small subset of genes with highly significant expression differences among the clones with high and low hematopoietic potentials. After annotating these genes, two strong candidates emerged: the transcription factors *Wt1* and *Lef1* both were expressed at lower levels in the clones with reduced hematopoietic potential, although *Wt1* expression in these clones was similar to that of B6/GFP ES cells (which robustly form hematopoietic progenitors). Validation of protein levels, and rescue experiments with ectopic expression of *Wt1* and *Lef1*, are currently in progress.

Wt1 encodes a zinc-finger transcription factor that is inactivated in a subset of embryonic kidney cancers termed Wilm's tumors.³³⁻³⁵ *Wt1* functions as a tumor suppressor, and its mutations are frequently linked to malignancies such as acute leukemia³⁶, breast cancer³⁷, lung cancer³⁸, retinoblastoma³⁹, and others. *Wt1* is expressed during mammalian embryonic development in many tissues, and disruption of the *Wt1* gene in mice has shown that it plays a critical role in

development of the heart, adrenal gland, spleen, kidney, and retina.^{40–43} Cunningham *et al.* recently reported that *Wt1*-deficient mESC exhibit a markedly reduced potential in hematopoietic differentiation *in vitro*, using two independent *Wt1* knockout ESC lines.⁴⁴ *Lef1* is a member of the LEF-1/FCF family of transcription factors, which have been identified as nuclear mediators of Wnt signaling⁴⁵. *Lef1* is expressed in developing B and T cells, as well as neural crest, mesencephalon, tooth germ cells, whisker follicles, and other sites during embryonic development^{46–49}. Lack of *Lef1* leads to abnormal patterning of somites, which give rise to skeletal muscle, cartilage, tendons, vertebrae, spinal nerves, and blood vessels.⁵⁰ In hematopoietic development, *Lef1* is not only critical for lymphopoiesis^{47–49}, but also for myelopoiesis^{51,52} and self-renewal of hematopoietic stem cells⁵³. Based on the Cunningham study, *Wt1* downregulation may represent the most likely candidate gene for hematopoietic phenotype. However, the level of expression in the poor differentiators is similar to that of B6 ESCs; clearly, direct testing of *Wt1* and *Lef1* replacement in these iPSC lines will be required to define their roles for this phenotype.

In summary, we have characterized the functional heterogeneity and mutational landscapes of 24 iPSC clones derived from the same mouse. We did not find exomic mutations that directly explained phenotypic heterogeneity for hematopoietic potential. However, we identified two hematopoietic transcription factors (*Wt1* and *Lef1*) that are dysregulated in clones with a reduced potential to form hematopoietic progenitors. To determine whether either gene is truly relevant for the phenotype, we will overexpress these cDNAs (both individually, and combined) in clones with normal and low hematopoietic potential, and knock them down in mESCs and miPSCs with normal hematopoietic potential. These studies may provide novel insights into the regulation of these transcription factors, and new information regarding the factors that govern hematopoietic lineage determination.

REFERENCES

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–76 (2006).
2. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–72 (2007).
3. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–20 (2007).
4. Bellin, M., Marchetto, M. C., Gage, F. H. & Mummery, C. L. Induced pluripotent stem cells: the new patient? *Nat. Rev. Mol. Cell Biol.* **13**, 713–26 (2012).
5. Gore, A. *et al.* Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63–7 (2011).
6. Laurent, L. C. *et al.* Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. *Cell Stem Cell* **8**, 106–18 (2011).
7. Abyzov, A. *et al.* Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* **492**, 438–42 (2012).
8. Young, M. a *et al.* Background mutations in parental cells account for most of the genetic heterogeneity of induced pluripotent stem cells. *Cell Stem Cell* **10**, 570–82 (2012).
9. Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T, Natesan S, Wagers AJ, Melnick A, Evans T, H. K. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat. Biotechnol.* **28**, 848–855 (2011).
10. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285–90 (2010).
11. Kim, K. *et al.* Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nat. Biotechnol.* **29**, 1117–9 (2011).

12. Boulting, G. L. *et al.* A functionally characterized test set of human induced pluripotent stem cells. *Nat. Biotechnol.* **29**, 279–288 (2011).
13. Hu, B.-Y. *et al.* Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4335–40 (2010).
14. Mills, J. a *et al.* Clonal genetic and hematopoietic heterogeneity among human-induced pluripotent stem cell lines. *Blood* **122**, 2047–51 (2013).
15. Kajiwara, M. *et al.* Donor-dependent variations in hepatic differentiation from human-induced pluripotent stem cells. *Proc. Natl. Acad. Sci.* **109**, 12538–12543 (2012).
16. Cahan, P. & Daley, G. Q. Origins and implications of pluripotent stem cell variability and heterogeneity. *Nat. Rev. Mol. Cell Biol.* **14**, 357–68 (2013).
17. Liang, G. & Zhang, Y. Genetic and epigenetic variations in iPSCs: potential causes and implications for application. *Cell Stem Cell* **13**, 149–59 (2013).
18. Bock, C. *et al.* Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439–52 (2011).
19. Chang CW, Lai YS, Pawlik KM, Liu K, Sun CW, Li C, Schoeb TR, T. T. Polycistronic Lentiviral Vector for “ Hit and Run ” Reprogramming of Adult Skin Fibroblasts to Induced Pluripotent Stem Cells. *Stem Cells* **27**, 1042–1049 (2009).
20. Vodyanik, M. a & Slukvin, I. I. Hematoendothelial differentiation of human embryonic stem cells. *Curr. Protoc. cell Biol.* **Chapter 23**, Unit 23.6 (2007).
21. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–60 (2009).
22. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–9 (2009).
23. Larson, D. E. *et al.* SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* **28**, 311–7 (2012).
24. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–76 (2012).

25. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–7 (2012).
26. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–303 (2010).
27. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–71 (2009).
28. Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods* **6**, 677–681 (2013).
29. Wartman, L. D. *et al.* Expression and function of PML-RARA in the hematopoietic progenitor cells of Ctsg-PML-RARA mice. *PLoS One* **7**, e46529 (2012).
30. Yang, S.-H., Cheng, P.-H., Sullivan, R. T., Thomas, J. W. & Chan, A. W. S. Lentiviral integration preferences in transgenic mice. *Genesis* **46**, 711–8 (2008).
31. Ruiz, S. *et al.* Analysis of protein-coding mutations in hiPSCs and their possible role during somatic cell reprogramming. *Nat. Commun.* **4**, 1382 (2013).
32. Ji, J. *et al.* Elevated coding mutation rate during the reprogramming of human somatic cells into induced pluripotent stem cells. *Stem Cells* **30**, 435–40 (2012).
33. Wagner, K.-D. The complex life of WT1. *J. Cell Sci.* **116**, 1653–1658 (2003).
34. Haber, D. a *et al.* An internal deletion within an 11p13 zinc finger gene contributes to the development of Wilms' tumor. *Cell* **61**, 1257–69 (1990).
35. Gessler, M., Poustka, A., Cavenee, W., Neve, R. L., Orkin, S. H. and Bruns, G. A. Homozygous deletion in Wilms tumours of a zinc-finger gene identified by chromosome jumping. *Nature* **343**, 774–778 (1990).
36. Owen, C., Fitzgibbon, J. & Paschka, P. The clinical relevance of Wilms Tumour 1 (WT1) gene mutations in acute leukaemia. *Hematol. Oncol.* **28**, 13–19 (2010).

37. Silberstein, G. B., Van Horn, K., Strickland, P., Roberts, C. T. & Daniel, C. W. Altered expression of the WT1 wilms tumor suppressor gene in human breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 8132–7 (1997).
38. Oji, Y. *et al.* Overexpression of the Wilms' tumor gene WT1 in de novo lung cancers. *Int. J. Cancer* **100**, 297–303 (2002).
39. Wagner, N. *et al.* The Wilms' tumor suppressor Wt1 is associated with the differentiation of retinoblastoma cells. *Cell Growth Differ.* **13**, 297–305 (2002).
40. Herzer, U., Crocoll, A., Barton, D., Howells, N. and Englert, C. The Wilms tumor suppressor gene wt1 is required for development of the spleen Ute Herzer , Alexander Crocoll , Debra Barton , Norma Howells. *Curr. Biol.* **9**, 837–840 (1999).
41. Kreidberg, J. a *et al.* WT-1 is required for early kidney development. *Cell* **74**, 679–91 (1993).
42. Moore, a W., McInnes, L., Kreidberg, J., Hastie, N. D. & Schedl, a. YAC complementation shows a requirement for Wt1 in the development of epicardium, adrenal gland and throughout nephrogenesis. *Development* **126**, 1845–57 (1999).
43. Wagner, K.-D. *et al.* The Wilms' tumor gene Wt1 is required for normal development of the retina. *EMBO J.* **21**, 1398–405 (2002).
44. Cunningham, T. J., Palumbo, I., Grosso, M., Slater, N. & Miles, C. G. WT1 regulates murine hematopoiesis via maintenance of VEGF isoform ratio. *Blood* **122**, 188–92 (2013).
45. Van Noort, M. & Clevers, H. TCF transcription factors, mediators of Wnt-signaling in development and cancer. *Dev. Biol.* **244**, 1–8 (2002).
46. Van Genderen, C. *et al.* Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. *Genes Dev.* **8**, 2691–2703 (1994).
47. Oosterwegel, M. *et al.* Cloning of murine TCF-1, a T cell-specific transcription factor interacting with functional motifs in the CD3-epsilon and T cell receptor alpha enhancers. *J. Exp. Med.* **173**, 1133–42 (1991).
48. Reya, T. *et al.* Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. *Immunity* **13**, 15–24 (2000).

49. Travis, a, Amsterdam, a, Belanger, C. & Grosschedl, R. LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. *Genes Dev.* **5**, 880–894 (1991).
50. Galceran, J., Sustmann, C., Hsu, S.-C., Folberth, S. & Grosschedl, R. LEF1-mediated regulation of Delta-like1 links Wnt and Notch signaling in somitogenesis. *Genes Dev.* **18**, 2718–23 (2004).
51. Skokowa, J. & Welte, K. LEF-1 is a decisive transcription factor in neutrophil granulopoiesis. *Ann. N. Y. Acad. Sci.* **1106**, 143–51 (2007).
52. Skokowa, J. *et al.* Interactions among HCLS1, HAX1 and LEF-1 proteins are essential for G-CSF-triggered granulopoiesis. *Nat. Med.* **18**, 1550–9 (2012).
53. Reya, T. *et al.* A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* **423**, 409–14 (2003).
54. Hussein, S. M. *et al.* Copy number variation and selection during reprogramming to pluripotency. *Nature* **471**, 58–62 (2011).
55. Martins-Taylor K, Nisler BS, Taapken SM, Compton T, Crandall L, Montgomery KD, Lalande M, X. R. Recurrent copy number variations in human induced pluripotent stem cells. *Nat. Biotechnol.* **29**, 488–491 (2011).
56. Cheng, L. *et al.* Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. *Cell Stem Cell* **10**, 337–44 (2012).
57. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68–73 (2011).
58. Phanstiel, D. H. *et al.* Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat. Methods* **8**, 821–7 (2011).

Figure Legends

Figure 3-1. Hematopoietic differentiation from mESCs/miPSCs

- A. Morphology of wild type mESC-derived cells after 7 days of OP9 coculture (unsorted) by Wright-Giemsa staining. A scale bar of 20 μ m is shown. Arrows indicate cells with the morphologic characteristics of primitive hematopoietic cells.
- B. Immunophenotyping of hematopoietic progenitor cells from wild type mouse bone marrow cells, wildtype mESCs after 7 days of OP9 coculture, and OP9 cells themselves (exposed to same culture condition without seeding mESCs): Lineage⁻ (Lin⁻), KLS (Lin⁻Kit⁺Sca⁺), progenitors (Lin⁻Kit⁺Sca⁻), CMPs (Lin⁻Kit⁺Sca⁻CD34⁺FCg⁻), GMPs (Lin⁻Kit⁺Sca⁻CD34⁺FCg⁺), and MEPs (Lin⁻Kit⁺Sca⁻CD34⁻FCg⁺).
- C. Morphology of day 7 OP9 cocultured mESC-derived cells after 7-8 days of additional culture in MethoCult media containing hematopoietic cytokines (SCF, IL-3, IL-6, and Epo). A scale bar of 20 μ m is shown. Arrows indicate mast cells, while other cells are predominantly erythrocytes at different stages of maturation.
- D. Immunophenotyping of day 7 OP9 cocultured mESC-derived cells after 7-8 days of additional culture in MethoCult media containing hematopoietic cytokines (SCF, IL-3, IL-6, and Epo), using myeloid and erythroid markers CD11b, Gr-1, CD34, Kit, and Ter119.

Figure 3-2. Hematopoietic differentiation potential of the 24 miPSC clones

- A. 100,000 cells from OP9 cocultured mESCs (B6Blu) or miPSCs were plated in MethoCult media containing hematopoietic cytokines (SCF, IL-3, IL-6, and Epo). CFUs were counted after 7 additional days of culture. The relative number of CFUs per 100,000 cells plated from Day7 miPSC-derived progenitors vs. Day7 mESC (B6/BLU)-derived progenitors is shown. iPSC clones are ranked from the highest to the lowest average of two independent experiments. Error bars represent the means +/- one standard deviation.
- B. Fractions of Lin⁻ cells, KLS cells, GMPs, CMPs, and MEPs from miPSCs relative to mESCs after 7 days of OP9 coculture (unsorted), presented in the same order as panel A.
- C. Fractions of CD11b⁺, CD34⁺Kit⁺, and Ter119⁺ cells obtained after 7 days of MethoCult culture containing hematopoietic cytokines (SCF, IL-3, IL-6, and Epo), comparing miPSC-derived progenitors relative to mESC-derived progenitors, in the same order as panel A.

Figure 3-3. Mutational landscapes of the 24 miPSC clones

- A. Mean exome sequencing coverage for the 24 miPSC clones and the fibroblast preparations from which they were derived

- B. Mean validation array coverage for the 24 miPSC clones and the fibroblast preparations from which they were derived
- C. The number of validated mutations detected by exon capture reagent per clone, in the same order as **Figure 3-2A**. No correlation is found between the number of validated mutations and their hematopoietic potential. ($r^2=0.0006065$)
- D. Variant allele frequencies of all validated mutations for each clone. Samples from miPSC preparations derived from the fibroblasts from Axilla 1 (Ax1) and Axilla 2 (Ax2) are shown in separate panels.

Figure 3-4. Expression array data comparing six miPSC clones with normal vs. low hematopoietic potential

- A. Unsupervised hierarchical clustering of RNA expression data from the six miPSC clones and two wildtype B6/GFP mESCs. The six clones do not organize by their hematopoietic phenotypes. miPSC clones with low hematopoietic potential are shown in red; miPSC clones with normal hematopoietic potential are shown in black; B6/GFP mESCs are shown in green.
- B. Differentially expressed genes in clones with distinct hematopoietic differentiation potentials. Two-way ANOVA was performed comparing the three clones with normal vs. the three clones with low hematopoietic potential. Z-scores of probesets with P -value less than 0.05 and fold change greater than 2 are shown, in order of the most downregulated to the most upregulated in clones with low hematopoietic potential. miPSC clones with low hematopoietic potential are shown in red; miPSC clones with normal hematopoietic potential are shown in black; B6/GFP mESCs are shown in green.
- C. Expression data (from the Affymetrix arrays) for candidate transcription factors with the most significantly altered expression in iPSC clones with low hematopoietic potential, and the pluripotency genes *Pou5f1*, *Nanog*, *Sox2*, for the six miPSC clones and two B6/GFP mESC samples. Individual data points represent expression data from each cell line.

Figure 3-1. Hematopoietic differentiation from mESCs/miPSCs

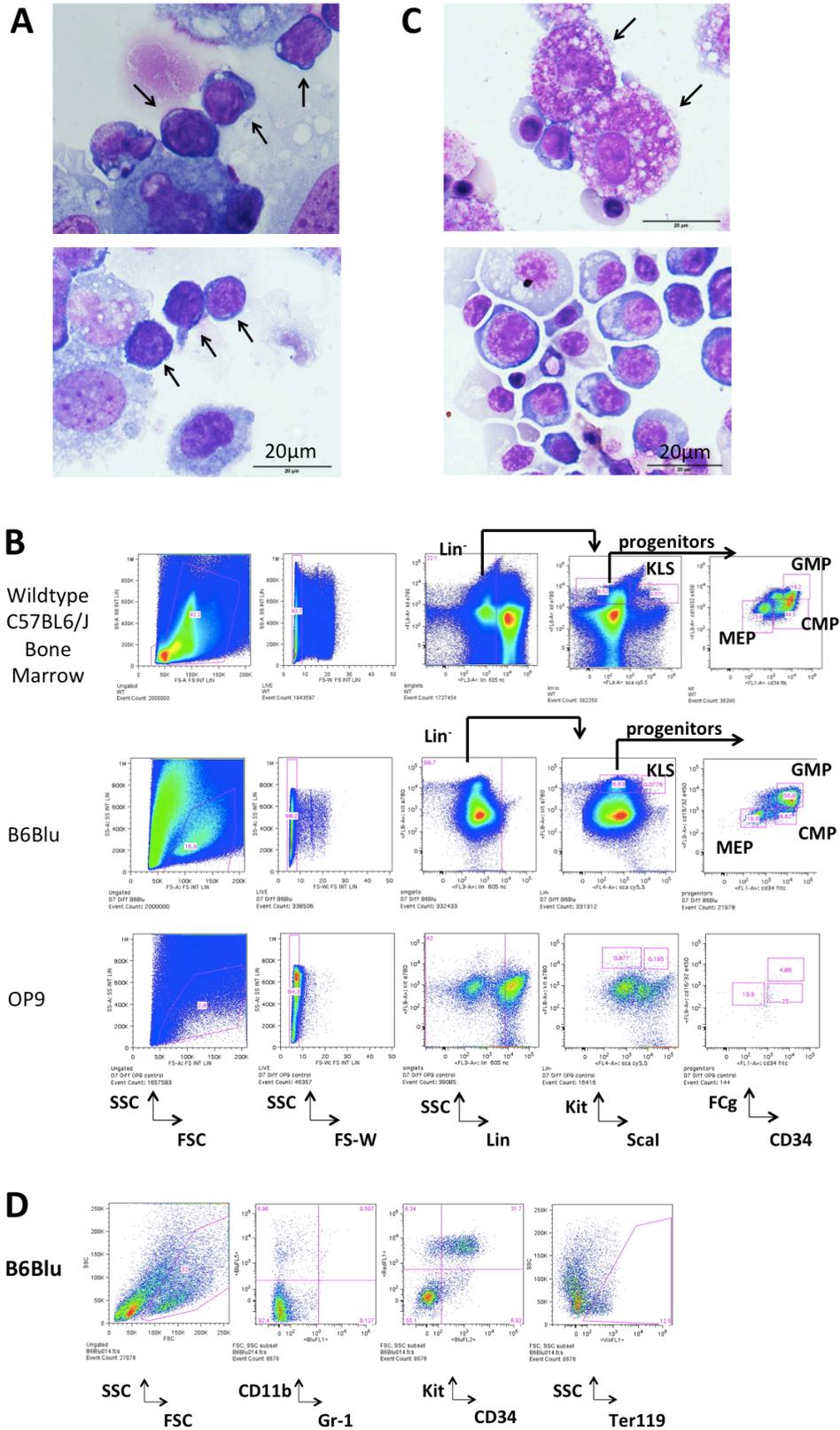


Figure 3-2. Hematopoietic potential of the 24 miPSC clones

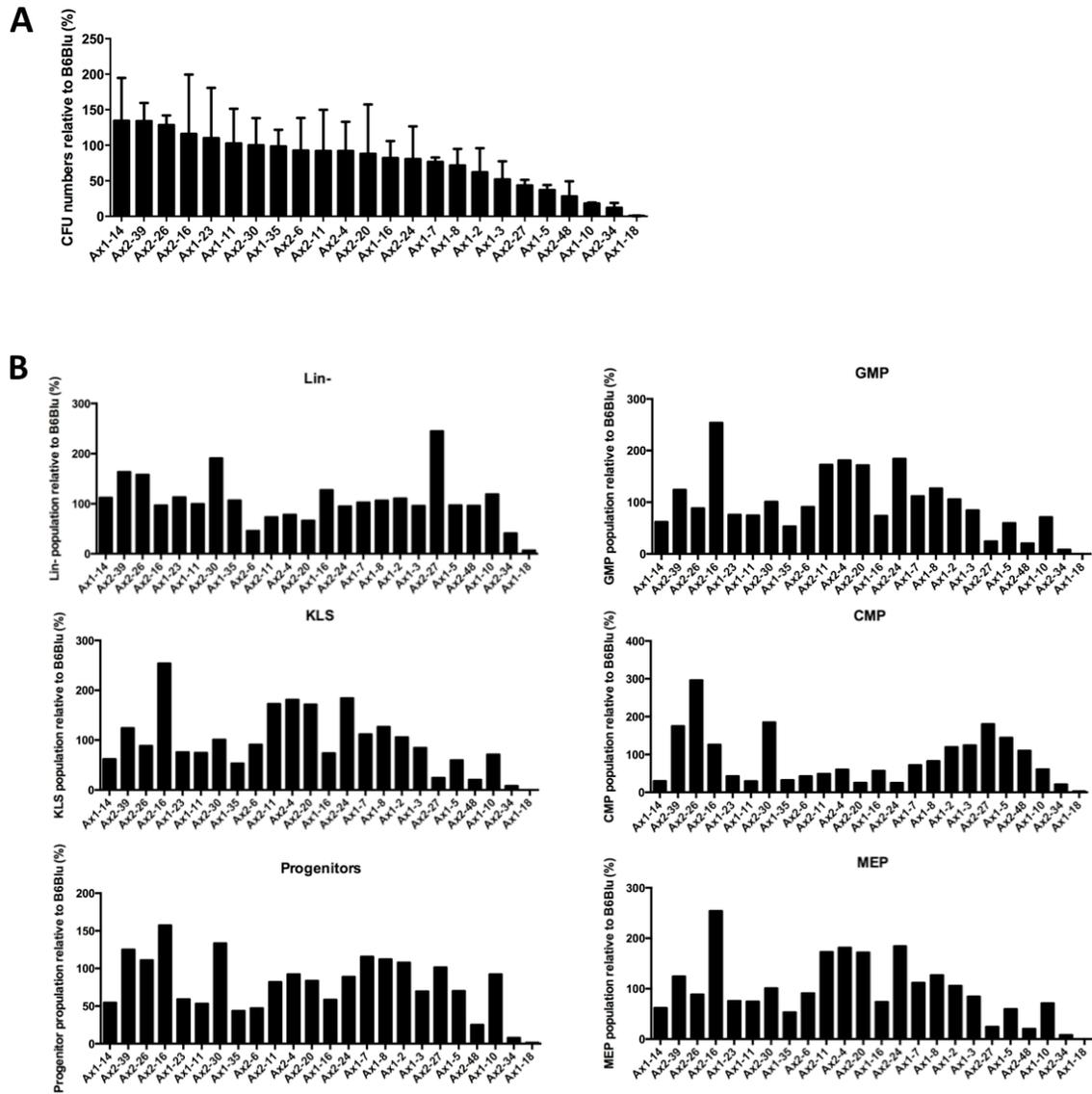
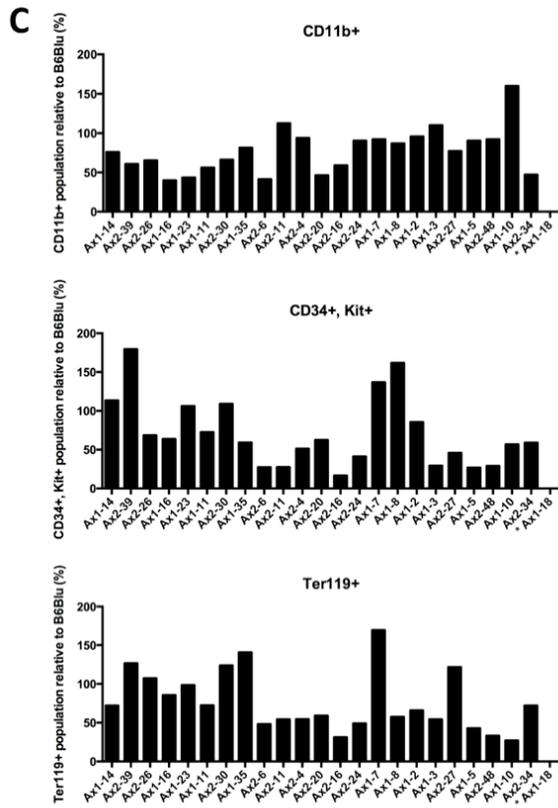


Figure 3-2. Hematopoietic potential of the 24 miPSC clones (Continued)



* Ax1-18 did not yield any colonies.

Figure 3-3. Mutation landscapes of the 24 miPSC clones

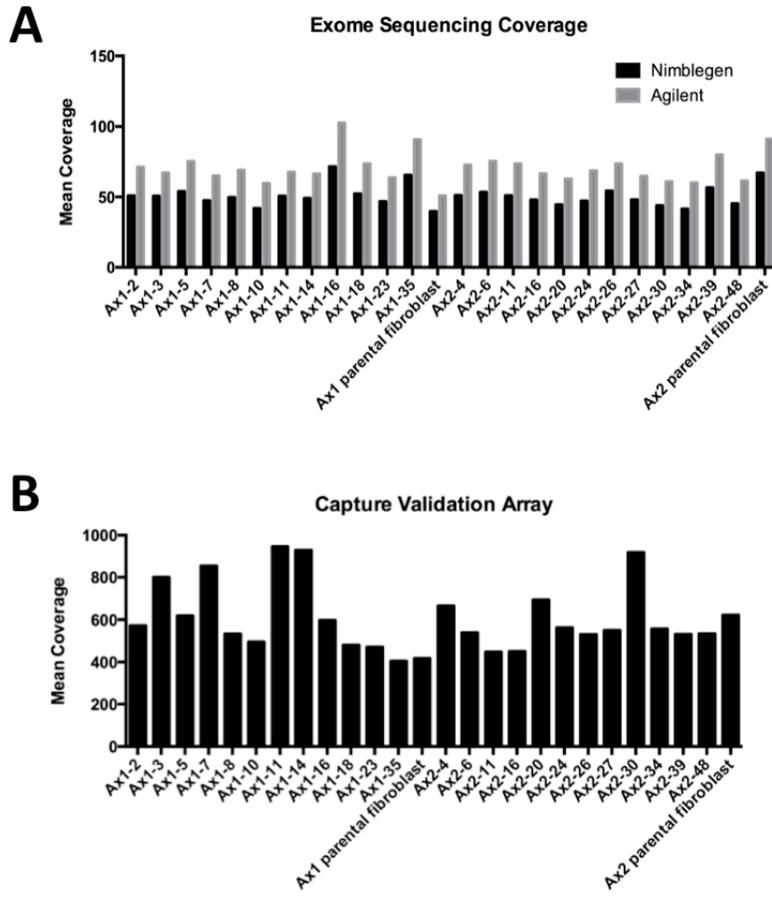


Figure 3-3. Mutation landscapes of the 24 miPSC clones (Continued)

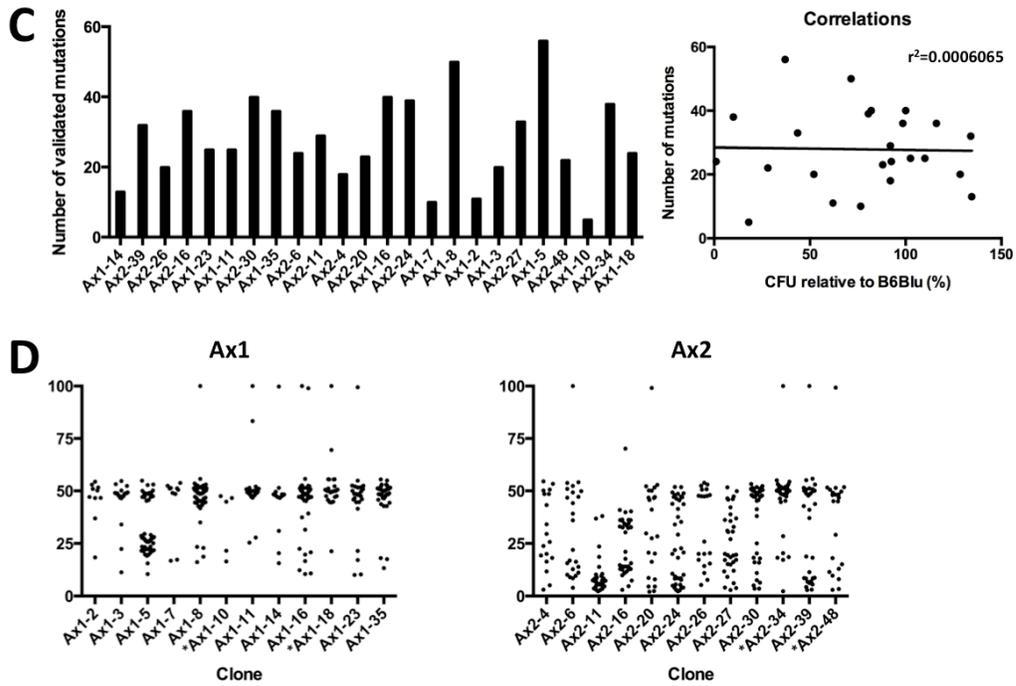


Figure 3-4. Expression array comparing six miPSC clones with normal vs low hematopoietic potential

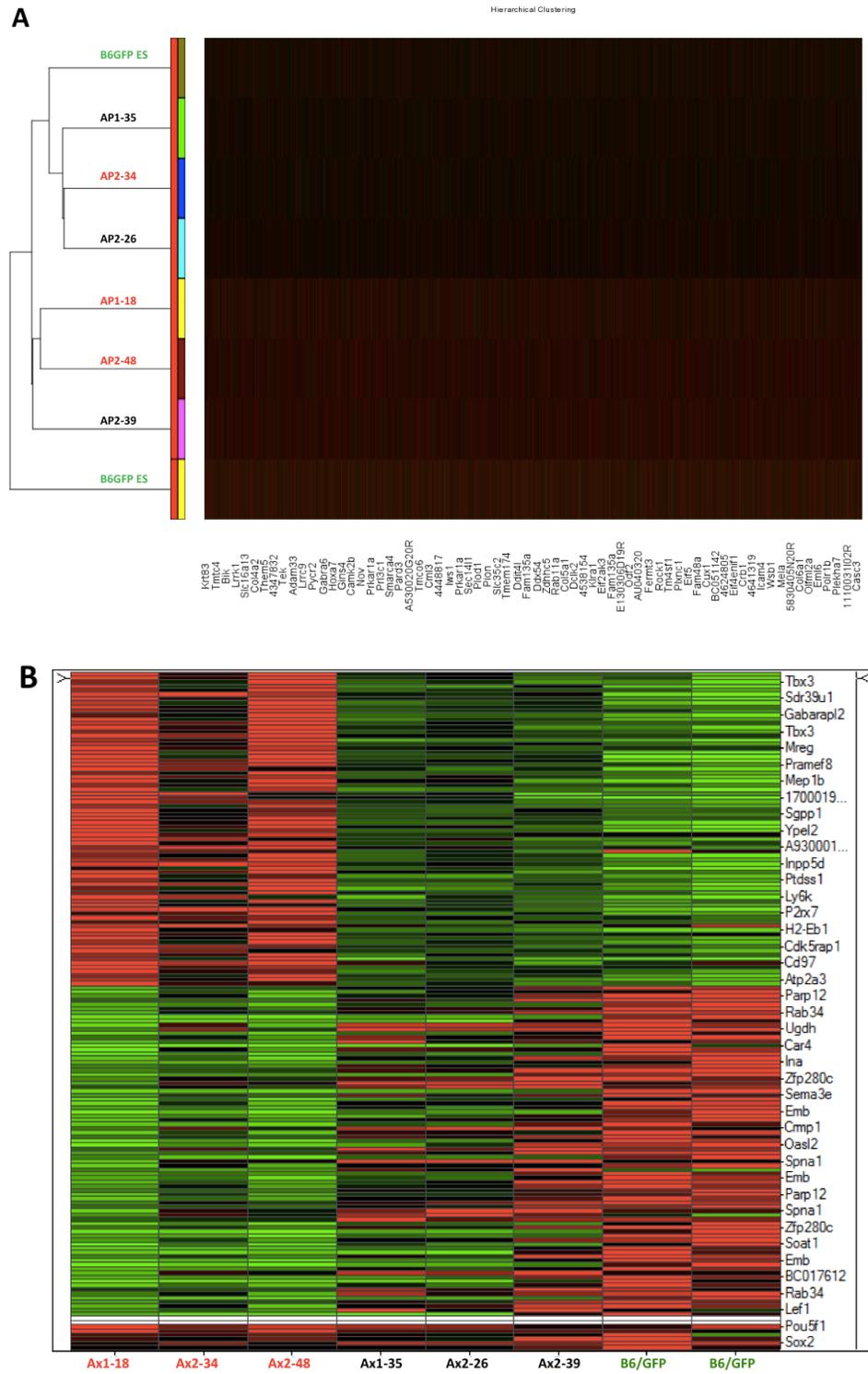


Figure 3-4. Expression array comparing six miPSC clones with normal vs low hematopoietic potential (Continued)

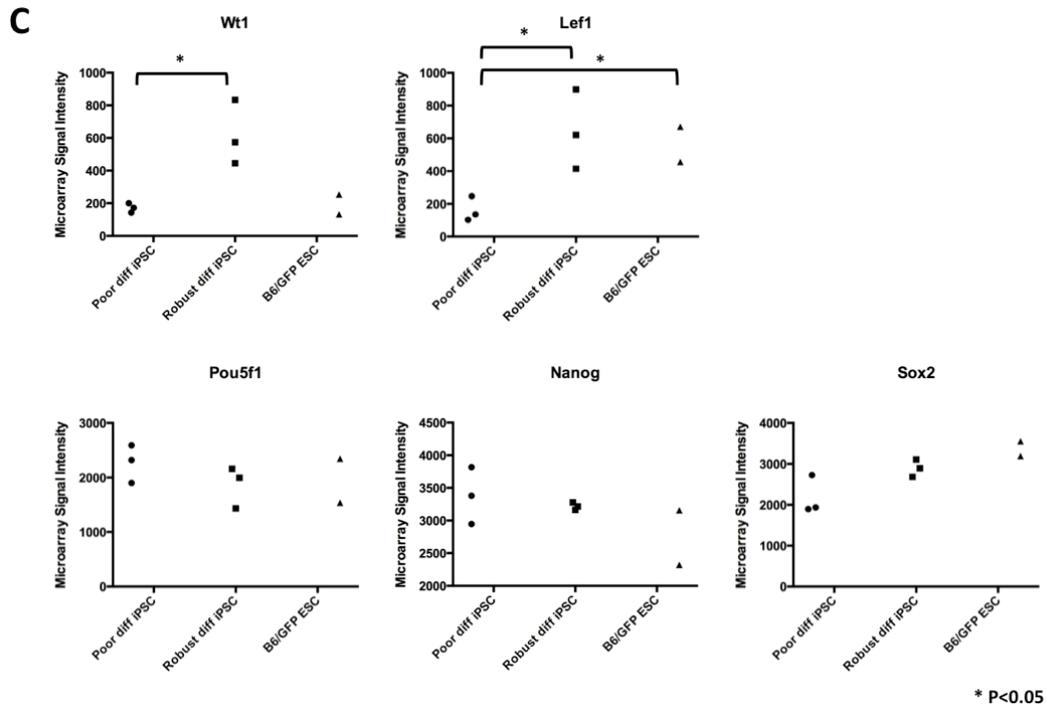


Table 3-1. Summary of the iPSC heterogeneity literature

Publication	Species	Lineage of functional heterogeneity	Maximum number of iPSC clones from one parental lines	Genetic heterogeneity	Epigenetic heterogeneity	Transcriptional heterogeneity	Proteomic heterogeneity
Boulting 2011 ¹²	Human	Neuron	3				
Hu 2010 ¹³	Human	Neuron	5				
Hussein 2011 ⁵⁴	Human		5	SNP array			
Laurent 2011 ⁶	Human		28	SNP array (parental sample NOT tested)			
Martins-Taylor 2011 ⁵⁵	Human		7	aCGH			
Ji 2012 ³²	Human		5	Exome Seq			
Gore 2011 ⁵	Human		2	Exome Seq			
Cheng 2012 ⁵⁶	Human		2	WGS			
Ruiz 2013 ³¹	Human		2	WGS			
Young 2012 ⁸	Mouse		4	WGS			
Mills 2013 ¹⁴	Human	Blood	3	SNP Karyotyping		Microarray	
Abyzov 2012 ⁷	Human		3	WGS		Microarray RNA-Seq	
Kajiwara 2012 ¹⁵	Human	Hepatocytes	5		Pyroseq	Microarray (2 clones)	
Bock 2010 ¹⁸	Human	EB	3		RRBS	Microarray	
Kim 2010 ¹⁰	Mouse	Blood Osteoblast	8		CHARM		
Kim 2011 ¹¹	Mouse		5		CHARM		
Lister 2011 ⁵⁷	Human		3		MethylC-Seq		
Phanstiel 2011 ⁵⁸	Human		4			Targeted bisulfide seq	Mass spec

Table 3-2. OP9 and FBS lot compatibility (Based on number of CFUs produced by B6 ESC-derived hematopoietic progenitors) with the hematopoietic differentiation assay

OP9 ATCC Lot#	FBS Brand & Lot#	Compatibility with hematopoietic differentiation assay
Lot # 58105522	FisherBrand – Research Grade Serum – Cat# 03-600-510 (Lot# FB-004)	✓
Lot # 58105522	Hyclone – FBS Characterized – Cat# SH30071 (Lot# AXF42326)	✓
Lot # 58105522	Hyclone – FBS Characterized – Cat # SH30071 (Lot# AXE41320)	✓
Lot # 58105522	Hyclone – FBS Characterized – Cat# SH30071 (Lot# AWC99942)	✓
Lot # 60484552	Hyclone – FBS Characterized – Cat# SH30071 (Lot# AWC99942)	✓
Lot # 60053041	Hyclone – FBS Characterized – Cat# SH30071 (Lot# AWC99942)	✓

Table 3-3. Whole exome sequencing coverage

Sample	NimbleGen Mean Coverage	NimbleGen Median Coverage	Agilent Mean Coverage	Agilent Median Coverage
Ax1-2	50.97	39.91	71.24	58.69
Ax1-3	50.72	37.85	67.02	55.19
Ax1-5	53.99	42.99	75.18	60.49
Ax1-7	47.38	35.97	65.11	53.14
Ax1-8	49.87	38.76	68.97	57.88
Ax1-10	42.04	31.79	59.70	48.90
Ax1-11	50.68	38.94	67.58	55.65
Ax1-14	49.18	37.68	66.41	55.68
Ax1-16	71.62	55.49	102.50	83.39
Ax1-18	52.35	40.72	73.54	59.80
Ax1-23	46.82	35.28	63.74	52.18
Ax1-35	65.58	50.83	90.71	74.75
Ax1 parental fibroblast	39.76	31.21	51.00	42.39
Ax2-4	51.18	40.33	72.68	60.48
Ax2-6	53.40	42.19	75.34	61.58
Ax2-11	51.03	39.96	73.56	59.24
Ax2-16	47.93	36.47	66.50	54.69
Ax2-20	44.55	33.72	62.85	51.43
Ax2-24	47.22	36.68	68.59	56.31
Ax2-26	54.30	42.27	73.54	62.44
Ax2-27	48.11	36.59	64.76	53.11
Ax2-30	43.88	33.61	60.87	49.81
Ax2-34	41.62	32.17	60.06	49.29
Ax2-39	56.74	44.10	80.00	64.48
Ax2-48	45.28	34.37	61.48	50.22
Ax2 parental fibroblast	67.05	53.26	91.18	75.36
Mean	50.89	39.35	70.54	57.95

Table 3-4. Validation array coverage

Sample	Mean Coverage	Median Coverage
Ax1-2	571.98	431.46
Ax1-3	800.73	606.67
Ax1-5	618.36	465.09
Ax1-7	853.38	652.16
Ax1-8	532.74	401.38
Ax1-10	495.42	371.31
Ax1-11	946.35	727.55
Ax1-14	928.85	703.41
Ax1-16	597.40	452.26
Ax1-18	480.24	348.80
Ax1-23	470.90	350.81
Ax1-35	404.18	303.20
Ax1 parental fibroblast	417.30	303.40
Ax2-4	665.04	500.78
Ax2-6	538.68	403.43
Ax2-11	447.50	335.88
Ax2-16	450.58	338.16
Ax2-20	694.01	532.81
Ax2-24	562.92	422.89
Ax2-26	530.64	396.95
Ax2-27	549.81	414.58
Ax2-30	916.81	695.59
Ax2-34	556.11	417.81
Ax2-39	531.33	396.36
Ax2-48	533.86	398.94
Ax2 parental fibroblast	621.84	454.89
Mean	604.50	454.87

Table 3-5. GFP and Oct3/4 expression of miPSC clones

Clone	GFP	Oct3/4
Ax1-2	96.29%	99.40%
Ax1-3	82.35%	94.00%
Ax1-5	86.86%	99.95%
Ax1-7	83.72%	99.70%
Ax1-8	88.81%	99.77%
Ax1-10	91.29%	91.00%
Ax1-11	76.23%	99.60%
Ax1-14	82.73%	99.48%
Ax1-16	74.90%	99.91%
Ax1-18	94.53%	99.91%
Ax1-23	93.77%	99.92%
Ax1-35	90.93%	99.96%
Ax2-4	79.28%	98.30%
Ax2-6	91.59%	98.96%
Ax2-11	88.64%	95.20%
Ax2-16	81.58%	98.20%
Ax2-20	82.96%	97.43%
Ax2-24	88.96%	98.10%
Ax2-26	41.31%	98.43%
Ax2-27	93.48%	98.86%
Ax2-30	95.32%	94.78%
Ax2-34	89.43%	98.94%
Ax2-39	70.65%	91.24%
Ax2-48	91.72%	97.31%

Table 3-6. Private mutations in all 24 miPSC clones

Clone	Chromosome name	start	stop	reference	variant	type	gene_name	Trv type	Amino acid change	Normal ref count	Normal var count	Normal VAF	Tumor ref count	Tumor var count	Tumor VAF
Ax1-2	1	75143105	75143105	A	C	SNP	Fam134a	missense	p.S368R	366	0	0	249	259	50.98
Ax1-2	1	95921940	95921940	T	G	SNP	Neu4	missense	p.L462V	357	0	0	221	263	54.34
Ax1-2	2	20583621	20583621	A	G	SNP	Etl4	missense	p.M202V	317	1	0.31	259	226	46.6
Ax1-2	2	40512974	40512974	A	G	SNP	Lrp1b	missense	p.C4239R	366	0	0	196	172	46.74
Ax1-2	3	65861204	65861204	C	T	SNP	Vep1	silent	p.V775	294	0	0	304	178	36.93
Ax1-2	5	35440918	35440918	T	G	SNP	Lrpap1	missense	p.K162T	355	0	0	237	239	50.21
Ax1-2	10	51795649	51795649	C	A	SNP	Ros1	silent	p.V1963	360	0	0	228	243	51.59
Ax1-2	14	75623932	75623932	C	T	SNP	Lcp1	silent	p.D519	324	0	0	214	241	52.97
Ax1-2	16	22987658	22987658	T	C	SNP	Kng2	missense	p.Q415R	402	0	0	316	282	47.16
Ax1-3	1	45914261	45914261	T	C	SNP	EG433297	silent	p.E119	258	0	0	220	193	46.73
Ax1-3	1	89316937	89316937	T	C	SNP	Gigyf2	missense	p.L660P	337	0	0	433	393	47.58
Ax1-3	1	137949518	137949518	A	C	SNP	Cacna1s	missense	p.K14T	329	0	0	357	341	48.85
Ax1-3	2	89033849	89033849	A	G	SNP	Olf1226	missense	p.L114P	367	0	0	296	286	48.97
Ax1-3	2	158206753	158206753	A	G	SNP	Shhg11	rna	NULL	334	0	0	345	313	47.57
Ax1-3	3	81912539	81912539	T	A	SNP	Gucy1a3	missense	p.N355Y	315	0	0	241	291	54.7
Ax1-3	3	126645598	126645598	A	G	SNP	Ank2	missense	p.I131T	335	0	0	281	284	50.18
Ax1-3	4	141313330	141313330	T	C	SNP	Agmat	silent	p.C256	326	0	0	377	360	48.85
Ax1-3	5	136577537	136577537	A	G	SNP	Rasa4	missense	p.K349R	291	0	0	333	290	46.55
Ax1-3	6	3658515	3658515	G	C	SNP	Calcr	missense	p.P224A	332	0	0	326	301	48.01
Ax1-3	8	107608921	107608921	A	C	SNP	Gm4738	silent	p.T53	352	0	0	324	291	47.32
Ax1-3	9	39278118	39278118	G	T	SNP	Olf955	missense	p.S64R	345	0	0	489	62	11.25
Ax1-3	11	35499816	35499816	C	T	SNP	Slit3	nonsense	p.Q1125*	287	0	0	418	216	34.07
Ax1-3	12	31624444	31624444	C	T	SNP	Sh3yl1	silent	p.S99	384	0	0	307	338	52.4
Ax1-3	13	100989769	100989769	A	G	SNP	Naip5	silent	p.L1098	249	0	0	278	267	48.9
Ax1-3	14	30916854	30916854	A	C	SNP	Cacna1d	missense	p.L1023R	306	0	0	307	298	49.26
Ax1-3	14	38535895	38535895	A	T	SNP	LOC382871	rna	NULL	300	0	0	317	299	48.54
Ax1-3	14	51979250	51979250	T	C	SNP	LOC633238	rna	NULL	314	0	0	275	266	49.08
Ax1-3	14	55566949	55566949	A	G	SNP	Myh6	missense	p.L1540P	275	0	0	296	337	53.24
Ax1-5	1	36894883	36894883	A	G	SNP	Tmem131	missense	p.V240A	337	0	0	300	267	47.01
Ax1-5	1	125273216	125273216	C	A	SNP	Dpp10	missense	p.D508Y	395	0	0	312	87	21.8
Ax1-5	1	180179510	180179510	G	T	SNP	Pppde1	nonsense	p.E166*	323	0	0	346	99	22.25
Ax1-5	2	32812773	32812773	T	A	SNP	Lrsam1	missense	p.K57I	318	0	0	361	148	29.08
Ax1-5	2	32845128	32845128	T	G	SNP	Garnl3	missense	p.K902T	338	0	0	391	144	26.92
Ax1-5	2	36924294	36924294	T	C	SNP	Olf360	silent	p.S156	234	0	0	189	176	48.22
Ax1-5	2	131066205	131066205	T	G	SNP	Mavs	missense	p.V91G	285	0	0	355	100	21.93
Ax1-5	3	22088832	22088832	C	A	SNP	Tbl1xr1	missense	p.A116D	283	0	0	503	185	26.81
Ax1-5	3	90063506	90063506	T	G	SNP	Crtc2	silent	p.G204	373	0	0	313	353	53

Ax1-5	3	97950412	97950412	A	G	SNP	Notch2	missense	p.E2156G	256	0	0	294	90	23.26
Ax1-5	4	63390783	63390783	T	G	SNP	Tnfsf15	missense	p.K200T	301	0	0	267	221	45.29
Ax1-5	4	107576929	107576929	A	T	SNP	Cpt2	missense	p.S629T	331	0	0	404	154	27.6
Ax1-5	4	120550586	120550586	G	T	SNP	Rims3	missense	p.G38V	314	0	0	339	142	29.52
Ax1-5	4	120726886	120726886	A	G	SNP	Col9a2	silent	p.G606	354	1	0.28	419	159	27.51
Ax1-5	4	122868094	122868094	A	C	SNP	Hpcal4	missense	p.K191T	271	0	0	332	117	26.06
Ax1-5	4	144212771	144212771	T	G	SNP	Gm13177	missense	p.L232V	321	0	0	247	234	48.65
Ax1-5	5	14978170	14978170	A	C	SNP	Speer8-ps1	rna	NULL	23	1	4.17	37	7	15.56
Ax1-5	5	74398750	74398750	T	C	SNP	Usp46	missense	p.S320G	365	0	0	283	252	47.1
Ax1-5	5	135589541	135589541	A	G	SNP	Mlxipl	missense	p.K113R	336	0	0	266	257	49.14
Ax1-5	6	67423914	67423914	T	C	SNP	Il23r	missense	p.S177G	272	0	0	386	135	25.91
Ax1-5	6	113397633	113397633	A	T	SNP	Jagn1	missense	p.Q157H	352	0	0	283	260	47.79
Ax1-5	7	114009125	114009125	T	A	SNP	Olf704	silent	p.A210	243	0	0	243	93	27.6
Ax1-5	7	127370926	127370926	A	G	SNP	Abca14	missense	p.Q427R	300	0	0	259	261	50.19
Ax1-5	8	3744088	3744088	T	C	SNP	Cd209a	missense	p.K238R	382	0	0	360	303	45.7
Ax1-5	8	15098375	15098375	T	G	SNP	Myom2	missense	p.V517G	346	0	0	339	88	20.56
Ax1-5	8	19682090	19682090	C	T	SNP	LOC624198	rna	NULL	66	10	13.16	157	43	21.5
Ax1-5	8	37166207	37166207	T	C	SNP	D8Erd82e	missense	p.S102P	235	5	2.07	262	91	25.78
Ax1-5	8	72388376	72388376	A	T	SNP	Pbx4	missense	p.E128V	380	0	0	392	157	28.6
Ax1-5	8	107258769	107258769	T	G	SNP	Ces6	missense	p.S83R	287	0	0	306	83	21.34
Ax1-5	8	126847946	126847946	C	A	SNP	Galnt2	missense	p.D175E	338	0	0	464	54	10.42
Ax1-5	8	127943120	127943120	T	G	SNP	Sipa112	missense	p.E1698D	308	0	0	375	115	23.42
Ax1-5	10	93052411	93052411	G	A	SNP	ENSMUSG0000086419	rna	NULL	224	1	0.44	280	69	19.77
Ax1-5	10	100075141	100075141	A	C	SNP	1700017N19Rik	missense	p.N300T	352	1	0.28	389	119	23.43
Ax1-5	10	126608769	126608769	A	C	SNP	B4galnt1	missense	p.K494N	295	0	0	368	133	26.44
Ax1-5	11	48910061	48910061	T	C	SNP	Ifi47	silent	p.S384	284	0	0	369	144	28.07
Ax1-5	11	67025077	67025077	A	G	SNP	Myh1	silent	p.Q897	383	1	0.26	482	131	21.34
Ax1-5	11	82713043	82713043	T	G	SNP	Fndc8	missense	p.F279V	373	0	0	308	278	47.44
Ax1-5	11	98018697	98018697	T	C	SNP	Med1	silent	p.G862	356	0	0	546	190	25.82
Ax1-5	11	115656677	115656677	T	C	SNP	2310067B10Rik	silent	p.T1001	308	0	0	327	291	47.09
Ax1-5	12	117381450	117381450	T	G	SNP	Vipr2	missense	p.F376V	315	0	0	341	128	27.29
Ax1-5	13	27215817	27215817	A	C	SNP	Pr13d2	missense	p.T70P	374	0	0	342	131	27.7
Ax1-5	14	50654785	50654785	A	G	SNP	Olf725	silent	p.L98	297	0	0	271	257	48.67
Ax1-5	14	65639321	65639321	T	G	SNP	Ints9	missense	p.I371S	315	0	0	452	172	27.56
Ax1-5	14	121885934	121885934	G	A	SNP	Slc15a1	missense	p.R161W	310	0	0	235	263	52.81
Ax1-5	15	42924847	42924847	A	G	SNP	Rspo2	silent	p.S45	381	0	0	349	134	27.74
Ax1-5	16	22895290	22895290	T	G	SNP	Ahsg	missense	p.L145V	339	0	0	233	283	54.84
Ax1-5	16	32750339	32750339	A	C	SNP	Muc4	missense	p.T32P	340	0	0	271	262	49.06
Ax1-5	17	20493783	20493783	T	G	SNP	Vmn2r104	missense	p.F161V	350	0	0	255	245	49
Ax1-5	17	50746557	50746557	A	G	SNP	Plcl2	missense	p.K423R	419	0	0	353	128	26.61
Ax1-5	17	74887067	74887067	T	A	SNP	LOC624159	rna	NULL	153	0	0	203	68	25.09
Ax1-5	18	37562659	37562659	A	G	SNP	Pcdhb9	missense	p.E684G	292	0	0	322	112	25.81
Ax1-5	19	41668313	41668313	A	C	SNP	AI606181	rna	NULL	402	0	0	515	126	19.66
Ax1-7	1	99614019	99614019	C	A	SNP	Hisppd1	missense	p.D1096Y	325	0	0	321	339	51.21

Ax1-7	2	172298214	172298214	G	A	SNP	1700029 J11Rik	missense	p.G41E	308	0	0	354	363	50.56
Ax1-7	4	146092582	146092582	T	A	SNP	Gm1305 1	missense	p.N394K	26	1	3.7	91	19	17.27
Ax1-7	5	116120300	116120300	G	A	SNP	Ccdc64	silent	p.L261	328	10	2.96	377	398	51.35
Ax1-7	7	4095211	4095211	G	A	SNP	Leng8	missense	p.V407I	276	0	0	319	307	49.04
Ax1-7	7	97395651	97395651	C	A	SNP	Ccdc83	missense	p.K73N	263	0	0	309	341	52.46
Ax1-7	12	102173940	102173940	C	A	SNP	Ccdc88c	missense	p.E1216 D	197	1	0.51	271	316	53.83
Ax1-7	13	34980005	34980005	A	G	SNP	Prpf4b	silent	p.R475	427	9	2.06	395	372	48.5
Ax1-7	17	66159963	66159963	T	C	SNP	Ppp4r1	silent	p.A127	373	0	0	421	441	51.16
Ax1-8	1	24037271	24037271	A	G	SNP	Fam135a	missense	p.L452P	353	1	0.28	200	206	50.74
Ax1-8	1	53700343	53700343	A	C	SNP	Dnahc7	missense	p.L233V	335	1	0.3	199	207	50.99
Ax1-8	1	54604841	54604841	A	G	SNP	Pgap1	missense	p.L140P	305	0	0	236	195	45.14
Ax1-8	1	74829327	74829327	A	G	SNP	Wnt6	silent	p.A164	350	0	0	215	237	52.43
Ax1-8	1	91153255	91153255	T	A	SNP	LOC383 542	rna	NULL	163	0	0	113	111	49.55
Ax1-8	1	175146497	175146497	C	A	SNP	Olf1404	silent	p.A238	304	0	0	161	156	49.06
Ax1-8	1	175394081	175394081	A	G	SNP	Aim2	missense	p.K83R	421	0	0	272	285	51.08
Ax1-8	2	26285618	26285618	A	G	SNP	Sec16a	missense	p.S1443 P	372	0	0	287	238	45.33
Ax1-8	2	69509108	69509108	A	T	SNP	Kbtbd10	missense	p.E285D	366	0	0	238	256	51.82
Ax1-8	2	125566371	125566371	A	C	SNP	Secisbp2 1	missense	p.L967R	255	0	0	190	137	41.77
Ax1-8	2	174283335	174283335	A	G	SNP	Tubb1	silent	p.E436	390	0	0	230	258	52.76
Ax1-8	3	135946888	135946888	C	T	SNP	Bank1	nonsense	p.W135*	346	0	0	242	195	44.62
Ax1-8	5	77391513	77391513	A	G	SNP	Paics	missense	p.E250G	402	0	0	260	289	52.64
Ax1-8	5	88922200	88922200	A	T	SNP	Enam	missense	p.D157V	325	1	0.31	246	235	48.76
Ax1-8	7	31657839	31657839	A	G	SNP	Cd22	missense	p.W410R	332	0	0	217	228	51.24
Ax1-8	7	120704954	120704954	T	A	SNP	Far1	missense	p.S398T	352	0	0	263	297	53.04
Ax1-8	7	130412558	130412558	T	C	SNP	Slc5a11	silent	p.L486	311	0	0	240	249	50.82
Ax1-8	7	148938654	148938654	C	-	DEL	Muc2	frame_shif t_del	p.K744fs	244	0	0	230	182	42.72
Ax1-8	8	54598973	54598973	T	C	SNP	Aga	missense	p.C69R	447	0	0	183	230	55.69
Ax1-8	8	128194789	128194789	T	C	SNP	4933403 G14Rik	silent	p.T340	335	0	0	209	215	50.71
Ax1-8	9	18299972	18299972	T	G	SNP	Muc16	missense	p.K8587 T	385	0	0	307	94	23.38
Ax1-8	9	19942252	19942252	T	A	SNP	Olf1869	missense	p.S203T	334	0	0	201	212	51.33
Ax1-8	9	120376656	120376656	A	C	SNP	Myrip	missense	p.S838R	299	0	0	195	169	46.43
Ax1-8	10	70844544	70844544	T	G	SNP	lpmk	missense	p.L390R	278	0	0	195	156	44.44
Ax1-8	10	80252098	80252098	T	C	SNP	Dot1l	missense	p.S1145 P	254	1	0.39	205	195	48.63
Ax1-8	10	82049452	82049452	T	C	SNP	LOC544 716	rna	NULL	402	0	0	219	252	53.5
Ax1-8	10	116715724	116715724	T	C	SNP	Lyz2	missense	p.R119G	327	0	0	235	195	45.35
Ax1-8	10	127747505	127747505	G	A	SNP	Pan2	missense	p.V400I	256	0	0	226	172	43.22
Ax1-8	11	3605099	3605099	T	G	SNP	Osbp2	missense	p.K899T	247	0	0	189	151	44.41
Ax1-8	11	20140320	20140320	A	G	SNP	Cep68	missense	p.S232P	260	1	0.38	168	186	52.54
Ax1-8	11	68921969	68921969	T	C	SNP	Per1	missense	p.S1185 P	257	1	0.39	187	187	50
Ax1-8	11	80700114	80700114	T	G	SNP	Accn1	missense	p.E475D	340	0	0	205	177	46.34
Ax1-8	13	9165677	9165677	T	A	SNP	Larp4b	missense	p.L551M	377	0	0	234	206	46.82
Ax1-8	13	54626205	54626205	T	C	SNP	4732471 D19Rik	missense	p.L335P	352	0	0	203	218	51.78
Ax1-8	14	53047666	53047666	C	A	SNP	TRAV1	missense	p.L9I	335	0	0	229	229	50
Ax1-8	14	56199914	56199914	A	G	SNP	Psme1	missense	p.E203G	340	0	0	243	261	51.79
Ax1-8	15	12082565	12082565	A	C	SNP	Zfr	missense	p.K487T	384	0	0	295	283	48.96

Ax1-8	15	15873418	15873418	G	T	SNP	EG23934 1	missense	p.R140S	254	0	0	182	98	35
Ax1-8	15	34358306	34358306	T	C	SNP	Matn2	missense	p.V687A	365	0	0	221	223	50.23
Ax1-8	15	76128418	76128418	A	T	SNP	Oplah	missense	p.L1024 H	297	0	0	228	237	50.75
Ax1-8	17	14138279	14138279	T	C	SNP	EG66482 1	missense	p.D206G	286	0	0	177	171	49.14
Ax1-8	17	37020207	37020207	A	C	SNP	Trim40	missense	p.L180V	341	0	0	283	254	47.21
Ax1-8	18	20629494	20629494	A	C	SNP	Dsg4	missense	p.K839T	327	0	0	235	187	44.31
Ax1-8	18	37090945	37090945	A	C	SNP	Pcdha11	missense	p.K336T	336	0	0	239	228	48.61
Ax1-8	18	37834988	37834988	A	G	SNP	Pcdhga1 1	missense	p.K280R	351	0	0	249	259	50.98
Ax1-8	19	5884607	5884607	A	G	SNP	Slc25a45	missense	p.K201R	382	0	0	309	311	50.16
Ax1-8	X	121247120	121247120	A	G	SNP	Vmn2r12 1	missense	p.I141T	48	1	2.04	61	14	18.67
Ax1-8	X	163680837	163680837	A	G	SNP	Tlr8	missense	p.L983P	209	0	0	0	226	100
Ax1-10	1	65119819	65119819	A	C	SNP	Crygc	missense	p.F57V	158	0	0	92	75	44.91
Ax1-10	2	89770728	89770728	T	G	SNP	Olfir1258	missense	p.F254C	340	0	0	196	171	46.59
Ax1-10	5	148426814	148426814	G	T	SNP	Flt1	missense	p.N714K	278	0	0	168	152	47.5
Ax1-11	1	93202033	93202033	T	G	SNP	Scly	splice_site	e4+2	348	0	0	406	423	51.03
Ax1-11	1	162971897	162971897	A	G	SNP	Dars2	missense	p.S638P	303	0	0	397	376	48.64
Ax1-11	2	87372495	87372495	A	G	SNP	Olfir153	missense	p.I102V	285	0	0	359	327	47.6
Ax1-11	2	88586688	88586688	C	T	SNP	Olfir1198	missense	p.R119Q	366	0	0	477	162	25.35
Ax1-11	3	92233277	92233277	G	A	SNP	ENSMU SG00000 042165	silent	p.P13	180	0	0	203	198	49.38
Ax1-11	4	46561647	46561647	T	A	SNP	Coro2a	missense	p.D154V	353	0	0	431	415	49.05
Ax1-11	6	49326191	49326191	G	T	SNP	D330028 D13Rik	missense	p.A110S	302	0	0	375	388	50.79
Ax1-11	6	97245592	97245592	A	G	SNP	Frm4b	silent	p.S890	240	0	0	359	345	49.01
Ax1-11	6	130044628	130044628	C	T	SNP	Klra11	missense	p.V23M	373	0	0	311	312	50.08
Ax1-11	7	28876894	28876894	T	C	SNP	Fcgbp	missense	p.F854L	297	0	0	437	403	47.92
Ax1-11	7	30680882	30680882	T	G	SNP	Zfp27	missense	p.T226P	350	0	0	450	444	49.66
Ax1-11	8	126426510	126426510	A	G	SNP	Nup133	missense	p.L1107 P	290	0	0	458	455	49.84
Ax1-11	9	120475484	120475484	C	T	SNP	Entpd3	missense	p.P472L	302	0	0	362	335	48.06
Ax1-11	10	126535960	126535960	T	G	SNP	Os9	missense	p.N352T	366	1	0.27	477	484	50.31
Ax1-11	11	54184417	54184417	G	T	SNP	4930404 A10Rik	missense	p.L52F	404	0	0	453	454	50
Ax1-11	11	70813237	70813237	T	C	SNP	Dhx33	missense	p.K213E	222	0	0	313	313	50
Ax1-11	12	7996197	7996197	G	T	SNP	Apob	missense	p.A493S	364	0	0	347	369	51.46
Ax1-11	15	78502820	78502820	G	C	SNP	Elf2	missense	p.D652E	300	0	0	385	339	46.76
Ax1-11	17	56985536	56985536	G	T	SNP	Acsbg2	nonsense	p.S636*	304	0	0	418	426	50.47
Ax1-11	18	37100731	37100731	C	T	SNP	Pcdha11 LOC674 488	missense	p.S587L	370	0	0	447	399	47.16
Ax1-11	19	5880022	5880022	T	C	SNP	Npm3	missense	p.K170T	303	0	0	411	396	49.01
Ax1-11	X	86651213	86651213	-	A	INS	Pet2	frame_shif t_ins	p.Q217fs	130	2	1.52	37	185	83.33
Ax1-11	X	138159389	138159389	T	A	SNP	Irs4	missense	p.K118M	139	0	0	0	267	100
Ax1-14	2	180735200	180735200	C	T	SNP	Col20a1	missense	p.P733L	328	0	0	390	366	48.35
Ax1-14	4	141188401	141188401	T	G	SNP	Plek2	missense	p.K343N	275	1	0.36	406	371	47.75
Ax1-14	4	146553590	146553590	C	T	SNP	Gm1313 9	silent	p.C438	48	2	4	81	15	15.62
Ax1-14	6	67675361	67675361	T	A	SNP	ENSMU SG00000 076503	missense	p.V23E	368	0	0	373	333	47.03
Ax1-14	7	17743811	17743811	A	G	SNP	Ceacam 3	missense	p.T377A	308	0	0	403	377	48.33

Ax1-14	9	75303490	75303490	A	T	SNP	Leo1	missense	p.K462N	254	0	0	358	326	47.66
Ax1-14	10	75874905	75874905	T	A	SNP	Pcnt	silent	p.R933	282	0	0	423	397	48.36
Ax1-14	15	85649257	85649257	A	C	SNP	Pkdrej	silent	p.R969	368	0	0	425	420	49.65
Ax1-14	17	56996331	56996331	G	T	SNP	Acsbg2	missense	p.P234T	365	0	0	443	472	51.47
Ax1-14	18	37455036	37455036	T	C	SNP	Pcdhb2	missense	p.L136P	369	0	0	368	323	46.68
Ax1-14	X	45316398	45316398	T	C	SNP	Ocr1	silent	p.S849	179	0	0	1	408	99.76
Ax1-16	2	25360071	25360071	T	C	SNP	Fbxw5	missense	p.L463P	238	0	0	228	201	46.85
Ax1-16	2	111026889	111026889	A	C	SNP	LOC623301	rna	NULL	359	1	0.28	244	231	48.63
Ax1-16	3	35907712	35907712	A	T	SNP	ENSMU SG0000086392	rna	NULL	353	0	0	244	269	52.44
Ax1-16	3	40661800	40661800	A	C	SNP	ENSMU SG0000074621	missense	p.K54T	304	0	0	208	190	47.62
Ax1-16	3	85942467	85942467	A	C	SNP	Rps3a	missense	p.I204S	296	0	0	251	234	48.25
Ax1-16	3	94231900	94231900	T	G	SNP	Tdrkh	missense	p.L336R	272	0	0	210	217	50.82
Ax1-16	3	98684383	98684383	G	T	SNP	Hao2	silent	p.R218	223	0	0	185	234	55.71
Ax1-16	4	82629308	82629308	T	C	SNP	Frem1	silent	p.G962	298	0	0	238	236	49.79
Ax1-16	4	108304211	108304211	A	C	SNP	Cc2d1b	missense	p.Q746H	318	0	0	274	260	48.69
Ax1-16	4	143556040	143556040	T	C	SNP	ENSMU SG0000078510	missense	p.Q111R	315	2	0.63	185	196	51.44
Ax1-16	5	88757576	88757576	T	C	SNP	Prol1	missense	p.S228P	433	2	0.46	273	244	47.2
Ax1-16	7	110926421	110926421	A	C	SNP	Olf68	missense	p.L146R	383	0	0	245	274	52.59
Ax1-16	7	114734198	114734198	A	C	SNP	Olfml1	missense	p.E318D	293	0	0	275	270	49.54
Ax1-16	8	113565421	113565421	T	G	SNP	Aars	missense	p.F175V	324	0	0	314	280	47.14
Ax1-16	8	122157205	122157205	A	G	SNP	Kcng4	missense	p.F111L	291	1	0.34	228	249	52.2
Ax1-16	9	3032385	3032385	A	T	SNP	ENSMU SG0000061971	silent	p.L99	58	1	1.69	57	8	12.31
Ax1-16	9	56469055	56469055	G	A	SNP	Lingo1	missense	p.P19L	234	0	0	199	181	47.63
Ax1-16	9	79499815	79499815	C	T	SNP	Col12a1	nonsense	p.W1866*	315	0	0	356	71	16.63
Ax1-16	11	23319006	23319006	G	A	SNP	Usp34	nonsense	p.W1732*	340	0	0	263	229	46.54
Ax1-16	11	82258626	82258626	C	T	SNP	Tmem132e	missense	p.R919C	284	0	0	240	252	51.22
Ax1-16	11	94278210	94278210	T	C	SNP	Cacna1g	missense	p.K1732R	299	0	0	269	161	37.44
Ax1-16	11	117710694	117710694	G	C	SNP	Birc5	missense	p.W10C	270	0	0	281	235	45.45
Ax1-16	13	12336225	12336225	G	A	SNP	Mtr	silent	p.S235	371	0	0	414	102	19.73
Ax1-16	13	21703077	21703077	T	G	SNP	Olf1362	missense	p.S254R	283	0	0	216	192	47.06
Ax1-16	13	43264151	43264151	G	A	SNP	Tbc1d7	missense	p.A62V	300	0	0	210	203	49.15
Ax1-16	14	55577447	55577447	G	T	SNP	Myh6	missense	p.H577Q	362	0	0	289	272	48.48
Ax1-16	15	57654086	57654086	A	C	SNP	Zhx2	missense	p.K432T	266	0	0	212	226	51.48
Ax1-16	15	88653536	88653536	A	G	SNP	Crel2	missense	p.N209S	369	0	0	303	331	52.21
Ax1-16	16	23966264	23966264	G	T	SNP	Bcl6	missense	p.Q691K	356	0	0	278	180	39.3
Ax1-16	16	72989899	72989899	C	T	SNP	Robo1	silent	p.I851	281	0	0	212	214	50.23
Ax1-16	17	35482121	35482121	A	G	SNP	H2-Q2	missense	p.D302G	312	0	0	235	204	46.47
Ax1-16	17	35577019	35577019	A	C	SNP	H2-Q7	missense	p.K167T	628	0	0	871	252	22.44
Ax1-16	17	37222430	37222430	T	G	SNP	Olf90	missense	p.S227R	283	0	0	233	233	50
Ax1-16	17	76683820	76683820	A	C	SNP	LOC100043886	missense	p.F128V	411	0	0	252	208	45.22
Ax1-16	X	7515520	7515520	A	C	SNP	Hdac6	silent	p.P415	174	0	0	0	309	100
Ax1-16	X	143825460	143825460	A	G	SNP	Il13ra2	missense	p.F204L	160	1	0.62	2	192	98.97
Ax1-18	1	20191259	20191259	C	A	SNP	Pkhd1	missense	p.Q3050H	342	0	0	157	132	45.67

Ax1-18	1	87411143	87411143	A	C	SNP	LOC100039794	missense	p.E415D	196	0	0	92	74	44.58
Ax1-18	1	133560888	133560888	A	G	SNP	Ctse	missense	p.Q125R	332	0	0	168	209	55.44
Ax1-18	2	67352880	67352880	T	C	SNP	Xirp2	silent	p.S2469	337	0	0	177	175	49.72
Ax1-18	2	120128566	120128566	A	G	SNP	Pla2g4f	silent	p.T629	402	0	0	228	227	49.89
Ax1-18	4	115314478	115314478	A	T	SNP	Cyp4b1	missense	p.Y99N	271	0	0	194	173	47.14
Ax1-18	5	38938318	38938318	T	A	SNP	Wdr1	missense	p.K81M	332	0	0	176	184	51.11
Ax1-18	5	149114848	149114848	G	T	SNP	Mtus2	nonsense	p.E12*	330	0	0	183	148	44.71
Ax1-18	6	69676796	69676796	T	C	SNP	ENSMU SG00000 076563	missense	p.S40G	359	0	0	140	127	47.57
Ax1-18	7	11165910	11165910	G	T	SNP	Vmn1r69	missense	p.A81E	376	0	0	174	190	52.2
Ax1-18	7	11300297	11300297	C	T	SNP	Nlrp4b	silent	p.C359	283	0	0	140	114	44.88
Ax1-18	7	111970448	111970448	C	A	SNP	LOC233 637	missense	p.S280Y	321	0	0	149	187	55.49
Ax1-18	9	19724859	19724859	G	C	SNP	Olf77	missense	p.V69L	312	0	0	154	158	50.64
Ax1-18	9	107454037	107454037	G	A	SNP	Rassf1	missense	p.G37D	321	1	0.31	137	141	50.54
Ax1-18	10	78480380	78480380	T	A	SNP	Olf1351	missense	p.F104L	215	0	0	115	105	47.51
Ax1-18	11	84113937	84113937	A	T	SNP	Acaca	missense	p.Q1338 H	373	0	0	216	493	69.53
Ax1-18	12	116404150	116404150	G	T	SNP	ENSMU SG00000 076715	missense	p.L47M	368	1	0.27	132	132	50
Ax1-18	13	100991748	100991748	T	G	SNP	Naip5	silent	p.R978	318	1	0.31	152	148	49.33
Ax1-18	15	84972004	84972004	C	T	SNP	Ribc2	silent	p.D182	271	0	0	0	383	100
Ax1-18	16	17257523	17257523	C	G	SNP	Hic2	missense	p.A41G	312	0	0	193	192	49.87
Ax1-18	17	29320950	29320950	C	G	SNP	Cpne5	missense	p.E219Q	388	0	0	193	242	55.5
Ax1-18	18	65408553	65408553	T	G	SNP	ENSMU SG00000 065402	rna	NULL	354	0	0	161	201	55.52
Ax1-18	19	47590702	47590702	T	A	SNP	Obfc1	nonsense	p.K164*	319	0	0	190	190	50
Ax1-23	1	57942740	57942740	A	T	SNP	Spats2l	missense	p.E131V	421	0	0	190	180	48.65
Ax1-23	1	135255968	135255968	G	T	SNP	Ren1	missense	p.G359C	314	0	0	187	201	51.67
Ax1-23	2	76609401	76609401	G	A	SNP	Ttn	silent	p.F16879	362	0	0	231	214	48.09
Ax1-23	3	96518016	96518016	A	G	SNP	Polr3c	splice_site	e12+2	334	0	0	254	205	44.66
Ax1-23	4	3186316	3186317	-	C	INS	LOC100 039044	rna	NULL	226	0	0	223	46	17.1
Ax1-23	5	81330151	81330151	A	C	SNP	LOC100 039384	rna	NULL	375	0	0	124	151	54.91
Ax1-23	6	65651546	65651546	A	C	SNP	A930038 C07Rik	missense	p.S100R	369	0	0	191	167	46.52
Ax1-23	7	27342992	27342992	A	C	SNP	Nlrp9a	missense	p.I339L	339	0	0	197	187	48.57
Ax1-23	9	20734598	20734598	A	C	SNP	Dnmt1	missense	p.S279R	365	0	0	206	213	50.84
Ax1-23	9	56904547	56904547	T	G	SNP	Ptpn9	missense	p.F386V	379	0	0	233	200	46.08
Ax1-23	10	20737728	20737728	A	C	SNP	Ahi1	missense	p.N851H	259	0	0	192	155	44.67
Ax1-23	10	39472453	39472453	T	C	SNP	EG66630 4	rna	NULL	252	0	0	179	127	41.5
Ax1-23	10	50473896	50473896	T	G	SNP	Ascc3	silent	p.T1762	377	0	0	197	167	45.88
Ax1-23	10	62285496	62285496	T	G	SNP	ENSMU SG00000 047146	missense	p.S90R	289	0	0	173	156	47.42
Ax1-23	10	102007890	102007890	G	T	SNP	Rassf9	silent	p.R166	324	0	0	154	171	52.45
Ax1-23	10	128687916	128687916	G	T	SNP	Olf1518	nonsense	p.E109*	301	0	0	148	164	52.56
Ax1-23	11	22920255	22920255	A	C	SNP	Fam161a	missense	p.K117N	373	0	0	223	221	49.77
Ax1-23	14	20964303	20964303	A	G	SNP	Kcnk5	missense	p.F114L	267	1	0.37	181	193	51.6
Ax1-23	17	34027423	34027423	C	A	SNP	LOC100 042970	missense	p.Q145H	171	0	0	183	50	21.46
Ax1-23	17	35279510	35279510	T	C	SNP	Bat3	silent	p.T512	264	0	0	169	169	50

Ax1-23	19	25235615	25235615	G	A	SNP	Dock8	silent	p.L1271	335	0	0	182	172	48.59
Ax1-23	X	40068274	40068274	G	A	SNP	Odz1	silent	p.C816	177	0	0	0	184	99.46
Ax1-35	2	12321721	12321721	A	G	SNP	Fam188a	silent	p.L168	387	0	0	172	175	50.43
Ax1-35	4	53055238	53055238	A	T	SNP	Abca1	missense	p.I1902N	334	0	0	161	181	52.92
Ax1-35	4	108180887	108180887	A	C	SNP	Zcchc11	missense	p.K586T	330	0	0	237	208	46.74
Ax1-35	4	130508754	130508754	A	C	SNP	Matn1	missense	p.K409T	325	0	0	166	140	45.75
Ax1-35	4	136574150	136574150	T	G	SNP	Zbtb40	missense	p.E162D	341	0	0	145	153	51.17
Ax1-35	5	15219883	15219883	C	G	SNP	Speer4c	missense	p.W46C	52	2	3.7	26	4	13.33
Ax1-35	5	120132643	120132643	T	A	SNP	Tbx3	missense	p.L590M	260	0	0	127	95	42.79
Ax1-35	6	34454428	34454428	A	T	SNP	Bpgm	missense	p.K253N	380	0	0	162	145	47.23
Ax1-35	7	39232345	39232345	T	A	SNP	EG62559 4	silent	p.P624	724	0	0	464	102	18.02
Ax1-35	7	49224724	49224724	T	G	SNP	Gprc2a- rs1	missense	p.K166T	351	0	0	211	195	48.03
Ax1-35	7	80608620	80608620	G	C	SNP	Chd2	missense	p.S1166 C	296	0	0	147	152	50.84
Ax1-35	7	111836866	111836866	C	A	SNP	Olf661	nonsense	p.C112*	337	0	0	171	173	50.29
Ax1-35	7	148179245	148179245	T	A	SNP	LOC665 143	rna	NULL	245	0	0	135	152	52.96
Ax1-35	8	16127148	16127148	T	C	SNP	Csmd1	missense	p.Q1367 R	373	0	0	153	158	50.8
Ax1-35	8	95898445	95898445	T	C	SNP	2310039 D24Rik	silent	p.L40	313	0	0	127	158	55.44
Ax1-35	9	38662724	38662724	C	A	SNP	Olf26	missense	p.P26T	326	0	0	160	149	48.22
Ax1-35	9	44196902	44196902	A	T	SNP	Hyou1	nonsense	p.R802*	378	0	0	191	199	51.03
Ax1-35	9	106556450	106556450	C	T	SNP	Grm2	missense	p.R57H	312	0	0	98	109	52.66
Ax1-35	9	107848526	107848526	A	T	SNP	Camkv	missense	p.K140M	340	0	0	142	154	52.03
Ax1-35	9	109396055	109396055	A	G	SNP	Fbxw19	silent	p.A120	340	0	0	157	122	43.73
Ax1-35	10	13610681	13610681	T	G	SNP	EG66685 9	rna	NULL	148	0	0	76	66	46.48
Ax1-35	10	79289955	79289955	A	T	SNP	9130017 N09Rik	missense	p.E487V	324	0	0	214	160	42.78
Ax1-35	11	69713019	69713019	C	A	SNP	2810408 A11Rik	missense	p.R209M	322	0	0	210	201	48.91
Ax1-35	11	87540958	87540958	C	A	SNP	Rnf43	missense	p.P64T	323	0	0	140	151	51.54
Ax1-35	11	99797964	99797964	T	G	SNP	Krtap31- 2	missense	p.C103G	96	0	0	48	45	48.39
Ax1-35	12	110956969	110956969	A	C	SNP	ENSMU SG00000 070105	rna	NULL	353	1	0.28	145	136	48.4
Ax1-35	13	22474628	22474628	C	T	SNP	Vmn1r19 9	silent	p.F74	376	0	0	149	183	54.95
Ax1-35	13	33924701	33924701	A	G	SNP	Serpnb6 e	missense	p.L254P	373	0	0	197	157	44.35
Ax1-35	13	113891125	113891125	A	G	SNP	Gzma	missense	p.L12P	307	0	0	191	179	48.38
Ax1-35	15	45811374	45811374	A	G	SNP	EG43295 1	silent	p.I70	360	0	0	130	129	49.81
Ax1-35	15	101862298	101862298	A	C	SNP	Krt18	missense	p.N393T	330	0	0	174	161	48.06
Ax1-35	17	47537242	47537242	A	T	SNP	Guca1a	missense	p.F43I	395	0	0	211	212	50.12
Ax1-35	17	84217761	84217761	A	G	SNP	Mta3	missense	p.S549G	332	0	0	186	175	48.48
Ax1-35	18	42146203	42146203	T	A	SNP	Grxcr2	missense	p.N242I	379	0	0	166	172	50.89
Ax2-4	2	76561769	76561769	T	G	SNP	Ttn	missense	p.K2913 6N	532	0	0	445	87	16.35
Ax2-4	2	144190261	144190261	C	T	SNP	LOC635 097	rna	NULL	87	1	1.14	52	49	48.51
Ax2-4	4	96436355	96436355	T	G	SNP	Gm1269 5	missense	p.Q89H	380	0	0	189	227	54.57
Ax2-4	4	108491783	108491783	A	G	SNP	Btf3l4	silent	p.L110	393	0	0	399	167	29.51
Ax2-4	4	143748122	143748122	C	T	SNP	Oog3	silent	p.G330	250	0	0	141	133	48.54
Ax2-4	5	10949006	10949006	C	A	SNP	LOC667	rna	NULL	373	0	0	156	160	50.63

Ax2-11	11	56272886	56272886	T	G	SNP	LOC624168	silent	p.T27	435	0	0	232	21	8.3
Ax2-11	11	70805689	70805689	T	G	SNP	Dhx33	missense	p.K541T	583	0	0	378	34	8.25
Ax2-11	12	54241494	54241494	C	A	SNP	Akap6	missense	p.S1568Y	470	0	0	330	24	6.78
Ax2-11	14	27419209	27419209	T	C	SNP	Fam116a	silent	p.L140	531	1	0.19	384	31	7.47
Ax2-11	14	45593979	45593979	A	C	SNP	ENSMU SG00000 085500	rna	NULL	557	0	0	349	17	4.64
Ax2-11	14	52074660	52074660	T	G	SNP	Vmn2r89	missense	p.F82V	508	0	0	371	29	7.25
Ax2-11	14	113714508	113714508	A	C	SNP	ENSMU SG00000 058126	missense	p.E204A	472	0	0	246	16	6.11
Ax2-11	17	15059264	15059264	C	T	SNP	Wdr27	splice_site	e7+1	496	0	0	236	145	38.06
Ax2-11	17	35562716	35562716	G	T	SNP	H2-Q6	missense	p.A179S	1004	1	0.1	618	22	3.43
Ax2-11	17	79172997	79172997	G	A	SNP	Heatr5b	missense	p.H1389Y	571	0	0	347	20	5.45
Ax2-11	18	28593311	28593311	G	T	SNP	EG62599 7	rna	NULL	276	11	3.83	144	17	10.56
Ax2-11	18	36908229	36908229	A	G	SNP	lk	silent	p.Q87	642	0	0	427	33	7.17
Ax2-11	18	42805516	42805516	T	C	SNP	Ppp2r2b	missense	p.R403G	494	0	0	313	33	9.54
Ax2-11	19	45080226	45080226	A	G	SNP	Mrpl43	missense	p.L148P	610	0	0	396	26	6.16
Ax2-11	19	47090218	47090218	A	G	SNP	Ina	missense	p.E325G	599	1	0.17	360	35	8.86
Ax2-16	1	10398403	10398403	T	G	SNP	Cpa6	missense	p.K263T	550	0	0	334	44	11.64
Ax2-16	1	22394576	22394576	C	A	SNP	Rims1	missense	p.R833I	571	0	0	261	129	33.08
Ax2-16	1	102110283	102110283	T	G	SNP	Cntnap5 b	silent	p.T562	537	0	0	284	31	9.84
Ax2-16	1	133496834	133496834	T	C	SNP	Avpr1b	missense	p.S173P	433	0	0	213	112	34.46
Ax2-16	2	64706352	64706352	G	T	SNP	LOC667 291	rna	NULL	424	0	0	230	28	10.85
Ax2-16	2	140226636	140226636	G	T	SNP	MacroD2	nonsense	p.E29*	568	0	0	253	128	33.6
Ax2-16	2	158210426	158210426	C	A	SNP	Shhg11	rna	NULL	552	0	0	260	148	36.1
Ax2-16	3	88787474	88787474	G	T	SNP	Ash1l	missense	p.A913S	404	0	0	327	185	36.13
Ax2-16	3	132989610	132989610	A	G	SNP	Ppa2	splice_site	e4-2	479	0	0	372	55	12.88
Ax2-16	4	88322198	88322198	T	C	SNP	Gm1328 0	missense	p.E188G	228	0	0	117	79	40.31
Ax2-16	5	13791671	13791671	T	C	SNP	Speer3	missense	p.L5P	373	0	0	237	34	12.55
Ax2-16	5	62997324	62997324	T	C	SNP	Arap2	missense	p.K1572R	524	0	0	178	101	36.2
Ax2-16	5	105757568	105757568	A	C	SNP	Gbp11	missense	p.C231W	518	0	0	279	48	14.68
Ax2-16	5	110105330	110105330	A	T	SNP	ENSMU SG00000 072763	silent	p.A424	477	0	0	349	54	13.4
Ax2-16	6	50166388	50166388	A	C	SNP	Dfna5	silent	p.T409	564	0	0	313	25	7.4
Ax2-16	8	13388812	13388812	T	C	SNP	Atp4b	missense	p.K188R	551	0	0	399	65	14.01
Ax2-16	8	19682091	19682091	C	T	SNP	LOC624 198	rna	NULL	146	11	7.01	64	14	17.72
Ax2-16	8	93557557	93557557	T	G	SNP	Chd9	silent	p.T2010	416	0	0	245	137	35.86
Ax2-16	8	119567182	119567182	C	A	SNP	Pkd112	missense	p.S1145I	530	0	0	266	128	32.41
Ax2-16	9	78026216	78026216	C	T	SNP	ENSMU SG00000 087566	rna	NULL	372	5	1.32	264	13	4.69
Ax2-16	11	45794573	45794573	A	T	SNP	Sox30	missense	p.D251V	465	0	0	333	62	15.62
Ax2-16	12	85760952	85760952	A	C	SNP	2900006 K08Rik	missense	p.K181T	553	0	0	366	54	12.86
Ax2-16	13	49326564	49326564	T	C	SNP	Susd3	missense	p.N251D	571	0	0	285	150	34.48
Ax2-16	14	31708632	31708632	A	G	SNP	Itih4	silent	p.Q574	455	0	0	353	51	12.59
Ax2-16	15	10908392	10908392	T	C	SNP	C1qtnf3	missense	p.S235P	506	1	0.2	232	154	39.9
Ax2-16	17	6033826	6033826	A	G	SNP	Synj2	silent	p.R1069	492	0	0	271	118	30.33

Ax2-16	17	25768682	25768682	C	A	SNP	Lmf1	missense	p.A219D	390	0	0	249	121	32.61
Ax2-16	17	34194363	34194363	G	T	SNP	Col11a2	missense	p.G815W	353	0	0	183	89	32.72
Ax2-16	18	37129340	37129340	G	A	SNP	Pcdha6	missense	p.R644H	449	0	0	286	47	14.07
Ax2-16	19	6931605	6931605	A	G	SNP	Ccdc88b	missense	p.L148P	367	0	0	189	131	40.94
Ax2-16	X	148654373	148654373	T	A	SNP	Igsec2	missense	p.F867I	330	0	0	62	146	70.19
Ax2-20	1	90131582	90131582	A	T	SNP	ENSMU SG00000 079429	missense	p.N330I	2125	0	0	1874	162	7.95
Ax2-20	2	5955219	5955220	-	G A A	INS	Upf2	in_frame_i ns	p.1034in _frame_i nsE	583	0	0	623	236	27.47
Ax2-20	2	78651820	78651820	C	A	SNP	LOC100 040494	rna	NULL	413	0	0	312	79	20.2
Ax2-20	3	54598373	54598373	T	C	SNP	Smad9	missense	p.L371P	466	0	0	309	271	46.64
Ax2-20	5	94862957	94862957	A	C	SNP	LOC665 521	rna	NULL	117	0	0	63	25	28.41
Ax2-20	5	148113556	148113556	G	T	SNP	Cdx2	silent	p.S266	433	0	0	248	265	51.66
Ax2-20	7	104420355	104420355	T	A	SNP	Gab2	silent	p.S142	388	1	0.26	231	252	52.17
Ax2-20	7	138080091	138080091	A	G	SNP	Htra1	silent	p.R102	401	0	0	244	245	50.1
Ax2-20	8	90863314	90863314	C	T	SNP	Brd7	silent	p.L479	546	0	0	573	12	2.05
Ax2-20	9	72741677	72741677	C	G	SNP	Prtg	missense	p.D827E	419	0	0	322	287	47.13
Ax2-20	12	45412313	45412313	A	C	SNP	Pnp1a8	missense	p.E686D	570	0	0	282	284	50.18
Ax2-20	13	22179937	22179937	A	G	SNP	Vmn1r18 8	missense	p.K64R	452	0	0	252	213	45.71
Ax2-20	16	57146236	57146236	G	C	SNP	Tom70 a	missense	p.E480Q	568	0	0	336	378	52.87
Ax2-20	17	18210694	18210694	C	A	SNP	EG63702 1	nonsense	p.S637*	323	0	0	285	27	8.63
Ax2-20	17	28212319	28212319	G	A	SNP	4930526 A20Rik	rna	NULL	508	0	0	364	320	46.78
Ax2-20	17	33786654	33786657	T C G G	-	DEL	Hnrmpm	frame_shif t_del	p.D619fs	423	0	0	282	197	41.13
Ax2-20	X	6562670	6562670	T	C	SNP	Ccnb3	missense	p.T1284 A	306	0	0	2	233	99.15
Ax2-20	X	166426921	166426921	C	T	SNP	Mid1	silent	p.P607	1592	0	0	2528	120	4.53
Ax2-20	X	166427034	166427034	C	T	SNP	Mid1	missense	p.A645V p.F1427 V	1599	2	0.12	2402	115	4.57
Ax2-24	1	66742009	66742009	T	G	SNP	Unc80	missense	p.F1427 V	568	0	0	428	51	10.65
Ax2-24	1	98861948	98861948	A	G	SNP	Slco6b1	missense	p.W312R	618	1	0.16	452	41	8.32
Ax2-24	1	144990579	144990579	A	C	SNP	LOC545 369	rna	NULL	475	0	0	169	182	51.85
Ax2-24	6	22922731	22922731	T	C	SNP	Ptprz1	silent	p.D213	566	1	0.18	264	136	34
Ax2-24	7	29075868	29075868	C	T	SNP	BC08949 1	missense	p.A77T	522	0	0	298	286	48.89
Ax2-24	7	58809247	58809247	G	T	SNP	Ano5	missense	p.K224N	571	0	0	299	163	35.28
Ax2-24	7	128536277	128536277	C	T	SNP	Hs3st2	missense	p.P12S	288	0	0	154	109	41.44
Ax2-24	7	133929701	133929701	T	A	SNP	Ppp4c	missense	p.K300M	484	0	0	572	16	2.72
Ax2-24	7	142843575	142843575	A	G	SNP	ENSMU SG00000 086609	rna	NULL	372	0	0	164	145	46.93
Ax2-24	9	3004676	3004676	C	-	DEL	ENSMU SG00000 079720	frame_shif t_del	p.F72fs	420	11	2.43	313	17	4.97
Ax2-24	9	3018000	3018000	C	-	DEL	ENSMU SG00000 074563	frame_shif t_del	p.S62fs	156	2	1.16	119	5	3.94
Ax2-24	9	3020866	3020866	T	-	DEL	ENSMU SG00000 079719	frame_shif t_del	p.S83fs	287	5	1.7	217	13	5.53
Ax2-24	9	3038250	3038250	T	G	SNP	ENSMU	missense	p.S203R	125	6	4.48	95	9	8.49

Ax2-27	6	125592936	125592936	T	G	SNP	Vwf	missense	p.F1520V	558	0	0	274	247	47.41
Ax2-27	7	49304251	49304251	C	A	SNP	EG637913	rna	NULL	360	0	0	233	31	11.74
Ax2-27	8	37635837	37635837	C	A	SNP	Dlc1	missense	p.D957Y	567	0	0	405	71	14.92
Ax2-27	8	63514744	63514744	C	A	SNP	Nek1	nonsense	p.S295*	583	0	0	317	140	30.63
Ax2-27	8	64466887	64466887	G	T	SNP	Ddx60	missense	p.K1064N	611	0	0	538	16	2.89
Ax2-27	9	45064503	45064503	A	C	SNP	Il10ra	missense	p.F280V	520	0	0	310	182	36.99
Ax2-27	10	77169463	77169463	G	T	SNP	ENSMU SG00000 069584	missense	p.Q25H	345	0	0	203	92	31.08
Ax2-27	10	93308983	93308983	A	T	SNP	Usp44	missense	p.K183N	569	0	0	409	179	30.44
Ax2-27	12	89328372	89328372	C	A	SNP	EG667350	silent	p.G268	814	0	0	481	236	32.87
Ax2-27	12	101358293	101358293	A	C	SNP	Psmc1	missense	p.K326N	603	0	0	315	273	46.43
Ax2-27	12	114216513	114216513	C	T	SNP	Btbd6	silent	p.D333	479	0	0	423	18	4.08
Ax2-27	14	63027536	63027536	T	G	SNP	Gucy1b2	missense	p.K580T	448	0	0	215	183	45.86
Ax2-27	16	10963499	10963499	C	T	SNP	Litaf	silent	p.V76	373	1	0.27	357	74	17.13
Ax2-27	16	38415523	38415523	C	A	SNP	Pla1a	missense	p.A112S	408	0	0	459	18	3.77
Ax2-27	19	47960649	47960649	T	G	SNP	Gsto2	missense	p.D239E	472	0	0	196	210	51.72
Ax2-30	1	155688897	155688897	G	C	SNP	Rgs1l	missense	p.P32R	629	0	0	472	476	50.21
Ax2-30	2	30139992	30139992	G	T	SNP	Dolk	missense	p.D520E	535	0	0	508	357	41.27
Ax2-30	3	96506400	96506400	T	C	SNP	Pias3	missense	p.C360R	377	0	0	413	369	47.19
Ax2-30	5	151157093	151157093	T	C	SNP	Fry	missense	p.L232P	435	0	0	379	397	51.16
Ax2-30	7	24678857	24678857	T	C	SNP	Vmn1r178	missense	p.C104R	318	0	0	182	193	51.33
Ax2-30	7	25076435	25076435	T	C	SNP	Zfp61	missense	p.S437G	485	0	0	366	363	49.79
Ax2-30	7	25300372	25300372	T	C	SNP	Zfp428	silent	p.A118	475	2	0.42	424	448	51.38
Ax2-30	7	52803417	52803417	C	T	SNP	Plekha4	missense	p.S433F	453	0	0	338	343	50.37
Ax2-30	8	64190236	64190236	G	T	SNP	Palld	missense	p.S406Y	582	0	0	376	404	51.79
Ax2-30	9	3018028	3018028	G	T	SNP	ENSMU SG00000 074563	silent	p.S71	176	3	1.63	202	8	3.52
Ax2-30	9	3020974	3020974	G	A	SNP	ENSMU SG00000 079719	missense	p.V118I	67	4	5.48	149	8	5.06
Ax2-30	9	37855568	37855568	G	A	SNP	Olfk884	missense	p.A254T	58	0	0	61	5	7.58
Ax2-30	10	6388470	6388470	T	G	SNP	Plekhg1	missense	p.K996T	517	0	0	413	369	47.13
Ax2-30	10	24635545	24635545	A	G	SNP	Arg1	silent	p.T281	563	0	0	407	344	45.68
Ax2-30	10	30411215	30411215	A	C	SNP	Ncoa7	silent	p.S424	537	0	0	375	378	50.13
Ax2-30	10	92527895	92527895	A	G	SNP	4930485 B16Rik	silent	p.L229	499	0	0	435	421	49.18
Ax2-30	11	67033505	67033505	T	G	SNP	Myh1	silent	p.A1550	537	0	0	455	416	47.76
Ax2-30	11	88790621	88790621	T	G	SNP	Scsep1	missense	p.K385T	567	0	0	435	496	53.28
Ax2-30	11	109097212	109097212	A	G	SNP	Rgs9	splice_site	e17+2	497	0	0	424	441	50.92
Ax2-30	11	119404552	119404552	A	C	SNP	Nptx1	missense	p.L341V	469	0	0	453	408	47.39
Ax2-30	12	77404959	77404959	T	C	SNP	Mthfd1	silent	p.G709	513	0	0	472	480	50.42
Ax2-30	13	12742276	12742276	A	T	SNP	ENSMU SG00000 075118	missense	p.S92T	222	0	0	149	151	50.33
Ax2-30	13	100991464	100991464	A	G	SNP	Naip5	missense	p.L1073P	384	0	0	312	304	49.35
Ax2-30	14	22667517	22667517	C	T	SNP	Comtd1	silent	p.Q81	524	0	0	441	413	48.36
Ax2-30	14	63859543	63859543	T	C	SNP	Gata4	missense	p.Q148R	357	0	0	286	238	45.33
Ax2-30	14	65540405	65540405	T	C	SNP	LOC100 043657	rna	NULL	183	2	1.08	300	11	3.54
Ax2-30	16	16862639	16862639	T	C	SNP	Igll1	splice_site	e2-2	518	0	0	448	439	49.49

Ax2-30	17	21398660	21398660	A	G	SNP	Vmn1r23 5	silent	p.K94	531	0	0	398	415	51.05
Ax2-30	17	25241578	25241578	G	T	SNP	Telo2	missense	p.L607I	551	0	0	433	476	52.31
Ax2-30	18	38134289	38134289	T	C	SNP	Arap3	missense	p.S1315 G	601	0	0	412	394	48.82
Ax2-30	19	6373651	6373651	A	C	SNP	Sf1	missense	p.E322D	476	0	0	415	386	48.07
Ax2-30	19	45497667	45497667	G	A	SNP	Btrc	silent	p.A31	429	0	0	349	214	38.01
Ax2-34	2	86530369	86530369	A	C	SNP	Olf1087	silent	p.T254	457	0	0	196	217	52.54
Ax2-34	4	62197547	62197547	C	A	SNP	4933430I 17Rik	missense	p.S121Y	426	0	0	434	10	2.25
Ax2-34	4	125910866	125910866	T	C	SNP	Mtap7d1	missense	p.T725A	304	0	0	141	153	52.04
Ax2-34	4	126194127	126194127	T	-	DEL	Eif2c4	frame_shif t_del	p.K238fs	414	0	0	266	221	44.74
Ax2-34	4	127973637	127973637	A	G	SNP	Csmd2	splice_site	e8-2	404	0	0	192	198	50.77
Ax2-34	5	4069126	4069126	A	T	SNP	Akap9	missense	p.K3544 N	471	0	0	243	266	52.26
Ax2-34	5	14680949	14680949	T	G	SNP	Pclo	missense	p.I3125S	509	1	0.2	200	201	50.12
Ax2-34	6	29663536	29663536	T	C	SNP	Tspan33	splice_site	e5+2	540	0	0	288	271	48.39
Ax2-34	6	37756264	37756264	T	G	SNP	LOC100 039636	missense	p.V60G	383	0	0	183	185	50.27
Ax2-34	7	51767463	51767463	T	G	SNP	Mybpc2	missense	p.Q531H	501	0	0	256	257	50
Ax2-34	8	91188098	91188098	A	C	SNP	Nod2	missense	p.E356A	434	0	0	165	175	51.32
Ax2-34	8	97427222	97427222	T	G	SNP	Ccdc102 a	missense	p.E530D	498	0	0	183	193	51.19
Ax2-34	10	31671085	31671085	C	T	SNP	Nkain2	missense	p.R105Q	607	0	0	241	217	47.38
Ax2-34	10	41460762	41460762	T	G	SNP	2410017 P07Rik	missense	p.S112R	638	0	0	300	284	48.63
Ax2-34	10	77514613	77514613	T	C	SNP	Dnmt3l	silent	p.T97	514	0	0	204	220	51.89
Ax2-34	11	7060254	7060254	A	T	SNP	Adcy1	nonsense	p.K833*	507	0	0	261	226	46.41
Ax2-34	11	68407816	68407816	A	C	SNP	Ccdc42	missense	p.N71T	458	0	0	238	245	50.72
Ax2-34	11	82817349	82817349	C	T	SNP	Sfn8	missense	p.M655I	563	1	0.18	275	272	49.73
Ax2-34	11	115119409	115119409	C	G	SNP	Grin2c	missense	p.A226P	350	0	0	192	218	53.04
Ax2-34	12	77710130	77710130	G	A	SNP	Spnb1	silent	p.H1388	466	0	0	411	164	28.47
Ax2-34	13	34838618	34838618	C	A	SNP	Fam50b	silent	p.T69	454	0	0	221	237	51.75
Ax2-34	13	46866897	46866897	T	A	SNP	Kif13a	missense	p.E1293 V	558	0	0	225	257	53.21
Ax2-34	14	123769206	123769206	T	C	SNP	Nalcn	missense	p.N691S	634	0	0	295	279	48.52
Ax2-34	15	76454546	76454546	G	A	SNP	Vps28	missense	p.A55V	515	0	0	186	198	51.56
Ax2-34	15	98057945	98057945	C	T	SNP	Olf286	missense	p.A39T	445	0	0	175	175	50
Ax2-34	16	20211717	20211717	A	G	SNP	Yeats2	missense	p.S929G	406	1	0.25	191	196	50.65
Ax2-34	16	25886297	25886297	T	A	SNP	Trp63	nonsense	p.L459*	407	0	0	133	163	55.07
Ax2-34	16	37511664	37511664	A	G	SNP	Gtf2e1	silent	p.H378	321	0	0	175	145	45.31
Ax2-34	17	20494742	20494742	A	G	SNP	Vmn2r10 4	missense	p.S251G	548	0	0	213	254	54.39
Ax2-34	18	57424100	57424100	C	T	SNP	Megf10	silent	p.G532	477	1	0.21	185	204	52.04
Ax2-34	18	84731343	84731343	T	C	SNP	Zfp407	missense	p.I346V	507	0	0	188	178	48.63
Ax2-34	19	43943363	43943363	G	T	SNP	Dnmbp	missense	p.T954N	486	0	0	179	166	48.12
Ax2-34	X	71150019	71150019	A	C	SNP	Arhgap4	missense	p.L248R	269	0	0	0	208	100
Ax2-39	1	90133278	90133278	A	T	SNP	ENSMU SG00000 079429	missense	p.K429M	2260	0	0	1619	155	8.74
Ax2-39	1	136319707	136319707	A	G	SNP	Adipor1	missense	p.K39R	583	0	0	261	240	47.81
Ax2-39	1	141727771	141727771	T	C	SNP	BC02678 2	silent	p.T39	486	0	0	373	28	6.98
Ax2-39	2	68427030	68427030	A	T	SNP	4933409 G03Rik	missense	p.N22I	417	0	0	148	182	55.15
Ax2-39	3	108784121	108784121	A	C	SNP	Fam102b	missense	p.L167R	503	0	0	481	32	6.24
Ax2-39	3	140874290	140874290	A	G	SNP	Pdha2	silent	p.T140	352	0	0	128	127	49.8

Ax2-39	4	153715442	153715442	T	C	SNP	Prdm16	missense	p.E665G	383	0	0	198	210	51.47
Ax2-39	7	109242800	109242800	G	A	SNP	Trpc2	missense	p.R948Q	411	0	0	202	192	48.73
Ax2-39	7	113654622	113654622	C	A	SNP	EG668139	rna	NULL	254	0	0	193	18	8.53
Ax2-39	7	151789396	151789396	T	C	SNP	Ano1	missense	p.K737R	492	0	0	241	242	50.1
Ax2-39	8	98289021	98289021	A	C	SNP	Cnot1	silent	p.T358	568	1	0.18	258	267	50.76
Ax2-39	9	37662895	37662895	A	G	SNP	Olf877	missense	p.Q164R	476	0	0	192	193	50.13
Ax2-39	10	126682531	126682532	-	T	INS	Kif5a	splice_site_ins	e8-1	306	0	0	161	112	41.03
Ax2-39	11	9299450	9299450	G	A	SNP	Abca13	missense	p.A3674T	532	3	0.56	199	209	51.23
Ax2-39	11	73167949	73167949	T	C	SNP	Olf20	silent	p.A231	434	0	0	177	223	55.75
Ax2-39	11	74724553	74724553	T	G	SNP	Srr	missense	p.S131R	557	0	0	461	27	5.52
Ax2-39	11	97535386	97535386	G	T	SNP	Mlt6	splice_site	e10+1	545	0	0	277	313	53.05
Ax2-39	12	113975371	113975371	T	C	SNP	AW555464	silent	p.V589	465	0	0	406	18	4.25
Ax2-39	13	51514903	51514903	T	A	SNP	S1pr3	missense	p.S250R	388	0	0	320	10	3.03
Ax2-39	14	66768162	66768162	G	A	SNP	Chrna2	missense	p.E307K	292	0	0	141	110	43.82
Ax2-39	16	14089095	14089095	A	C	SNP	2900011008Rik	missense	p.N95T	506	0	0	468	34	6.77
Ax2-39	17	34670083	34670083	C	A	SNP	Btn17	missense	p.R497S	587	0	0	268	201	42.77
Ax2-39	18	37495397	37495397	T	C	SNP	Pcdhb6	silent	p.T572	424	0	0	170	162	48.8
Ax2-39	18	62343715	62343715	G	T	SNP	Gm9949	silent	p.L94	403	0	0	227	134	37.12
Ax2-39	19	56983759	56983759	G	A	SNP	Vwa2	missense	p.V669I	448	0	0	169	174	50.73
Ax2-39	X	98079710	98079710	A	C	SNP	Tex11	missense	p.L596V	308	0	0	0	201	100
Ax2-48	5	20116935	20116935	G	A	SNP	Magi2	missense	p.E1158K	583	0	0	273	240	46.78
Ax2-48	5	20402328	20402328	A	T	SNP	Rsb1l	silent	p.T677	475	0	0	255	246	49.1
Ax2-48	7	86528579	86528579	T	C	SNP	Rlbp1	silent	p.A20	420	0	0	215	177	45.15
Ax2-48	7	115198851	115198851	T	A	SNP	Olf479	nonsense	p.Y118*	437	0	0	273	60	18.02
Ax2-48	7	134702596	134702596	T	C	SNP	LOC100043597	missense	p.L2615P	476	0	0	238	243	50.52
Ax2-48	7	150856737	150856737	T	G	SNP	Tnfrsf23	missense	p.K136T	423	0	0	209	223	51.62
Ax2-48	8	19225552	19225552	T	C	SNP	Defb6	silent	p.S16	488	0	0	390	12	2.99
Ax2-48	9	3032387	3032387	A	T	SNP	ENSMUSG0000061971	missense	p.Y100F	101	3	2.88	56	6	9.68
Ax2-48	10	79775100	79775100	A	T	SNP	Apc2	missense	p.Q1052L	543	0	0	283	232	45.05
Ax2-48	11	68091000	68091000	G	T	SNP	Ntn1	missense	p.R378S	406	0	0	208	197	48.17
Ax2-48	11	85323208	85323208	T	A	SNP	Bcas3	silent	p.P500	446	0	0	212	229	51.81
Ax2-48	12	34114754	34114754	C	T	SNP	Twistnb	missense	p.L79F	216	1	0.46	191	34	15.11
Ax2-48	12	112689142	112689142	A	G	SNP	ENSMUSG0000087280	rna	NULL	529	0	0	237	219	48.03
Ax2-48	14	56701831	56701831	T	C	SNP	Mcpt8	silent	p.Q159	512	0	0	223	236	51.42
Ax2-48	17	16806615	16806615	C	T	SNP	LOC100041407	silent	p.P206	340	1	0.29	163	152	48.25
Ax2-48	17	20761379	20761379	G	T	SNP	Fpr-rs3	missense	p.L155M	526	0	0	255	105	29.09
Ax2-48	19	5753928	5753928	G	C	SNP	Ltbp3	splice_site	e18-1	441	0	0	243	241	49.69
Ax2-48	X	5277425	5277425	T	C	SNP	LOC627412	rna	NULL	266	0	0	1	136	99.27

Table 3-7. Common mutations shared among miPSC clones

Clones	chromosome_name	start	stop	reference	variant	type	gene_name	trv_type	amino_acid_change	Normal_ref_count	Normal_var_count	Normal_VAF	Tumor_ref_count	Tumor_var_count	Tumor_VAF
Ax1-10	1	90139354	90139354	C	T	SNP	ENSMUSG00000079429	silent	p.I756	55	6	9.84	81	16	16.49
Ax1-14	1	90139354	90139354	C	T	SNP	ENSMUSG00000079429	silent	p.I756	55	6	9.84	107	48	30.97
Ax1-16	1	90139354	90139354	C	T	SNP	ENSMUSG00000079429	silent	p.I756	55	6	9.84	80	21	20.79
Ax1-2	1	90139354	90139354	C	T	SNP	ENSMUSG00000079429	silent	p.I756	55	6	9.84	80	18	18.37
Ax1-35	1	90139354	90139354	C	T	SNP	ENSMUSG00000079429	silent	p.I756	55	6	9.84	61	13	17.57
Ax1-2	2	69587414	69587500	GAGAGTCC AAACAGAA AAGTAGAA AAGGAAAA GAAAGCTA AAGACCAT AAATCTGA AAGCAAAG AGAGAGAC ATCAGAAG AAATTCA	-	DEL	Ppig	in_frame_del	p.SPNRK VEKEKKA KDHKSES KERDIRR NSE413i n_frame_del	NA	NA	NA	NA	NA	NA
Ax1-35	2	69587414	69587500	GAGAGTCC AAACAGAA AAGTAGAA AAGGAAAA GAAAGCTA AAGACCAT AAATCTGA AAGCAAAG AGAGAGAC ATCAGAAG AAATTCA	-	DEL	Ppig	in_frame_del	p.SPNRK VEKEKKA KDHKSES KERDIRR NSE413i n_frame_del	NA	NA	NA	NA	NA	NA
Ax1-11	4	147049481	147049481	T	C	SNP	ENSMUSG00000062518	missense	p.K280R	23	2	8	39	15	27.78
Ax1-16	4	147049481	147049481	T	C	SNP	ENSMUSG00000062518	missense	p.K280R	23	2	8	26	12	31.58
Ax1-23	10	52024482	52024482	A	G	SNP	Dcbld1	missense	p.I216M	325	16	4.69	180	196	52.13
Ax1-5	10	52024482	52024482	A	G	SNP	Dcbld1	missense	p.I216M	325	16	4.69	253	253	50
Ax2-34	10	52024482	52024482	A	G	SNP	Dcbld1	missense	p.I216M	498	16	3.11	215	211	49.53
Ax2-24	10	52024482	52024482	A	G	SNP	Dcbld1	missense	p.I216M	498	16	3.11	223	206	48.02

Ax1-10	10	57685755	57685755	G	A	SNP	Dux	missense	p.P664S	121	12	9.02	146	40	21.51
Ax1-18	10	57685755	57685755	G	A	SNP	Dux	missense	p.P664S	121	12	9.02	126	34	21.25
Ax1-3	10	57685755	57685755	G	A	SNP	Dux	missense	p.P664S	121	12	9.02	200	58	22.39
Ax1-5	10	57685755	57685755	G	A	SNP	Dux	missense	p.P664S	121	12	9.02	170	61	26.41
Ax1-8	10	57685755	57685755	G	A	SNP	Dux	missense	p.P664S	121	12	9.02	145	43	22.87
Ax1-16	10	117032338	117032339	-	ATGA	INS	Kifc5c	frame_shi ft_ins	p.C434fs	150	6	3.85	206	25	10.82
Ax1-23	10	117032338	117032339	-	ATGA	INS	Kifc5c	frame_shi ft_ins	p.C434fs	150	6	3.85	166	19	10.27
Ax1-5	11	3046912	3046913	-	ACA	INS	LOC10004 4660	in_frame_ ins	p.260in_ frame_in sV	77	13	14.4 4	142	40	21.98
Ax1-8	11	3046912	3046913	-	ACA	INS	LOC10004 4660	in_frame_ ins	p.260in_ frame_in sV	77	13	14.4 4	125	24	16.11
Ax1-14	15	3003747	3003747	G	A	SNP	LOC67420 7	rna	NULL	42	4	8.7	90	23	20.35
Ax1-5	15	3003747	3003747	G	A	SNP	LOC67420 7	rna	NULL	42	4	8.7	76	18	19.15
Ax1-7	15	3003747	3003747	G	A	SNP	LOC67420 7	rna	NULL	42	4	8.7	84	17	16.83
Ax2-39	1	90163167	90163167	C	A	SNP	Hjulp	missense	p.G198V	126	24	16	82	19	18.81
Ax2-26	1	90163167	90163167	C	A	SNP	Hjulp	missense	p.G198V	126	24	16	72	18	20
Ax2-24	1	90163167	90163167	C	A	SNP	Hjulp	missense	p.G198V	126	24	16	68	19	21.84
Ax2-16	1	90163167	90163167	C	A	SNP	Hjulp	missense	p.G198V	126	24	16	60	16	21.05
Ax2-11	1	90163167	90163167	C	A	SNP	Hjulp	missense	p.G198V	126	24	16	65	20	23.53
Ax2-4	2	144190261	144190261	C	T	SNP	LOC63509 7	rna	NULL	87	1	1.14	52	49	48.51
Ax2-27	2	144190261	144190261	C	T	SNP	LOC63509 7	rna	NULL	87	1	1.14	48	35	42.17
Ax2-34	4	146092582	146092582	T	A	SNP	Gm13051	missense	p.N394K	39	1	2.5	44	10	18.52
Ax2-30	4	146092582	146092582	T	A	SNP	Gm13051	missense	p.N394K	39	1	2.5	55	14	20.29
Ax2-27	4	146092582	146092582	T	A	SNP	Gm13051	missense	p.N394K	39	1	2.5	36	9	20
Ax2-6	4	146553590	146553590	C	T	SNP	Gm13139	silent	p.C438	48	1	2.04	68	10	12.82
Ax2-48	4	146553590	146553590	C	T	SNP	Gm13139	silent	p.C438	48	1	2.04	46	7	12.96
Ax2-30	4	146553590	146553590	C	T	SNP	Gm13139	silent	p.C438	48	1	2.04	90	20	18.18
Ax2-27	4	146553590	146553590	C	T	SNP	Gm13139	silent	p.C438	48	1	2.04	52	10	16.13
Ax2-24	4	146553590	146553590	C	T	SNP	Gm13139	silent	p.C438	48	1	2.04	54	12	18.18
Ax2-16	4	146553590	146553590	C	T	SNP	Gm13139	silent	p.C438	48	1	2.04	44	7	13.73
Ax2-4	4	146958038	146958038	A	T	SNP	Gm13154	missense	p.K418I	29	0	0	36	9	20
Ax2-27	4	146958038	146958038	A	T	SNP	Gm13154	missense	p.K418I	29	0	0	31	7	18.42
Ax2-20	4	146958038	146958038	A	T	SNP	Gm13154	missense	p.K418I	29	0	0	41	8	16.33
Ax2-16	4	146958038	146958038	A	T	SNP	Gm13154	missense	p.K418I	29	0	0	14	8	36.36
Ax2-4	4	146958044	146958044	G	A	SNP	Gm13154	missense	p.S420N	24	1	4	29	15	34.09
Ax2-27	4	146958044	146958044	G	A	SNP	Gm13154	missense	p.S420N	24	1	4	27	10	27.03
Ax2-20	4	146958044	146958044	G	A	SNP	Gm13154	missense	p.S420N	24	1	4	33	14	29.79
Ax2-6	4	147049019	147049019	T	C	SNP	ENSMUSG 00000062 518	missense	p.K434R	33	1	2.94	32	9	21.95

Ax2-4	4	147049019	147049019	T	C	SNP	ENSMUSG 00000062 518	missense	p.K434R	33	1	2.94	46	11	19.3
Ax2-27	4	147049019	147049019	T	C	SNP	ENSMUSG 00000062 518	missense	p.K434R	33	1	2.94	29	7	19.44
Ax2-34	4	147129614	147129614	A	G	SNP	ENSMUSG 00000078 495	silent	p.C296	22	1	4.35	27	6	18.18
Ax2-30	4	147129614	147129614	A	G	SNP	ENSMUSG 00000078 495	silent	p.C296	22	1	4.35	33	11	25
Ax2-27	4	147129614	147129614	A	G	SNP	ENSMUSG 00000078 495	silent	p.C296	22	1	4.35	22	14	38.89
Ax2-24	4	147129614	147129614	A	G	SNP	ENSMUSG 00000078 495	silent	p.C296	22	1	4.35	17	5	22.73
Ax2-48	7	148825006	148825006	T	A	SNP	Muc6	missense	p.H1640 L	704	16	2.22	766	27	3.4
Ax2-4	7	148825006	148825006	T	A	SNP	Muc6	missense	p.H1640 L	704	16	2.22	838	26	3.01
Ax2-20	7	148825006	148825006	T	A	SNP	Muc6	missense	p.H1640 L	704	16	2.22	901	22	2.38
Ax2-16	7	148825006	148825006	T	A	SNP	Muc6	missense	p.H1640 L	704	16	2.22	686	21	2.97
Ax2-11	7	148825006	148825006	T	A	SNP	Muc6	missense	p.H1640 L	704	16	2.22	601	19	3.06
Ax2-39	8	19878743	19878743	G	A	SNP	2610005L 07Rik	silent	p.V10	48	1	2.04	45	10	18.18
Ax2-30	8	19878743	19878743	G	A	SNP	2610005L 07Rik	silent	p.V10	48	1	2.04	63	12	16
Ax2-6	9	3018946	3018946	C	A	SNP	ENSMUSG 00000074 563	missense	p.T202K	103	6	5.13	81	10	10.1
Ax2-39	9	3018946	3018946	C	A	SNP	ENSMUSG 00000074 563	missense	p.T202K	103	6	5.13	65	9	11.39
Ax2-30	9	3018946	3018946	C	A	SNP	ENSMUSG 00000074 563	missense	p.T202K	103	6	5.13	149	19	10.98
Ax2-26	9	3018946	3018946	C	A	SNP	ENSMUSG 00000074 563	missense	p.T202K	103	6	5.13	80	14	14.29
Ax2-16	9	3018946	3018946	C	A	SNP	ENSMUSG 00000074 563	missense	p.T202K	103	6	5.13	53	9	14.06
Ax2-6	9	3018970	3018970	T	A	SNP	ENSMUSG 00000074 563	missense	p.F210Y	147	8	5.16	99	16	13.91
Ax2-27	9	3018970	3018970	T	A	SNP	ENSMUSG 00000074 563	missense	p.F210Y	147	8	5.16	102	14	12.07
Ax2-39	9	3020852	3020852	G	-	DEL	ENSMUSG 00000079 719	frame_shi ft_del	p.S77fs	239	4	1.58	165	5	2.89
Ax2-24	9	3020852	3020852	G	-	DEL	ENSMUSG 00000079 719	frame_shi ft_del	p.S77fs	239	4	1.58	192	8	3.9
Ax2-48	9	3030594	3030594	G	-	DEL	ENSMUSG 00000074 559	frame_shi ft_del	p.S87fs	52	2	3.39	57	8	11.43
Ax2-34	9	3030594	3030594	G	-	DEL	ENSMUSG 00000074 559	frame_shi ft_del	p.S87fs	52	2	3.39	44	12	20.34

Ax2-30	9	3030594	3030594	G	-	DEL	ENSMUSG 00000074 559	frame_shi ft_del	p.S87fs	52	2	3.39	69	14	15.91
Ax2-26	9	3030594	3030594	G	-	DEL	ENSMUSG 00000074 559	frame_shi ft_del	p.S87fs	52	2	3.39	28	8	20
Ax2-20	9	3030594	3030594	G	-	DEL	ENSMUSG 00000074 559	frame_shi ft_del	p.S87fs	52	2	3.39	52	14	20.59
Ax2-39	9	3038207	3038207	C	A	SNP	ENSMUSG 00000074 558	missense	p.S189Y	223	12	5.06	141	13	8.39
Ax2-27	9	3038207	3038207	C	A	SNP	ENSMUSG 00000074 558	missense	p.S189Y	223	12	5.06	168	17	9.09
Ax2-24	9	3038207	3038207	C	A	SNP	ENSMUSG 00000074 558	missense	p.S189Y	223	12	5.06	147	17	10.12
Ax2-27	9	3038247	3038247	C	T	SNP	ENSMUSG 00000074 558	silent	p.F202	123	15	10.8 7	81	20	19.61
Ax2-26	9	3038247	3038247	C	T	SNP	ENSMUSG 00000074 558	silent	p.F202	123	15	10.8 7	94	24	20.34
Ax2-24	9	3038247	3038247	C	T	SNP	ENSMUSG 00000074 558	silent	p.F202	123	15	10.8 7	87	23	20.72
Ax2-48	14	121112967	121112967	T	A	SNP	LOC66883 5	rna	NULL	165	3	1.79	171	15	8.06
Ax2-39	14	121112967	121112967	T	A	SNP	LOC66883 5	rna	NULL	165	3	1.79	162	15	8.47
Ax2-30	14	121112967	121112967	T	A	SNP	LOC66883 5	rna	NULL	165	3	1.79	292	20	6.41
Ax2-24	14	121112967	121112967	T	A	SNP	LOC66883 5	rna	NULL	165	3	1.79	155	13	7.74
Ax2-11	14	121112967	121112967	T	A	SNP	LOC66883 5	rna	NULL	165	3	1.79	129	8	5.8
Ax2-34	16	32755261	32755262	-	T	INS	Muc4	frame_shi ft_ins	p.R326fs	37	5	11.9	48	10	17.24
Ax2-30	16	32755261	32755262	-	T	INS	Muc4	frame_shi ft_ins	p.R326fs	37	5	11.9	69	15	17.86
Ax2-24	16	32755261	32755262	-	T	INS	Muc4	frame_shi ft_ins	p.R326fs	37	5	11.9	36	9	20
Ax2-27	X	121244729	121244730	-	TGA	INS	Vmn2r121	splice_site _ins	e5-1	85	6	6.59	48	10	17.24
Ax2-11	X	121244729	121244730	-	TGA	INS	Vmn2r121	splice_site _ins	e5-1	85	6	6.59	48	11	18.64

Table 3-8. OSK lentiviral integration sites

Clone	Chromosome name	Start	End	Supporting Reads	Gene location
Ax1-2	2	98506736	98507278	4	
Ax1-2	18	42403351	42403680	7	
Ax1-3	2	5804912	5805384	15	Cdc123
Ax1-3	2	98502410	98507353	9	
Ax1-3	4	130274243	130274619	14	Serinc2
Ax1-3	9	3002047	3007568	3	
Ax1-3	9	3024074	3027195	6	
Ax1-3	17	25063321	25063633	12	lft140 and Tmem204
Ax1-5	2	98502397	98507406	40	
Ax1-5	4	129341285	129341600	3	Zbtb8os
Ax1-5	9	3000297	3017965	25	
Ax1-5	9	3020843	3032855	18	
Ax1-5	9	56271726	56272024	16	Peak1
Ax1-5	12	3109872	3109978	8	
Ax1-7	2	98506703	98507261	3	
Ax1-7	4	98180096	98180353	10	
Ax1-8	9	71779801	71780302	12	
Ax1-8	14	8395160	8395240	4	
Ax1-10	1	164653171	164653431	9	
Ax1-10	5	51426546	51426987	4	
Ax1-10	13	55365658	55366178	8	
Ax1-11	2	98502678	98507455	14	
Ax1-11	8	35736558	35736868	11	
Ax1-11	9	3000351	3014078	7	
Ax1-11	9	3024426	3034834	5	
Ax1-11	15	101150866	101151174	14	
Ax1-14	2	98502403	98507339	4	
Ax1-14	12	7921237	7921438	8	AK146888
Ax1-14	X	143932298	143932828	12	Dcx
Ax1-16	9	3000902	3004077	3	
Ax1-16	9	77355602	77355808	3	
Ax1-16	X	100516732	100525473	45	
Ax1-18	12	56243747	56243968	4	
Ax1-18	12	56318811	56319071	10	
Ax1-18	12	67829660	67829899	6	
Ax1-18	X	100516732	100525474	60	
Ax1-23	2	98507251	98507281	3	
Ax1-23	9	3003341	3018499	3	
Ax1-35	15	50590987	50591387	5	
Ax2-4	2	98506401	98507270	11	
Ax2-4	3	97567477	97568309	13	Chd1l
Ax2-4	5	104682819	104683209	12	

Ax2-4	9	3000533	3017095	5	
Ax2-6	4	108996772	108997136	7	
Ax2-11	1	95594871	95595225	3	St8sia4
Ax2-11	X	100516736	100525464	10	
Ax2-16	2	98506404	98507324	4	
Ax2-16	3	126629794	126629958	15	Camk2d
Ax2-16	9	3000531	3014067	3	
Ax2-16	16	85940412	85940466	3	
Ax2-20	11	96208908	96209155	8	
Ax2-24	4	93684035	93684178	4	
Ax2-26	2	98502408	98507283	5	
Ax2-26	3	88776847	88777275	8	Gon4l
Ax2-26	3	117503426	117503817	6	
Ax2-26	9	3000474	3017976	3	
Ax2-26	9	72125875	72125963	4	
Ax2-27	2	98502844	98507285	4	
Ax2-27	17	28731783	28731998	6	Mapk14
Ax2-30	2	98502394	98507363	29	
Ax2-30	2	146134656	146134990	13	4930529M08Rik
Ax2-30	7	79841973	79856380	15	Anpep
Ax2-30	8	66526382	66526800	16	
Ax2-30	9	3000478	3020221	23	
Ax2-30	9	3023512	3032852	10	
Ax2-30	19	58930990	58931290	13	Hspa12a
Ax2-30	X	100516733	100525478	85	
Ax2-34	2	98502400	98507273	6	
Ax2-34	8	75425873	75426388	8	
Ax2-34	9	3000345	3014062	5	
Ax2-34	9	107927909	107928244	8	AK014951
Ax2-39	2	100796359	100796568	5	
Ax2-39	X	99161831	99162249	6	
Ax2-39	X	99570928	99571034	4	M55023
Ax2-48	2	98502810	98507283	6	
Ax2-48	7	19714217	19714426	8	Tomm40
Ax2-48	9	3000924	3014061	3	
Ax2-48	10	23250911	23251078	7	Eya4
Ax2-48	16	77574612	77574902	5	2810055G20Rik
Ax2-48	X	100516734	100525462	61	

Table 3-9. Integration “hotspots”

Clone	Chromosome name	Start	End	Supporting Reads
Ax1-2	2	98506736	98507278	4
Ax1-3	2	98502410	98507353	9
Ax1-5	2	98502397	98507406	40
Ax1-7	2	98506703	98507261	3
Ax1-11	2	98502678	98507455	14
Ax1-14	2	98502403	98507339	4
Ax1-23	2	98507251	98507281	3
Ax2-4	2	98506401	98507270	11
Ax2-16	2	98506404	98507324	4
Ax2-26	2	98502408	98507283	5
Ax2-27	2	98502844	98507285	4
Ax2-30	2	98502394	98507363	29
Ax2-34	2	98502400	98507273	6
Ax2-48	2	98502810	98507283	6
Ax1-3	9	3002047	3007568	3
Ax1-3	9	3024074	3027195	6
Ax1-5	9	3000297	3017965	25
Ax1-5	9	3020843	3032855	18
Ax1-11	9	3000351	3014078	7
Ax1-11	9	3024426	3034834	5
Ax1-16	9	3000902	3004077	3
Ax1-23	9	3003341	3018499	3
Ax2-4	9	3000533	3017095	5
Ax2-16	9	3000531	3014067	3
Ax2-26	9	3000474	3017976	3
Ax2-30	9	3000478	3020221	23
Ax2-30	9	3023512	3032852	10
Ax2-34	9	3000345	3014062	5
Ax2-48	9	3000924	3014061	3
Ax1-16	X	100516732	100525473	45
Ax1-18	X	100516732	100525474	60
Ax2-11	X	100516736	100525464	10
Ax2-30	X	100516733	100525478	85
Ax2-48	X	100516734	100525462	61

Table 3-10. Expression changes in genes located, upstream, or downstream of integration sites

Clone	Chromosome name	Start	End	Supporting Reads	Gene location	Expression	Upstream gene	Expression	Downstream gene(s)	Expression
Ax1-18	12	56243747	56243968	4			Brms1l	NC	Nkx2-1	NA
Ax1-18	12	56318811	56319071	10			Brms1l	NC	Nkx2-1	NA
Ax1-18	12	67829660	67829899	6			Mdga2	NA	Rps29/Lrr1/Rpl36a1/Mgat2	NA/NA/NA/NC
Ax1-18	X	100516732	100525474	60			Igfbp1	NC	Dgat216	NA
Ax1-35	15	50590987	50591387	5			Csmd3	NA	Trps1	NC
Ax2-26	2	98502408	98507283	5			Lrrc4c	NA	Rag1/Rag2/Traf6/Prr5l	NA/NA/NC/NC
Ax2-26	3	88776847	88777275	8	Gon4l	NA	Syt11	NA	Gon4l/Msto1/Dap3	NA/NA/NA
Ax2-26	3	117503426	117503817	6			D3Bwg0562e	NA	Snx7	NA
Ax2-26	9	3000474	3017976	3					Alkbh8	NC
Ax2-26	9	72125875	72125963	4			Tcf12	NC	Zfp280d	DN ~50%
Ax2-34	2	98502400	98507273	6			Lrrc4c	NA	Rag1/Rag2/Traf6/Prr5l	NA/NA/NC/NC
Ax2-34	8	75425873	75426388	8			Rasd2	NA	Nr3c2	NA
Ax2-34	9	3000345	3014062	5					Alkbh8	NC
Ax2-34	9	107927909	107928244	8	AK014951	NA	Mst1r	NA	Actl11	NA
Ax2-39	2	100796359	100796568	5			Lrrc4c	NA	Rag1/Rag2/Traf6/Prr5l	NA/NA/NC/NC
Ax2-39	X	99161831	99162249	6			Efnb1	NA	Pja1	NC
Ax2-39	X	99570928	99571034	4	M55023	NA	Pja1	NC	Tmem28	NA
Ax2-48	2	98502810	98507283	6			Lrrc4c	NA	Rag1/Rag2/Traf6/Prr5l	NA/NA/NC/NC
Ax2-48	7	19714217	19714426	8	Tomm40	NA	ApoE	NC	Pvrl2	NA
Ax2-48	9	3000924	3014061	3					Alkbh8	NC
Ax2-48	10	23250911	23251078	7	Eya4	NC	Tcf21	NA	Rps12/Snora33/Snord100	NA/NA/NA
Ax2-48	16	77574612	77574902	5	2810055G20Rik	NA	Usp25	NC	Cxadr	NC
Ax2-48	X	100516734	100525462	61			Igfbp1	NC	Dgat216	NA

NA: No probesets found for the gene, or the microarray signal intensity for all samples are <200.
 NC: No change in gene expression.
 DN: Gene expression is downregulated.

Chapter 4

Summary and Future Directions

The role of CSD proteins in normal and leukemic hematopoiesis

In chapter 2, we investigated the roles of the cold shock domain-containing proteins, Ybx1 and Msy4, in normal and leukemic hematopoiesis. We found that hematopoiesis in *Msy4*^{-/-} mice and *Ybx1*^{-/-} fetal liver cells is essentially normal. However, when we knocked out Ybx1 in adult hematopoietic cells using a conditional *Ybx1* allele, the long-term repopulating potentials of both *Ybx1*^{lox/+} x *Vav1-cre*^{+/-} and *Ybx1*^{lox/-} x *Vav1-cre*^{+/-} bone marrow cells were found to be significantly reduced, suggesting that *Ybx1*^{lox/+} x *Vav1-cre*^{+/-} bone marrow cells are not equivalent to those that are simply haploinsufficient for *Ybx1*. Therefore, the *Ybx1* conditional knockout model requires further characterization and the data derived from these mice needs to be interpreted with caution.

Using TAT-cre induced floxing and a multiplexed qPCR platform to monitor the floxed population, we were able to track the *Ybx1*-floxed population in *Ybx1*^{lox/-} and *Ybx1*^{lox/-} x *Msy4*^{-/-} bone marrow cells after they were transduced with MSCV-MLL-AF9-IRES-GFP, which is a leukemia-initiating fusion protein that results in a serial replating phenotype. While the *Ybx1*-floxed population in *Ybx1*^{lox/-} samples (*Ybx1*^{-/-} x *Msy4*^{+/+}) was stable after 5 weeks of serial replating, the *Ybx1*-floxed population in *Ybx1*^{lox/-} x *Msy4*^{-/-} samples (*Ybx1*^{-/-} x *Msy4*^{-/-}) was either reduced or undetectable after 3 weeks of replating. Collectively, these data demonstrated that expression of either *Ybx1* or *Msy4* is required to maintain the high rates of self-renewal and proliferation induced by MLL-AF9.

One possible explanation for these discordant results is that the floxed *Ybx1* allele has a dominant negative effect against wildtype Ybx1 and Msy4. Indeed, the Uchiumi *et al.* *Ybx1* knockout model (which targets the same exons [5 and 6] as the Mandinova conditional allele) do have a phenotype in heterozygous *Ybx1* knockout ES cells¹, while haploinsufficient mice with the Lu *et al.* *Ybx1* allele (targeting exons 3) have no detectable phenotype in any assay². Even though we were able to ectopically express this predicted truncated version of Ybx1 in 293T cells, we have not able to detect it in

the bone marrow cells (or other tissues) from $Ybx1^{lox/+} \times Vav1\text{-cre}^{+/-}$ or $Ybx1^{lox/-} \times Vav1\text{-cre}^{+/-}$ mice, with or without treatment with a proteosomal inhibitor *in vitro* (data not shown).

To determine whether the predicted truncation protein encoded by the floxed *Ybx1* allele has a dominant negative effect against wildtype *Ybx1* and/or *Msy4*, we will clone the ORF of the truncated or full length *Ybx1* into an MSCV-IRES-GFP vector, and transduce it into wildtype, $Ybx1^{+/-}$, and $Msy4^{-/-}$ bone marrow cells (along with an empty vector control), inject the transduced cells into lethally irradiated C57BL6/J recipients, and track the fate of GFP⁺ cells in the peripheral blood every 4 weeks. We will also use Western blotting to confirm the expression of the truncated *Ybx1* in the transduced bone marrow cells. We expect that wildtype, $Ybx1^{+/-}$, and $Msy4^{-/-}$ bone marrow cells transduced with MSCV-*Ybx1*-IRES-GFP vector to have a stable percentage of GFP⁺ cells in their peripheral blood, since overexpression of full length, wildtype *Ybx1* is not expected to cause a phenotype in cells with high levels of CSD proteins. If the truncation mutant has dominant negative activity, we expect that wildtype, $Ybx1^{+/-}$, and $Msy4^{-/-}$ bone marrow cells transduced with MSCV-*Ybx1*Trunc (truncated)-IRES-GFP will have a reduction in the percentage of GFP⁺ cells in the peripheral blood over time. This phenotype may be different in mice with different germline “doses” of the CSD proteins, depending on the targets of the dominant negative activity produced by the truncated protein.

We also found inconsistent results in the retroviral MLL-AF9 leukemia model using the *Ybx1* conditional knockout mouse. $Ybx1^{lox/-} \times Vav1\text{-cre}^{+/-}$ bone marrow cells, which had a reduced long-term repopulating potential, survived normally after MLL-AF9 transduction and serial replating *in vitro*, and displayed unaltered leukemia-free survival *in vivo*. In contrast, after TAT-cre induced floxing, the *Ybx1*-floxed population in $Ybx1^{lox/-} \times Msy4^{-/-}$ samples ($Ybx1^{-/-} \times Msy4^{-/-}$) declined in abundance after two weeks of replating *in vitro*, using the same retroviral MLL-AF9 leukemia system. It is not yet clear whether

the discrepancies are due to differences caused by the different methods for inducing floxing, or the different model systems.

To address these issues, we plan to do the following experiments: 1) we will inject TAT-cre induced and MSCV-MLL-AF9-IRES-GFP transduced (or MSCV-IRES-GFP control transduced) bone marrow cells from *Ybx1^{lox/-}* or *Ybx1^{lox/-} x Msy4^{-/-}* mice into lethally irradiated C57BL6/J recipients, and track the *Ybx1*-floxed population in the peripheral blood (and leukemia cells arising in these mice). Based on the serial replating data *in vitro*, we expect that the *Ybx1*-floxed population in *Ybx1^{lox/-} x Msy4^{-/-}* samples (*Ybx1^{-/-} x Msy4^{-/-}*) will not be present in the MLL-AF9 induced tumor cells (i.e. leukemic bone marrow and spleen); 2) We will also test these cells in two other *in vivo* acute leukemia models, using retroviruses that express c-Myc³ or Nup98-HoxA9⁴; 3) We will breed *Ybx1^{lox/-} x Msy4^{-/-}* mice to *Vav1-Cre^{+/-}* mice, which will definitively test whether the loss of both *Ybx1* and *Msy4* in HSPCs will prevent or delay the onset of MLL-AF9 or c-Myc induced leukemia *in vivo*, or abolish the replating phenotype in the retroviral MLL-AF9, c-Myc, or Nup98-HoxA9 models *in vitro*.

If these studies confirm the hypothesis that *Ybx1* and *Msy4* combine to prevent senescence in rapidly proliferating AML cells, and that the cold shock domain contains a dominant negative activity against the full length proteins, this would represent a novel approach for the treatment of AML (and other cancers). We would explore alternative ways to inhibit the activity of cold shock domain proteins, perhaps by screening for small molecules that bind to the CSD domain, or by delivering the CSD into AML cells with a TAT-CSD fusion protein. Biochemical studies of the mechanism underlying the dominant negative effect (prevention of CSD dimer or oligomer formation, etc.) would also improve our understanding of the function of these proteins, and how they might be inhibited with drugs.

Functional and genetic heterogeneity in iPSC clones

In chapter 3, we evaluated functional differences among the hematopoietic potentials of 24 miPSC clones derived from two independent preparations of skin fibroblasts from the same 8-week-old C57BL6/J mouse. Among the 24 clones, we observed varying abilities to produce hematopoietic progenitor cells that could produce colonies in methylcellulose plating experiments. We defined the mutational landscapes in the exomes of all 24 clones in an attempt to define genetic mechanisms that might be relevant for phenotypic variation; this is the largest collection of exome sequencing data of iPSCs from a single source. However, this analysis did not identify specific mutations that clearly explained why some clones had reduced hematopoietic potential. Finally, we compared the expression profiles of iPSC clones with extreme outlier phenotypes for hematopoiesis *in vitro*; this study yielded a small set of candidate genes (including *Wt1* and *Lef1*) that could be relevant for hematopoietic differentiation in mouse iPSCs.

Wt1 and *Lef1* have both been shown to be critical for embryonic and hematopoietic development.⁵⁻¹⁵ Cunningham *et al.* recently reported that *Wt1*-deficient mouse ES cells exhibit a markedly reduced potential in hematopoietic differentiation *in vitro*, using two independent *Wt1* knockout ESC lines.¹⁶ Based on this study, *Wt1* downregulation may represent the most likely candidate gene for hematopoietic phenotype. However, the level of expression in the poor differentiators is similar to that of B6 ESCs; clearly, direct testing of *Wt1* and *Lef1* replacement in these iPSC lines will be required to determine whether they are truly relevant for the phenotype.

To address this issue, we will first validate the protein levels of both *Wt1* and *Lef1* in flow-sorted B6/GFP ESCs, and the six iPSC clones used for the expression array study. Next, we will overexpress *Wt1* and *Lef1* cDNAs (individually, and combined) cloned into MSCV-ires-YFP and MSCV-ires-mCherry vectors to allow each transduced population to be tracked in living cells. These retroviruses will be transduced into iPSC clones with normal vs. low hematopoietic potentials. We will first determine whether CFUs increase with the overexpression of either or both proteins. The percentage of YFP⁺, mCherry⁺,

or YFP⁺mCherry⁺ double positive population in ESC- or iPSC-derived hematopoietic cells will further confirm the relevance of *Wt1* and *Lef1* expression for the “rescue” of hematopoietic potential in these clones. Finally, we will knock down *Wt1* and *Lef1* (individually, and combined) using pooled shRNAs in a vector with a puromycin selection marker (Sigma) in ESCs and iPSCs with normal hematopoietic potential. We will differentiate puromycin-selected ESCs or iPSCs and determine whether the CFUs decrease with the knock down of either or both *Wt1* and *Lef1*, compared to scrambled shRNA controls. Collectively, these studies may provide novel insights into the regulation of these transcription factors, and new information regarding the factors that govern hematopoietic lineage determination.

REFERENCES

1. Shibahara, K. *et al.* Targeted disruption of one allele of the Y-box binding protein-1 (YB-1) gene in mouse embryonic stem cells and increased sensitivity to cisplatin and mitomycin C. *Cancer Sci.* **95**, 348–53 (2004).
2. Lu, Z. H., Books, J. T. & Ley, T. J. YB-1 Is Important for Late-Stage Embryonic Development, Optimal Cellular Stress Responses, and the Prevention of Premature Senescence. *Mol. Cell. Biol.* **25**, 4625–4637 (2005).
3. Luo, H. *et al.* c-Myc rapidly induces acute myeloid leukemia in mice without evidence of lymphoma-associated antiapoptotic mutations. *Blood* **106**, 2452–61 (2005).
4. Kroon, E., Thorsteinsdottir, U., Mayotte, N., Nakamura, T. & Sauvageau, G. NUP98-HOXA9 expression in hemopoietic stem cells induces chronic and acute myeloid leukemias in mice. *EMBO J.* **20**, 350–61 (2001).
5. Herzer, U., Crocoll, A., Barton, D., Howells, N. and Englert, C. The Wilms tumor suppressor gene wt1 is required for development of the spleen Ute Herzer, Alexander Crocoll, Debra Barton, Norma Howells. *Curr. Biol.* **9**, 837–840 (1999).
6. Kreidberg, J. a *et al.* WT-1 is required for early kidney development. *Cell* **74**, 679–91 (1993).
7. Moore, a W., McInnes, L., Kreidberg, J., Hastie, N. D. & Schedl, a. YAC complementation shows a requirement for Wt1 in the development of epicardium, adrenal gland and throughout nephrogenesis. *Development* **126**, 1845–57 (1999).
8. Wagner, K.-D. *et al.* The Wilms' tumor gene Wt1 is required for normal development of the retina. *EMBO J.* **21**, 1398–405 (2002).
9. Galceran, J., Sustmann, C., Hsu, S.-C., Folberth, S. & Grosschedl, R. LEF1-mediated regulation of Delta-like1 links Wnt and Notch signaling in somitogenesis. *Genes Dev.* **18**, 2718–23 (2004).
10. Oosterwegel, M. *et al.* Cloning of murine TCF-1, a T cell-specific transcription factor interacting with functional motifs in the CD3-epsilon and T cell receptor alpha enhancers. *J. Exp. Med.* **173**, 1133–42 (1991).
11. Reya, T. *et al.* Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. *Immunity* **13**, 15–24 (2000).
12. Travis, a, Amsterdam, a, Belanger, C. & Grosschedl, R. LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. *Genes Dev.* **5**, 880–894 (1991).

13. Skokowa, J. & Welte, K. LEF-1 is a decisive transcription factor in neutrophil granulopoiesis. *Ann. N. Y. Acad. Sci.* **1106**, 143–51 (2007).
14. Skokowa, J. *et al.* Interactions among HCLS1, HAX1 and LEF-1 proteins are essential for G-CSF-triggered granulopoiesis. *Nat. Med.* **18**, 1550–9 (2012).
15. Reya, T. *et al.* A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* **423**, 409–14 (2003).
16. Cunningham, T. J., Palumbo, I., Grosso, M., Slater, N. & Miles, C. G. WT1 regulates murine hematopoiesis via maintenance of VEGF isoform ratio. *Blood* **122**, 188–92 (2013).

Cheng “Cynthia” Li

660 S. Euclid Avenue, Campus Box 8007, St. Louis, MO 63110

chengli@wustl.edu | 314-489-7813

EDUCATION

- 2014 **Washington University School of Medicine in St. Louis**
Saint Louis, Missouri
Ph.D. in Molecular Cell Biology
- 2007 **Huazhong University of Science and Technology**
Wuhan, China
B.S. in Biotechnology
- 2006 **National University of Singapore**
Singapore, Singapore
Exchange student

AWARDS & HONORS

- 2013 American Society of Hematology Abstract Achievement Award
- 2008 Selected trainee in The Lucille P. Markey Special Emphasis Pathway in Human Pathobiology at Washington University
- 2006 Excellent Student in Hubei Province, China
- 2006 Undergraduate Student Model with National Scholarship

RESEARCH EXPERIENCE

- 2008-2014 **Ph.D. Candidate in Molecular Cell Biology, Washington University School of Medicine in St. Louis, Laboratory of Dr. Timothy Ley**
- 2005-2007 **Research assistant, Laboratory of Jianfeng Liu at Huazhong University of Science and Technology, China**
- 2006 **Research assistant, Laboratory of Shazib Pervaiz at National University of Singapore, Singapore**

PUBLICATIONS

CONFERENCE ABSTRACT

2013 *55th American Society of Hematology Annual Meeting Poster Abstracts.*

Li C, Helton NM, George DR, Mudd JL, Klco JM, Ley TJ. **Functional Early Hematopoietic Progenitor Cells Derived From Mouse Embryonic Stem Cells and induced Pluripotent Stem Cells.**

2012. *54th American Society of Hematology Annual Meeting Poster Abstracts.* Li

C, George DR, Havey NM, Klco JM, Ley TJ. **Functional Hematopoietic Cells Derived From Mouse Embryonic Stem Cells.**

PEER REVIEWED JOURNAL ARTICLES

Wartman LD, Larson DE, Xiang Z, Ding L, Chen K, Lin L, Cahan P, Klco JM, Welch JS, Li C, Payton JE, Uy GL, Varghese N, Ries RE, Hoock M, Koboldt DC, McLellan MD, Schmidt H, Fulton RS, Abbott RM, Cook L, McGrath SD, Fan X, Dukes AF, Vickery T, Kalicki J, Lamprecht TL, Graubert TA, Tomasson MH, Mardis ER, Wilson RK, Ley TJ. **Sequencing a mouse acute promyelocytic leukemia genome reveals genetic events relevant for disease progression.** *J Clin Invest.* 2011;121(4):1445–1455. doi:10.1172/JCI45284.

Li C, George DR, Mandinova A, Ley TJ. **The role of the cold shock proteins Ybx1 and Msy4 in hematopoiesis.** In preparation.

Li C, Klco JM, George DR, Havey NM, Ley TJ. **Genetic and functional heterogeneity of induced pluripotent stem cells derived from adult skin fibroblasts.** In Preparation.

TEACHING EXPERIENCE

Fall 2009 **Graduate Teaching Assistant**
Washington University in St. Louis
Course: Laboratory on DNA Manipulation

MEMBERSHIP IN PROFESSIONAL ORGANIZATIONS

American Society of Hematology