Fall 2014

Molecular and cellular basis of hematopoietic stem cells maintenance and differentiation

Khanh Linh Duong
University of Iowa

MOLECULAR AND CELLULAR BASIS OF HEMATOPOIETIC STEM CELLS MAINTENANCE AND DIFFERENTIATION

by

Khanh Linh Duong

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Molecular and Cellular Biology
in the Graduate College of
The University of Iowa

December 2014

Thesis Supervisor: Assistant Professor Dana N. Levasseur

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Khanh Linh Duong

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Molecular and Cellular Biology at the December 2014 graduation.

Thesis Committee: _____
Dana N. Levasseur, Thesis Supervisor

_____
Frederick Domann

_____
Marc Wold

_____
Dawn Quelle

_____
John Colgan

To my family and friends.

Character cannot be developed in ease and quiet. Only through experience of trial and suffering can the soul be strengthened, ambition inspired, and success achieved.

- Helen Keller

ACKNOWLEDGMENTS

First, I would like to thank my family for their unwavering support for me during my graduate training. Their love and support kept me going, especially those difficult days when things don't seem to work. Their positive outlooks help me to stay focused and grounded.

Second, I would like to thank my thesis advisor, Dr. Dana Levasseur, for allowing me to be independent and exploring different scientific techniques to answer the important biological questions. Thank you for showing me that no matter how difficult things get, your love for science still remains.

Third, I would like to thank my thesis committee for investing the past four years to help me develop as a scientist and for always challenging me to be the best that I can be. Dr. Frederick Domann, as a MCB program director, you have helped us students secure funding, lab space, and resources to ensure success in our training. As a committee member you have been my second mentor, academically and professionally; you always have the best advices and solutions. I will remember to apply them toward my future endeavors. Thank you Dr. Colgan and Justin Fishbaugh (UI Flow Cytometry Core) for teaching me about flow parameters, gating, and analysis; I went from zero knowledge to designing, processing, and analyzing 7-color flow cytometry.

Finally, I am forever indebted to the Molecular and Cellular Biology program and the University of Iowa Graduate College for all of the mental and financial supports that I received during my graduate training. Thank you to the MCB and Graduate College staff, especially Paulette Vilhauer, for providing useful administrative assistance and helping

me to remember important deadlines. Finally, a big thank to Dr. Mark Anderson for providing additional funding toward my graduate training.

I am truly grateful for many UI faculty, staff, and colleagues for providing me the invaluable supports; thank you for letting me bug you with my scientific questions and methodologies. Thank you everyone for making my graduate training such a wonderful journey.

ABSTRACT

The blood system consists of two main lineages: myeloid and lymphoid. The myeloid system consists of cells that are part of the innate immune response while the lymphoid system consist of cells that are part of humoral response. These responses protect our bodies from foreign pathogens. Thus, malignancies in these systems often cause complications and mortality. Scientists worldwide have been researching alternatives to treat hematologic disorders and have explored induced pluripotent stem cells (iPSCs) and the conversion of one cell type to another.

First, iPSCs were generated by overexpression of four transcription factors: Oct4, Sox2, Klf4 and cMyc. These cells closely resemble embryonic stem cells (ESCs) at the molecular and cellular level. However, the efficiency of cell conversion is less than 0.1%. In addition, many iPS colonies can arise from the same culture, but each has a different molecular signature and potential. Identifying the appropriate iPSC lines to use for patient specific therapy is crucial. Here we demonstrate that our system is highly efficient in generating iPSC lines, and cell lines with silent transgenes are most efficient in differentiating to different cell types.

Second, we are interested in generating hematopoietic stem cells (HSCs) from fibroblasts directly, without going through the pluripotent state, to increase efficiency and to avoid complications associated with a stem cell intermediate. However, a robust hematopoietic reporter system remains elusive. There are multiple hematopoietic reporter candidates, but we demonstrate that the CD45 gene was the most promising. CD45 is expressed early during hematopoiesis on the surface of HSCs; and as HSCs differentiate CD45 levels increase. Furthermore, the CD45 reporter is only active in hematopoietic

cells. We were able to confirm the utility of the CD45 reporter using an *in vitro* and an *in vivo* murine model.

In conclusion, the goal of this research was to expand the knowledge of stem cell reprogramming, specifically the reprogramming of iPSCs. Furthermore, it is our desire that the CD45 reporter system will undergo further validation and find utility in clinical and cell therapy environments.

PUBLIC ABSTRACT

Many human diseases are linked to cellular malfunction. In their search for ways to replace cells, scientists use skin cells to generate “induced” stem cells, which, under the right conditions, can mature into a variety of cell types (e.g., pancreas, liver, eye, brain, and blood). Our goal is to generate high quality and quantity blood cells that have direct clinical applications.

To generate blood cells, we used two approaches with skin cells taken from mice. We used blood-related genes to convert skin cells directly into blood cells. We have yet to identify the most effective genetic factors, but we have narrowed the search. We used skin cells to make induced pluripotent stem cells (iPSCs) that can be coaxed to become blood cells. We employed fluorescent proteins to track iPSC formation and to characterize reprogrammed colonies with or without exogenous genes. iPSCs that lack fluorescence after upregulation of the endogenous genes are more efficient in differentiating to other cell types, thus fluorescent protein usage should be considered for studies gearing toward “patient” cell-based therapy. Finally, to track the reprogramming process of skin cells and stem cells into blood cells, we have identified a blood reporter that was effective in marking blood cells exclusively. This is important to researchers studying the safety of blood cells generated from pluripotent stem cells in therapeutic applications.

Although much research remains before we will be able to replace malfunctioning cells in patients, our research on cell generation produces key findings that lay the foundation for disease treatment.

TABLES OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EB: Embryoid bodies

ESCs: Embryonic stem cells

HSCs: Hematopoietic stem cells

HSPCs: Hematopoietic stem and progenitor cells

iPSCs: induced pluripotent stem cells

LTR: Long terminal repeat

MEFs: Mouse embryonic fibroblasts

MOI: Multiplicity of infection

NSCs: Neural stem cells

OSMK: Oct4/Sox2/cMyc/Klf4

PB: Piggybac

PCR: Polymerase chain reaction

PORN: Poly-DL-ornithinie hydrobromide

qRT-PCR : Quantitative real-time PCR

RF: Reprogramming factor

SCNT: Somatic cell nuclear transfer

TF: Transcription factor

TTFs: Tail tip fibroblasts

VPA: Valproic acid

CHAPTER I

INTRODUCTION

Chapter summary

This thesis will present work centered on advancing understanding of and techniques for clinical utilization of the hematopoietic system. The second chapter will include the identification of a hematopoietic reporter to aid the purification and tracking of hematopoietic cells, and the following of pluripotent cell differentiation or direct reprogramming. The third chapter will discuss work involved in the generation and understanding of iPSCs. The overall goal here is to expand the understanding of cellular reprogramming to advance "personalized" regenerative medicine.

Hematopoiesis and blood development

The blood, also known as the hematopoietic system, is a tissue consisting of both solids (cells) and liquid (plasma). The solids are made of red blood cells (RBC), white blood cells, and platelets, while the liquid portion consists of salts, water, and proteins. In animals, the blood is a key component of the circulatory system, supplying vital nutrients and oxygen to organs such as the brain, heart, lungs, and kidneys. It is crucial for the body to have healthy, active blood, which supplies fresh oxygen, clears carbon dioxide, and affords the body ability to fight off germs by supplying white blood cells. One cell type, the hematopoietic stem cell (HSC), can generate all of the other types of blood cells.

HSC are rare, multipotent stem cells that have the capacity to divide indefinitely (the quality of self-renewal) and give rise to all cell types in the blood system (the quality of stemness). Remarkably, even a single, transplanted HSC has been shown capable of reconstituting an entire

host hematopoietic system [1, 2]. Depending on the location in the body and age of the animal, HSCs exhibit various properties. HSCs found in fetal liver are active, while HSCs found in adult bone marrow are quiescent [3]. In humans, young donors' bone marrow yields HSCs that maintain function better and are much more efficient in reconstituting a recipient's blood system [4, 5].

In vertebrates, hematopoiesis occurs in two waves, the first is considered the primitive wave and the second the definitive wave, as presented in Figure 1. Primitive hematopoiesis occurs in the yolk sac, which provides a large number of red blood cells to supply oxygen to the fast-growing embryo. This phase is transient, and rapidly supplanted by the definitive hematopoiesis phase of the adult. The definitive phase begins when cells are in the aorta-gonad-mesonephros (AGM) region, and progresses as the cells migrate from the fetal liver, to the thymus, to the spleen, and then finally collect in the bone marrow (BM).

During vertebrate development, as hematopoiesis progresses, the cells move through the body, making it difficult to determine where many developmental changes occur. It is widely accepted, however, that HSCs in the BM can give rise to two major immune cell lineages: myeloid and lymphoid (**Figures 2 and 3**). The myeloid lineage mounts the body's first line of response, the innate immune response, when the body first encounters foreign pathogens. This lineage consists of mast cells, eosinophils, macrophages, basophils, neutrophils, and platelets. Pathogens that escape the first line of defense will encounter the body's second line of defense, the humoral immune response, which is mounted by cells from the lymphoid lineage, including natural killer cells, T and B cells. The function and involvement of each cell in different diseases will be described in the next section.

Diseases, prevalence and treatments

Many different cell types comprise the blood, each associated with specific functions. Malfunctions of these cells often lead to disease [6], disorders that can be acute or chronic, genetic or sporadic. Depending on the type of disease and cell involved, the prevalence and prognosis varies. According to the Center for Disease Control (CDC), deep vein thrombosis (VTE) and pulmonary embolism are hematopoietic diseases associated with clotting malfunction and account for over 100,000 deaths per year. VTE has a prevalence of one million in the US. Blood-clotting diseases (e.g., cerebrovascular disease and peripheral artery disease) also cause many deaths; and hereditary hemochromatosis, a symptomatic disease that can lead to liver cirrhosis and cancer if not detected early, affects ~one million Americans. Three to four millions American are carriers for sickle-cell disease. One in 1,500 people have some rare blood disease. Although, that frequency is low in compared to cardiovascular and lung diseases, blood diseases are still a major burden in the Nation's straining budget. According to National Heart, Lung, and Blood Institute (NHLBI) in 2009 anemias alone cost the Nation's economy $6 billion. Direct health expenditures cost $5 billion, and $1 billion in indirect cost of mortality. The World Health Organization (WHO) showed that each year over 300,000 babies born with severe hereditary hemoglobin disorders mainly sickle cell and thalassemia diseases. Collectively, these hematopoietic diseases caused great burden nationally and worldly especially with lack of effective treatments for many rare blood diseases. Thus, increasing in awareness and understanding on these diseases will minimize the cost and mortality especially dealing with the great burdens associating with treatments and management of these diseases. Below is a brief discussion on different blood cells, their functions and involvement in different hematopoietic diseases found in humans.

The mast cell contains a single large nucleus with large granules on its surface. The granules contain histamine and heparin among other cytokines. Mast cells develop in the bone marrow and migrate to other tissues, such as skin, lymph nodes, liver, spleen, and the lining of the lungs, stomach and intestine. In these tissues, mast cells help the immune system protect tissues from foreign pathogens by releasing the contents within each granule to recruit other immune cells. Mast cells are often associated with allergy and anaphylaxis, since their granules contain histamine and heparin. This granulocyte is also involved in wound healing and defends against pathogens by releasing cytokines that recruit other blood cells. While mast cells help the body fight off invading pathogens, if too abundant can lead to mastocytosis, which manifests in two forms: skin and systematic.

When mast cells infiltrate the skin they can cause the disease known as urticarial pigmentosa. When systemic mastocytosis can damage the liver, spleen, bone marrow, and intestines; and malignant cases are often associated with cancer or a blood disorder. Mastocytosis is considered an "orphan disease" since it is not well documented and the exact etiology is still not well established. Treatment is still limited to antihistamine, steroids, and/or chemotherapy.

Eosinophils, bi- or tri-lobe nucleus granulocytes, are produced in bone marrow before migrating into blood or other secondary organs, such as the medulla of the thymus, lower gastrointestinal tract, spleen, ovary, uterus and lymph nodes. Once in the blood these cells have life span of 8-12 hours, but in other tissues they can survive up to 8 days un-stimulated. Like mast cells, eosinophils also contain small granules on the cell surface. These cells are mostly known for fighting parasitic infections, but also play a minor role in allergy and asthma. The granules contain cytokines, histamines, reactive oxygen species, basic proteins, cationic proteins, lipid mediators

and growth factors. Eosinophils also show antiviral activity since their granules contain cationic ribonucleases.

Genetic differences are associated with the various disorders; and overall, most of these disorders are life long, and have dietary restrictions to control their disease. In healthy individuals, eosinophils compromise approximately 1-6% of white blood cells and an increase in this percentage might be an indication of intestinal parasitic infection or diseases such as rheumatoid arthritis, Hodgkin's disease, dermatitis, Addison disease, and esophagitis. Eosinophils can cause hypereosinophilic syndrome (HES), which affects the skin, lungs, gastrointestinal tract, heart, and nervous system. HES affects 0.36 to 6.3 people per 100,000 in US and is a life-long condition with neither a known cause nor a cure. Eosinophils can also build up in the gastrointestinal tract, causing a disorder known as eosinophilic gastrointestinal disorder (EGID). EGID affects up to one in 1,000 and often runs in families. One of the most serious conditions associated with increased eosinophils count is eosinophilic granulomatosis with polyangitis (EGPA). This is characterized by asthma and inflammation of blood vessels, which can trigger a heart attack, indeed half of the deaths attributed to EGPA are from heart attack.

Basophils are white blood cells containing granules with bi-lobed nuclei. The basophil got its name from the fact that the cell can be stained with basic dyes. These granulocytes are produced and mature in the bone marrow. These cells are usually confined to the circulation [7] and for the most part involved in the allergic reaction. Upon activation, basophils granules release histamine, cytokines, leukotrienes, proteoglycans, and proteolytic enzymes. Under normal condition basophils only consist of 0.01-0.3% circulating in the blood, making detecting abnormally low basophils somewhat difficult but a low basophil count causes a condition known as basopenia. Some studies suggested that basopenia is related to autoimmune urticarial, a condition involved

with chronic itching [8]. An abnormally high number of basophils causes a rare condition called basophilia, a common feature of myeloproliferative disorders and prominent in chronic myelogenous leukemia.

Monocytes differentiate in the tissues to give raise to macrophages which function in both innate and adaptive immunity. The primary role of macrophages is to phagocytose dead cells and pathogens. In the ladder, the macrophage acts as an antigen presenting cell (APC) and, upon phagocytosis, displays specific features of the pathogen which then activates lymphocytes – cells of the adaptive immune system. In tissues these macrophages have a life span of several months. In addition to the removal of cell debris and pathogens through phagocytosis, macrophages are also involved in wound healing, iron homeostasis, muscle and limb regeneration. The ability of macrophages to engulf other cells and pathogens also makes these cells vulnerable to various diseases. In cases of tuberculosis, leishmaniasis, chigungunya and HIV infection, macrophages act as a host/reservoir where the pathogens proliferate. Macrophages can also cause progressive lesion plaques of atherosclerosis in heart disease [9]. Elevated numbers of macrophages correlated with poor cancer prognosis especially in cancers of the breast, cervix, bladder and brain [10]. Macrophages enhance tumor growth and progression by releasing cytokines and growth factors that cause chronic inflammation and angiogenesis [10]. Interestingly, increasing pro-inflammatory macrophages in adipose cells has been linked to obesity complications such as insulin resistance and type 2 diabetes [11]. It also been found that defective macrophage function is associated with Crohn's disease [12].

The smallest blood cells are plateletes, also known as thrombocytes, which are derived from megakaryocytes and lack a nucleus. Each megakaryocyte produces about 1,000 to 3,000 platelets, and about 10^{11} platelets are produced daily in healthy adult. Each platelet has a life-span

of eight to nine days in circulation. Platelets exhibit a “disk” shape unless they are stimulated by a broken blood vessel, after which they change form and assume a “spider-like” appearance, releasing fibrin when they reach the site of damage. Platelets prevent excessive bleeding through a process of coagulation in which they bind together to form a blood clot at the site of wound or injury to the blood vessel. A normal platelet count ranges from 150,000 to 450,000 platelets per microliter of blood. A platelet count below 150,000 is considered as thrombocytopenia, while above the 450,000 count is considered as thrombocytosis; neither condition is favorable since having few platelets can lead to excessive bleeding, while too many platelets cause over clotting, which can cause stroke. The cause for these conditions can be either congenital or extrinsic factors, such as medications, kidney infection, excessive alcohol intake, cancer, or chemotherapy. Drugs can suppress hyperactive platelet function or promote additional production of platelets. Transfusion therapy can also deliver additional platelets or eliminate extra platelets, to restore a patient’s platelet count to a normal range.

Erythrocytes, also known as red blood cells (RBCs), function to deliver oxygen (O_2) to the body tissues via blood circulation [13, 14]. In addition to transporting O2, RBC also modulates blood vessels by releasing hydrogen sulfide, nitric oxide, and APT; the hemoglobin in these cells can also release free radicals that can break down a pathogen’s cell wall and membrane. RBCs are bi-concaved, lacking a nucleus, and their cytoplasm contains the biomolecule hemoglobin, which binds to oxygen. This molecule gives the cells the red color appearance. RBCs, developed in the bone marrow, circulate in the periphery for 100-120 days before being digested by macrophages. RBCs are produced in a process called erythropoiesis. Upon sensing the low level of oxygen, the liver produces erythropoietin which then travels to the blood stream to target red blood cell precursors which have the receptors that recognize and bind to erythropoietin. The receptor and

ligand interaction leads to the expansion of RBCs. Mutations or malfunctions can cause the number of RBCs to increase tremendously. An elevated RBCs count is known as polycythemias and can be caused by extrinsic or intrinsic factors. This condition is often serious since it can cause hemorrhages, blood clots, and leukemia, and high blood pressure. Other malfunction in RBCs can lead to different types of anemias – the most common types are iron deficiency, sickle cell disease, and thalassemia. While iron deficiency can be treated with an oral supplement, sickle cell anemia is as yet untreatable. In sickle cell disease, instead of a biconcave disc shape, RBCs exhibit a flattened, crescent shape that adheres to blood vessel walls, blocking the flow of blood to other tissues [15]. The only potential treatment for this double recessive disease is bone marrow transplant, but in the future, potential gene therapy might be able to correct the disease.

Natural killer cells (NK cells) are mononuclear cytotoxic lymphocytes involved in both the innate and adaptive immune responses [16-18]. NK cells primarily target tumor cells and microbial infections by virture of small cytoplasmic granules carrying perforin and granzyme proteases. They minimize the spread of infections and tumor invasion by releasing perforin, which creates small holes on the surface of the target cell; granzymes penetrate through these holes and induce apoptosis. This innate function of NK cells, to destroy harmful microbes and cancerous cells, is tightly regulated. Over activation or malfunction of NK cells often causes autoimmune diseases such as arthritis, autoimmune thyroid disease, rheumatoidathitis (RA), multiple sclerosis (MS), systemic lypus erythematosus (SLE), Sjogren's syndrome, and type I diabetes mellitus (T1DM), psoriasis, juvenile dermatoyositis, and systemic-onset juvenile idiopathic arthritis (JIA). Aberrant expansion of immature NK cells causes functional deficits in the blood and is found in a chronic condition call NK cell lymphocytosis. This condition is often associated with autoimmune

syndromes, including vasculitis peripheral neuropathy, and arthritis. Conversely, patients deficient in NK cells are highly susceptible to early phases of herpes virus infections.

T lymphocytes are another type of white blood cell that plays a role in adaptive immunity. They are produced in the bone marrow but mature in the thymus, hence the name T lymphocytes. In the thymus, T cells undergo negative and positive selection; and the two percent that pass the selection process exit the thymus to become mature immunocompetent T cells. However, the amount of naïve T cells produced in the thymus decreases as the thymus shrinks in a maturing animal. The T cells express T cell receptor (TCR) on their surface, which is primarily made of α and β glycoproteins. There are several types of T cells, each with distinct function: helper T cells (T_H cells), cytotoxic T cells (CTL or T_c cells), memory T cell, regulatory T cells (T_{reg} cells), natural killer cells (NKT cells), mucosal associated invariant T cells (MAITs), and a gamma delta T cells subtype. Helper T cells (T_H cells), positive for the glycoprotein CD4, activates macrophages and assist in the maturation of B cells into memory B cells and plasma cells. Cytotoxic T cells are positive for CD8 glycoprotein on their surface, and recognize MHC I presented by target cells. Tc cells are mostly responsible for targeting virally infected cells and tumor cells, although they also trigger transplant rejection in patients. Memory T cells function as the name implies; they are long-term cells created after the body first encounters antigens. Memory T cells retain memory of a particular pathogen in the event that the body re-encounters it, after which they expand rapidly, to trigger the body's defense system. Treg cells are another type of T cells and are responsible for keeping T cell activity in check. Tregs shutdown T cell mediated immune responses, and suppress auto-active T cells. Finally, natural killer T cells (NKT cells) express the glycosylated CD1d receptor and are activated by glycolipid antigens. When activated, NKT cells can behave like T_h and T_c cells, by releasing cytokines and cytolytic molecules. Improper T cell function can result

in severe combined immunodeficiency (SCID), Omenn syndrome, Cartilage-hair hypoplasia, acquired immune deficiency syndrome (AIDS), DiGeorge syndrome (DGS), chromosomal breakage syndromes (CBS), ataxia telangiectasia (AT) and Wiskott-Aldrich syndrome (WAS). Abnormalities in T cells can also lead to T cell lymphoma.

B lymphocytes are another type of white blood cells that are part of the adaptive immune system. B cells develop in the bone marrow and, upon encountering antigens or other stimulus, become active in the periphery. B cells produce antibodies, phagocytose antigens, and act as antigen presenting cells (APC) to activate T cells; they also produce cytokines for signaling. There are many types of B cells including plasma B cells, memory B cells, B-1 cells, B-2 cells, marginal zone B cells, follicular B cells, and regulatory B cells. Unlike T cells, B cells receptors recognize the native form of soluble antigen in blood or lymph. B lymphocytes can be activated with or without T lymphocytes and are unique in that they express the B cell receptor (BCR) which allows them to recognize and bind to antigens. This receptor consists of two components: the ligand binding, membrane bound immunoglobulin (IgD, IgG, IgM, IgE, or IgA) and the signal transducing moiety, the Ig-a/Ig-B (CD79) heterodimer. The BCR has been implicated in several diseases, including B cell derived lymphoma and autoimmune diseases. Aberrant production of antibodies from B cells can cause rheumatoid arthritis and systemic lupus erythematosus autoimmune diseases.

As mentioned previously, blood diseases could be acute or chronic depending on the type of cells involved. Hence, treatments for the acute diseases can be as simple as taking supplements, while finding the right treatment for chronic or rare blood diseases can be challenging. Diseases such as sickle cell anemia require bone marrow transplant, while lymphoma requires chemotherapy. Nevertheless, these regimens do not cure the underlying cause. Often times,

genetics play a significant role in blood diseases. In serious cases, patients require fresh healthy hematopoietic cells by means of a bone marrow transplant. Finding a suitably matched donor can be difficult; and the robustness and function of HSCs decrease as the donor age increases. BM from older donors must be re-transplanted more frequently; and hematopoiesis of HSC from older people is imbalanced compared to BM received from younger donors [4, 5].

To avoid the complications of transplanting cells from one person to another, new therapies filter patient BM, using an autologous process whereby diseased cells are removed from the peripheral blood and only healthy HSC are permitted to re-enter. This method is efficient in young patients, but efficiency declines with age. Moreover, diseased cells can evade detection, and be reintroduced into the patient to eventually give rise to more malignant cells. To circumvent these issues many researchers have turned to cell-based therapies, to identify ways to convert one cell fate into another, as mentioned in Figure 4 and 5 below. These approaches would render patient-specific treatments, taking mutant cells from a patient, treating the underlying genetic disease and then converting these cells into hematopoietic cells before re-transplanting them into the patient.

The overall goal of the work described in this dissertation is to understand the molecular and cellular basis of cellular reprogramming and to develop safer and more reliable techniques for making stem cells more clinically useful.

Cell-based therapy using pluripotent stem cells and alternative reprogramming.

Cellular re-programming: a brief history

Researchers have long experimented with mature cells to gain understanding of cell plasticity and differentiation ability. The first success was the somatic cell nuclear transfer

(SCNT), introduced by Briggs and King, a method in which nuclei isolated from differentiated blastula cells are transplanted into enucleated eggs [19-21]. Using SCNT, Gurdon et al. generated whole frogs by isolating nuclei from frog intestinal cells and transplanting them into amphibian eggs [22-27]. This early work unequivocally demonstrated that differentiated cells and early embryonic cells have the same genetic information, and that the epigenome of differentiated cells can be reset to a pluripotent state. Thus, epigenetic changes are reversible, including the DNA and histone modification that alter gene expression during development.

In 1997 the first cloned mammal, Dolly the sheep, was reported by Wilmut et al. Soon, other species were also cloned [28]. Beyond such uses of SCNT as proof of principle for cellular plasticity, use of SCNT for practical purposes has met obstacles. For one, animals cloned via SCNT manifest subtle to severe phenotypic and gene expression abnormalities [29, 30]; in addition, the method is technically challenging and the process is not accessible for molecular studies. To bypass these issues, Stevens et al., and Kleinsmith et al. were the first to create embryonal carcinoma cells (ECCs), immortal pluripotent cell lines from teratocarcinomas [31, 32]. Although these cells confer a pluripotent state, they are aneuploid (they contain an abnormal number of chromosomes) and when injected into host blastocysts, populate adult somatic tissues poorly.

Problems in finding a suitable source of cells were solved by using embryonic stem cells (ESCs) isolated from the inner cell mass (ICM) of mouse blastocysts [33, 34]. Pluripotent cells can also be isolated from other sources such as the epiblast of embryonic day 5.5-7.5 of postimplantation mouse embryos, primordial germ cells [PGCs] of the mid gestation embryo, and multipotent germline stem cells [maGSCs] from mouse testicular cells; however, out of all pluripotent stem cells, ESCs (isolated from ICM) are considered the most “naïve” and express appropriate levels of Oct4, Sox2, and Nanog. These cells have the capacity to generate all cell

types in the body except for the extra-embryonic membrane (Figure 6). This might be the reason why only ESCs passed a tetraploid embryo complementation test: the most stringent test of true pluripotentcy. In the extra-embryonic tetraploid-embryo complementation test, diploid ESCs were fused to morula from a tetraploid embryo; the resulting blastocyst was derived primarily from ESCs, while the extraembryonic tissue from the tetraploid host morula [35], demonstrating the complete pluripotency of ESCs.

Despite the advantages of ESCs, ethical concerns around ESC isolation hampered wide application. In 2006, a breakthrough study by Yamanaka et al. created a man-made pluripotent cell line simply by expressing, in fibroblasts, four transcription factors: Oct4, Sox2, Klf4, and cMyc [36]. Later, the Yamanaka group repeated the same procedure with human fibroblasts; and these fibroblasts could de-differentiate to become iPSCs as well. This phenomenon strongly suggests not only that somatic cells have plasticity, but also that humans and mice share the same pluripotency network. The iPSCs (induced pluripotent stem cells) this protocol generated tested as nearly identical to ESCs, both molecularly and functionally. Remarkably, when injected into tetraploid blastocysts, iPSC lines gave rise to “all-iPSC” mice [37-40]. Just like ESCs, iPSCs can divide indefinitely (they have the capacity for self-renewal) and can give rise to any tissue (they are pluripotent).

Although iPSCs were initially generated from mouse fibroblasts, studies showed that many cell types could generate iPSCs. A variety of differentiated cell types have been used: neural cells, keratinocytes, melanocytes, adipose-derived cells, amniotic cells, pancreatic cells, and blood cells [41-47]. Although, some are more permissible for reprogramming than others, the ease of isolating somatic cells for cellular reprogramming and the pluripotent potentials have made iPSCs attractive to clinical researchers. Therapeutic strategies using iPSCs have garnered great interest since they

could be tailored to the patient (e.g., patient skin biopsies could generate iPSCs to treat a genetic disease). Technical advancements in gene correction via zinc finger, CRISPR or homologous recombination, which offer a tractable method for manipulating the iPSC genome have reinvigorated interest in iPSCs since, once mutations are corrected these cells might be put back into the patient.

Pluripotency can be restored in terminally differentiated cells by efficient integration of viruses expressing the four pluripotent factors. Generated in this way, iPSCs not only can overcome the immune barrier (since the cells would be taken from the patient) but can also overcome the ethical concerns of retrieving ESCs. The most efficient method for generating iPSCs, however, uses viral integration of the four OMSK factors. This limits the attractiveness of iPSCs in the clinic since viral integration can be tumorgenic depending on the place of insertion of the viral vector. Furthermore, it was shown that in mouse offspring from the chimeric adults generated from iPS implantation, there was a 20% chance of c-Myc reactivation [48]. To ensure the safety and efficacy of the iPS system, we must find alternatives to using viral integration as well as c-Myc.

Other studies used polycystronic vectors, in which the four reprogramming factors are expressed from a single construct instead of four separate vectors [49]. Since lentiviral system enables large DNA fragments to be inserted, this design minimizes the number of integrations and number of vector used. Another advantage the lentiviral system offers is the establishment of LoxP sites, which reduces the chance of reactivation of the four factors as well as the endogenous gene. Even if vectors with LoxP sites integrate near the promoter region of a tumorigenic gene, they are blocked by the self-inactivating LTR [50-52]. Chapter III describes this system further, detailing how the loxp site was embedded within the 3’ LTR region, which

was also truncated to minimize any promoter/enhancer activity. Upon the integration into the host genome, loxP sites are duplicated near the 5'LTR, which allows excision of the transgenes and removes any possibility of reactivation. Nevertheless, the excision of the exogenous vector is not complete, leaving behind fragments flanking the loxP sites, which might interfere with normal regulation and/or gene expression at the viral integration site.

An alternative method that allows the complete removal of the foreign genetic element is the piggyback system. This transient system incorporates the four Yamanaka genes into a single viral, piggyback (PB) vector, which is subsequently removed from the genome by transient transfection of PB transposase [53-55]. This technique is robust and has high reprogramming efficiency. Unlike the cre-loxP recombinase system, the PB transponsase system removes all the integrated vector material from the genome.

With these approaches, researchers over the years have come up with various methods for generating iPSC lines, with the goal of making them more applicable in the clinic. Each of these methods aimed to generate safer and more reliable iPSCs. A variety of other methods rely on non-integrating viruses, RNAs and proteins, miRNAs, and small molecules.

iPSC generated with non-integrating viruses

A safe and reliable method for generating iPSC lines has been the goal of many researchers. Several modifications to viruses could potentially yield clinically useful cells. Replication-incompetent adenovirus has been used to generate iPSCs from mouse fibroblasts and hepatocytes [56, 57] as well as human fibroblasts [58]. This ubiquitous DNA virus allows transient expression of the exogenous genes without integration into the host genome and can accommodate up to 7.5 kilobase pairs (kD) of inserted DNA fragment [59]. Integration and

tumors have not been reported associated with iPSC lines generated via adenoviral vectors. Despites these advantages, adenovirus packaging capacity is low, so it cannot accommodate polycistronic insertion of foreign genes. Additionally, since this system is transient and lacks internalization of DNA, repeated rounds of transfection are needed generate iPS colonies. Also, some cell types are more admendable to adenovirus infection than others. For example, although Stadfelt et al. were able to generate iPSCs by using adenovirus to overexpress the four factors [57], Okita et al. were unable to generate murine hepatocyte iPS clones with adenovirus alone, and so needed to use retroviruses [56]. Thus, our understanding of cell-specific plasticity must be refined so we can identify which cell type to use with which overexpression system. Importantly, efficiency is significantly low in generating iPSCs compared to other systems.

Another popular choice of vehicle in delivering the exogenous genes is Sendai virus (SeV). This RNA virus does not integrate into the host genome, rather it replicates in the cytoplasm of infected cells, allowing stable expression of reprogramming factors; and SeV does not goes through a DNA phase [60]. This system is as efficient as retrovirus and better than other non-integrating systems in cellular reprogramming, as it requires fewer transductions. iPSCs generated from human fibroblasts using Sendai virus infection were able to differentiate into cardiomyocytes, neurons, bones and pancreatic cells [47]. SeV has also been used to generate human iPSCs from terminally differentiated circulating T cells [61]. To halt cytoplasmic replication of the virus, sendai infected cells can be removed from cell culture by antibody-mediated negative selection against those expressing hemagglutinin-neuraminidase (HN) [47]. Alternatively, others developed a temperature sensitive mutant, created to inhibit viral replication [62]. SeV has several unattractive features, however, since the cost for making it is relatively high, more stringent biosafety containment measures are required, and multiple passages and

clonal expansion of iPS colonies are required to generate very few iPSCs with no detectable trace of virus.

iPSC generated by ectopic expression of mRNA and protein

Synthetic, modified mRNAs have been highly efficient in reprogramming human cells to become iPSCs [63]. The efficiency is as high as lentivirus transduction since using this method can cause the reprogramming factors to be robustly expressed. The drawback is that reprogramming via modified RNAs is technically difficult and labor intensive. Another alternative is the use of protein to overexpress the reprogramming factors, OSMK. Kim et al. demonstrated that protein delivery of the four factors could reprogram human cells into iPSCs [64]. Granted, this method does not involve genomic integration, but the efficiency is extremely low. The introduction of the OSMK protein products (without the use of any viral DNA) has successfully generated iPSCs. Studies generated recombinant proteins of transcription factors that can overcome the cell-membrane barrier by fusion to polyarginine – a cell-penetrating peptide (CPP). In other cases, proteins extracted from ESCs were able to convert mouse fibrobasts into iPSCs [55]. This method of generating iPSCs seem the most promising and safe since it does not involve use of virus and transgenes, making patient-specific cell therapy safe. Nevertheless, the efficiency of using proteins in reprogramming is extremely low, 0.001%.

iPS generated by ectopic expression of miRNA

Recent studies demonstrated that small non-coding RNAs, miRNAs, play important roles in regulating the differentiation of ESCs and iPSCs [65-69]. Deletion of the miRNA processing enzymes Dicer or DGCR8 perturbs both the pluripotency and differentiation of mouse ESCs [70-73]. Accordingly, pluripotency was shown to be regulated by two miRNAs: miR-302 maintains

pluripotency; and miR-145, when overexpressed, causes hESCs to differentiate [65-68]. miR-302 expression is restricted to ESCs and iPSCs [68, 74-76]; and miR-145 is highly upregulated during differentiation of ESCs. Studies showed that expression of the miRNA302/367 cluster was two fold more efficient in generating iPSCs than the traditional OSMK [77], suggesting miR-302 and miR-145 have opposing roles in regulating pluripotency.

miR-302 is part of a positive feedback loop with OCT4/SOX2/KLF4 that prevents ESCs from differentiating [68, 74]. As a consequence, pluripotency persists. OCT4 binds the miR-302 promoter to increase its transcription, while miR-302 activity upregulates OCT4 expression. miR-302 enhances OCT4 transcription by downregulating nuclear receptor subfamily 2, group F, number 2 (NR2F2) post-trancriptionally. NR2F2 directly binds to the OCT4 promoter and inhibits its transcription [68]. miR-302 can facilitate cellular reprogramming by targeting multiple cellular processes, including epigenetic regulation, the cell cycle, and the epithelial-mesynchemal transition (EMT) [68, 74, 75, 78]. Conversely, these proteins cause ESCs to differentiate. Interestingly, knockdown of the tumor suppressor gene p53 (p53) greatly enhanced reprogramming in somatic cells [65, 79-81], likely because p53 binds to the miR-145 promoter and thereby enhances miR-145 expression [65, 82-84]. Transient transfection of miR-291-3p, miR-294, and miR-295 in combination with Oct4, Sox2 and Klf4 increased reprogramming efficiency [85]. Interestingly, inhibiting the miRNAs only expressed in differentiated tissue (such as let-7) simultaneously with expressing only three of the OSMK factors (i.e., Oct4, Sox2, and Klf4) was sufficient to reprogram cells to become iPSCs, and efficiency was increased by 4.3 fold. Thus miRNAs play an important role in reprogramming and can act as reprogramming factors.

iPS generated with small molecules

Studies have tested the effects of small molecule chemicals involved in chromatin modification as a means of increasing reprogramming efficiency. In presence of the four Yamanaka factors, inhibitors of histone deacetylase (HDAC) and DNA methyltransferase significantly improved conversion of mouse fibroblasts into iPSCs [86, 87]. The methyltransferase of interest is 5'-azacytidine; and the HDACs are well known inhibitors – suberoylanilide hydroxamic acid (SAHA), trichostatin A (TSA), and valproic acid (VPA). VPA was the most potent as it increased the reprogramming efficiency by 100 fold and also aided in cellular reprogramming with only Sox2 and Oct4 [87, 88]. It also been shown that BIX01294, an inhibitor of G9a histone methyltransferase, and BayK8644, an L-channel calcium agonist, affect epigenetic modeling indirectly through upstream signaling pathways. These two molecules enhanced reprogramming efficiency of mouse neural progenitors and fibrobrasts in presence of only Oct4 and Klf4 [89-92]. G9a histone methyltransferases methylate histone H3 at lysine 9 (H3K9) and inhibit Oct4 expression by de novo DNA methylation at the Oct4 promoter. Other molecules shown to enhance efficiency include rho-associate kinase (ROCK) inhibitor, Y-27632, inhibitor of Wnt signaling, MEK, FGF, and TGF-b receptor. Despite positive effects on reprogramming kinetics, most of these small molecules also affect the whole genome [54, 93], so although these compounds promote activation of pluripotent genes, their off target effects can activate tumorigenic genes. The latter situation is not ideal. Nevertheless, if we could identify compounds with a more specific and narrow target, the efficiency of cellular reprogramming might increase without the need for viral integration or oncogenes, making cell therapy much more attractive and clinically applicable.

Bypassing the pluripotent state to go directly to the multipotent state

As mentioned above, iPSCs are generated by ectopically expressing the Oct4, Sox2, Myc, and Klf4 (OSMK) transcription factors in mouse embryonic fibroblasts (MEF) and adult human fibroblasts [36, 94, 95]. To achieve and sustain pluripotency, expression of these transcription factors must be upregulated and maintained at optimal levels [96]. Despite promising progress, a fraction of iPSCs might fail to differentiate, and then give rise to teratomas [97, 98], which are benign tumors that contain tissues representing the three embryonic germ layers (ectoderm, mesoderm, and endoderm). These benign tumors have the potential to become malignant [97]. Another pitfall of using iPSCs in cell therapy is the amount of time involved in obtaining the ideal target cells. Human somatic cells require more time for cellular reprogramming compared to mouse. Thus, a significant amount of time will be spent generating iPSCs, and then differentiating them into specific cell types. If the patient is in need of healthy cells immediately, iPSCs might not be the best option. Many studies explored another alternative that can convert one somatic cell type directly to another, without going through the pluripotent state. Likewise, the ultimate goal of this research was to eliminate of the tumorigenicity of iPSCs and shorten the time required to generate desirable cell types, specifically HSPCs for cell-based therapy.

Many cell types have been used as a platform for cellular reprogramming, but fibroblasts are used most commonly, because they are highly accessible and share genetic marks with the type of cells researcher are trying to generate. Fibroblasts can de-differentiate into hematopoietic progenitor cells and Bhatia et al.[99] showed that overexpression of Oct4 in fibroblasts, in the presence of hematopoietic cytokines, transformed them into cells positive for blood markers, including the blood exclusive CD45 marker. The use of reprogrammed fibroblasts for cell therapy

has great potential, as it circumvents the pluripotent state, a state that has prevented clinical use of pluripotent stem cells. Fibroblasts have been successfully converted to cardiomyocytes and neuronal cells by overexpression of particular transcription factors so, one could argue, this strategy might be employed to generate HSPCs as well. This thesis will describe an attempt at using 16 transcription factors and epigenetic regulators to generate HSPCs from fibroblasts, as described in detail in chapter four.

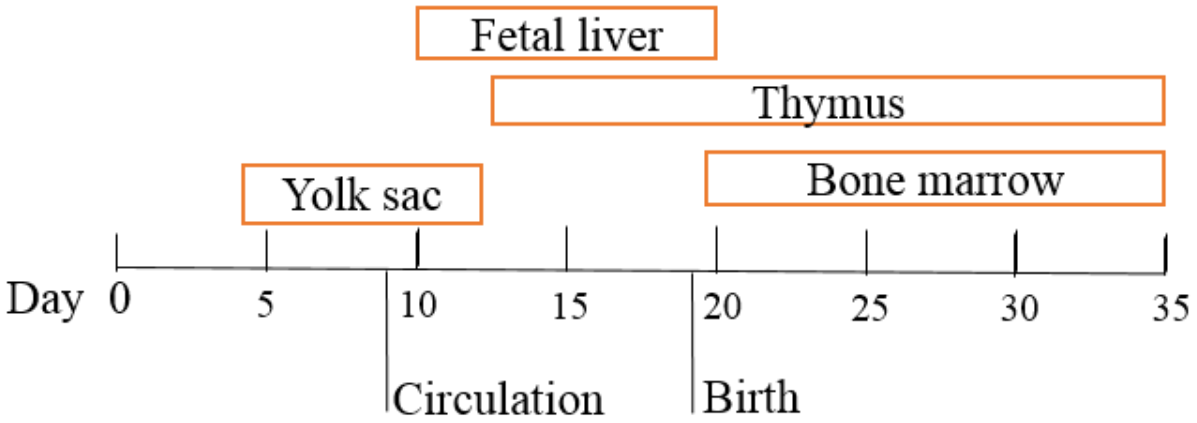


Figure 1. Overview of mouse blood development. Murine hematopoiesis moves through various stages, from the early embryonic to the fetal stage. Specific hematopoietic processes are associated with each stage.

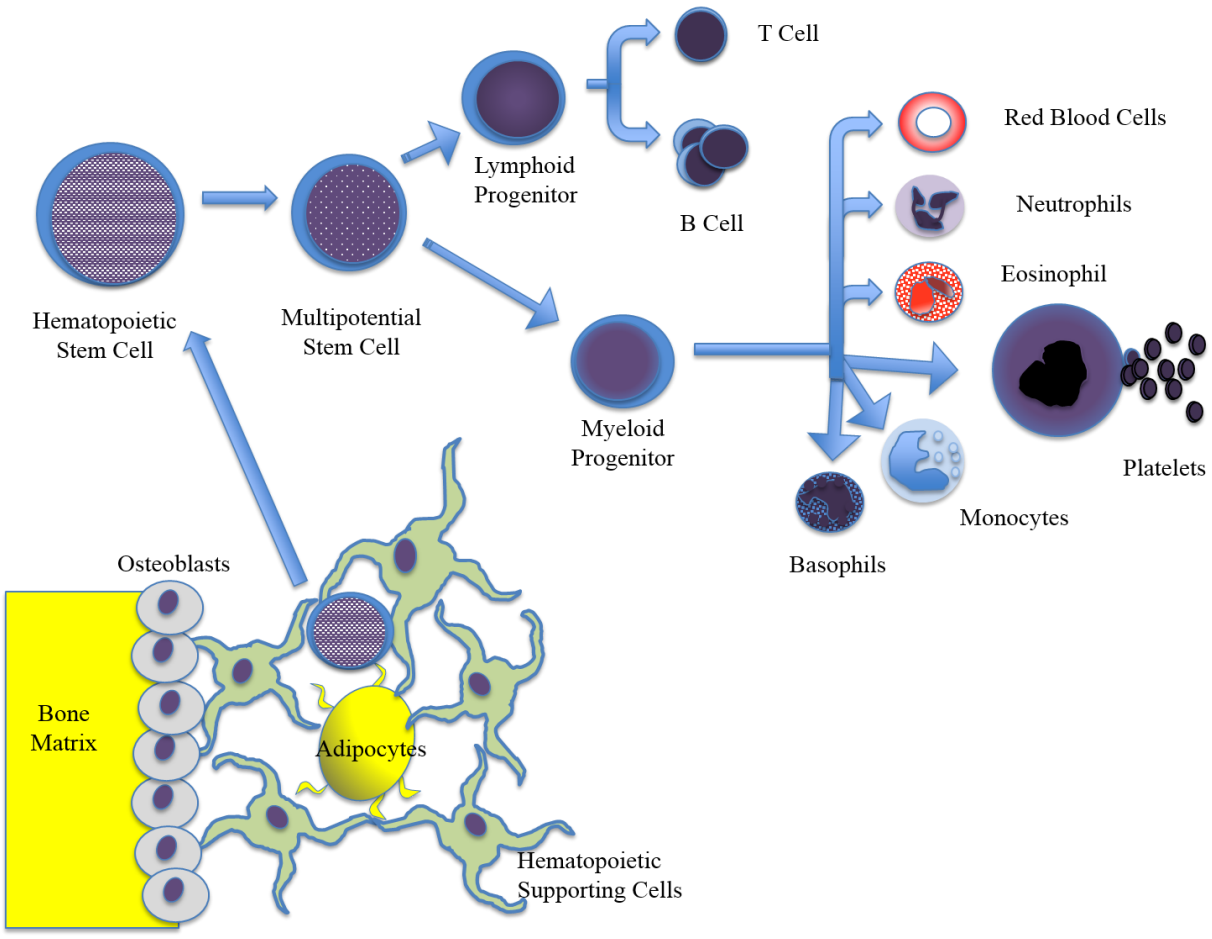


Figure 2. Blood cell development. Bone marrow is the niche that gives rise to hematopoietic stem cells (HSCs) that, within or outside the BM, then differentiate to multiple lineages, including lymphoid and myeloid lineages.

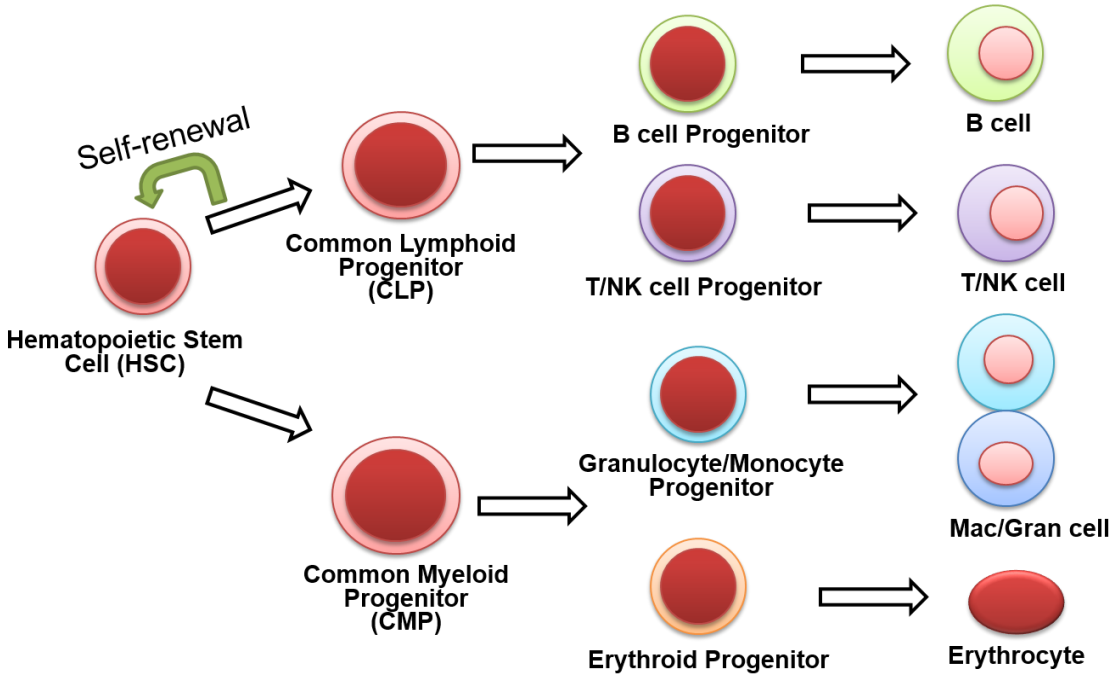


Figure 3. Hematopoiesis of a single hematopoietic stem cell (HSC), exhibiting both stemness and clonality. HSCs give rise to a common myeloid progenitor (CMP) and common lymphoid progenitor (CLP). These precursor cells will further differentiate into specific blood cells.

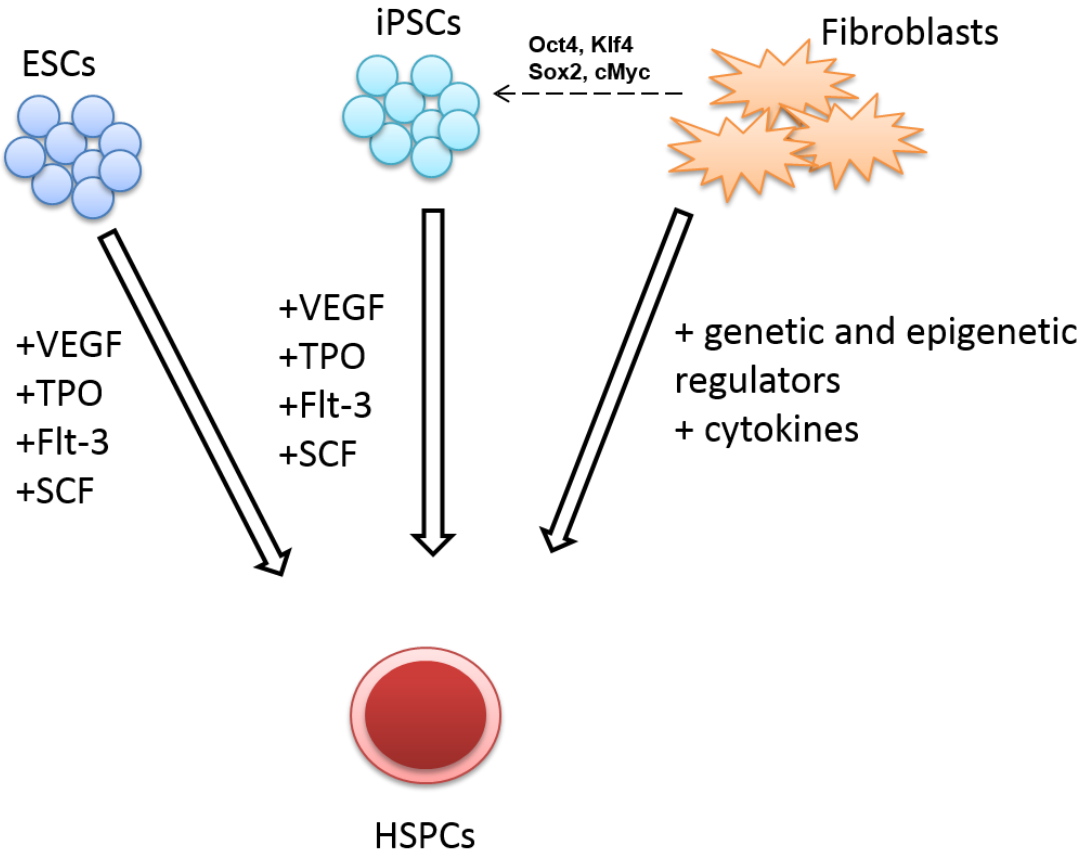


Figure 4. Multiple paths generate hematopoietic stem and progenitor cells (HSPCs). ESC (blue circles) and iPSCs (cyan) can be coaxed to become HSPC, in the presence of HoxB4 and cytokines. The goal here is to identify epigenetic and genetic regulators that convert these cells into HSPCs.

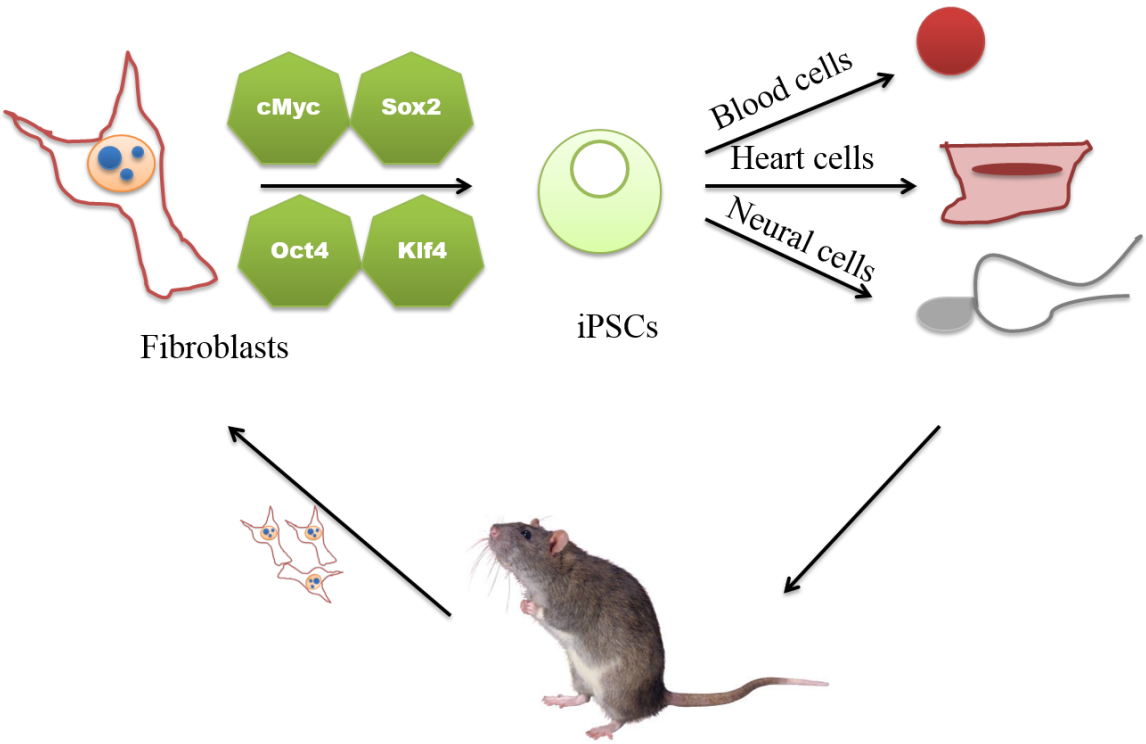


Figure 5. Cell therapy using pluripotent stem cells (iPSCs). Fibroblasts were converted into iPSCs by overexpression of Oct4, Sox2, cMyc, and Klf4 (OSMK). These iPSCs can be differentiated into various cell types, including blood cells, heart cells, and neural cells. The latter can be used to replace damaged cells.

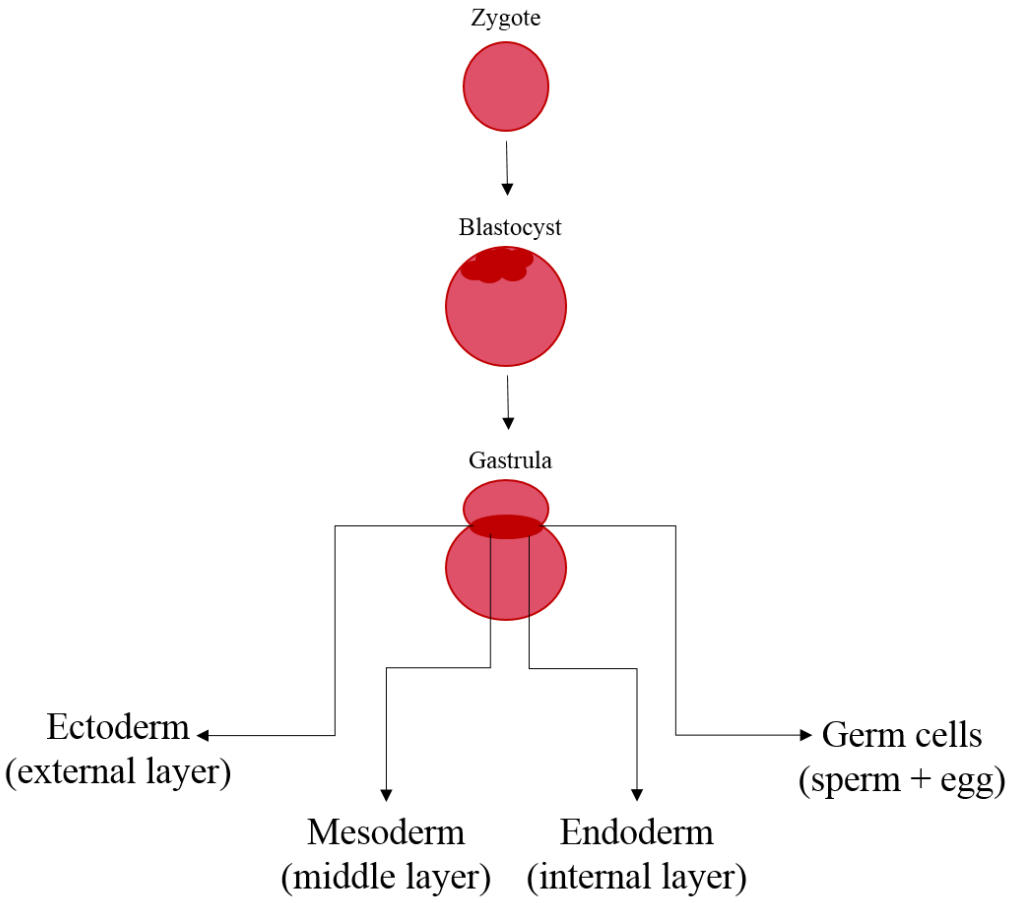


Figure 6. Differentiation of embryonic stem cells. Upon fertilization of the egg and sperm, the zygote develops, rapidly growing and dividing to become the blastocyst. A mass of pluripotent cells within the blastocyst—the inner cell mass—has the capability to give rise to germ cells (sperm and egg) as well as the three germ layers: ectoderm, mesoderm, and endoderm.

CHAPTER II

IDENTIFICATION OF A HEMATOPOIETIC REPORTER

Chapter summary

The development of a hematopoietic reporter is crucial for determining the fate of cells. It will enable safer embryonic stem (ES) and induced pluripotent stem (iPS) cell-based derivation of blood lineages, and facilitate the development of efficient cellular reprogramming strategies based on direct fibroblast conversion. In our search for an ideal hematopoietic reporter candidate gene, we examined several candidates including myb, vav1, Sca-1/Ly6A, CD34, and CD45. We focus extensively on Sca-1, CD34 and CD45 since these candidates showed promising results. The data suggested that the protein tyrosine phosphatase CD45 is the most appropriate choice as a reporter since Sca-1 expression is ubiquitous, and the CD34 promoter is not strong. We used transcription factor chromatin occupancy (ChIP-seq) and promoter nuclease sensitivity (DNase-seq) to discover the reporter genes regulatory elements and to define minimally sufficient sequences required for expression. Subsequently, molecular manipulation distinguished two transcriptional initiation regions essential for high-level expression of CD45. Reporter constructs containing these regions demonstrated high levels of GFP expression in lymphoid, myeloid and nucleated erythroid cell lines derived from mice and humans. These findings were further supported in primary B cell differentiation assay and bone marrow transplantation studies. High GFP expression levels were detected for five months following post-secondary transplantation. Interestingly, when fibroblasts or ESCs were transduced with the reporter, there was a lack of GFP expression. However, GFP expression was observed after ESCs were differentiated into

hematopoietic cells, suggesting that this hematopoietic reporter system would be useful in validating potential autologous blood cell therapies.

Introduction

An effective treatment regimen for hematologic disease and malignancy has been challenging due to a lack of suitably-matched donors [100]. To circumvent this issue, efforts have focused on generating hematopoietic stem and progenitor cells (HSPCs) from embryonic stem (ES) [101, 102] and induced pluripotent (iPS) cells [103, 104]. However, the wide-spread application of pluripotent stem cells is currently hampered by their tumorigenic potential. A proposed alternative is the direct conversion of fibroblasts into HSPCs and blood cells [99, 105]. Although lineage specification and reconstitution potential are currently inefficient [99, 105], evidence suggests that improvements in direct hematopoietic reprogramming could provide a viable strategy for hematopoietic based therapeutics. Aided by GFP reporters, recent studies demonstrated that overexpression of specific transcription factors facilitated generation of neurons and cardiomyocytes from fibroblasts [106, 107], suggesting the conversion of fibroblasts into functional HSPCs was plausible. These studies highlight the importance of having a reporter system for hematopoietic marking and a method to track cellular reprogramming.

Since the blood cell therapy field lacks a reliable reporter for hematopoietic production after differentiation of ES and iPSCs [104, 108], the development of a hematopoietic restricted marking system is essential. Furthermore, a fluorescent reporter system enables real time tracking of full reprogramming [109, 110], permits the study of reprogramming intermediates [111, 112], and may facilitate the eventual use of small molecules for direct reprogramming, as demonstrated recently for iPSC derivation [113]. Additionally, a reporter construct could aid in

the purification and removal of undifferentiated pluripotent cells to minimize teratoma formation upon transplantation.

An effective reporter should be inactive in fibroblasts and pluripotent stem cells, but turned on in the desired reprogrammed cell fate. Transcription factors such as Gata2, Hoxb4 and Evi1 were previously employed as reporters due to their essential roles in HSPC genesis, maintenance and/or amplification [3, 114-118]. However these reporters were not limited to blood cell lineages, and this limited their utility [119, 120]. Additionally, these transgenes used in the production of reporter mice [121] cannot be virally introduced into hematopoietic reconstituting cells because the reporter is too large for the viral backbone [122]. In this study we chose to look at three hematopoietic associated genes: Sca-1/ly6A, CD34 and CD45 as the foundation for our reporter.

The stem cell antigen-1 (Sca-1) is a member of lymphocyte antigen 6 (Ly-6) family. This super family encompasses at least 18 highly conserved genes. Sca-1 is encoded by the *Ly6a* gene and the protein is a 18 kDa glycosyl phosphatidylinositol-anchored membrane protein. Sca-1 has been used as one selectable marker along with others to isolate HSCs from other cells [6, 114]. This protein is expressed at early definitive stage of murine hematopoietic development and specifically on HSC surface. Sca-1 expression fluctuates in both myeloid and lymphoid development. As myeloid progenitors differentiate, Sca-1 is downregulated, but when these became colony forming unit (CFU) Sca-1 expression is upregulated [52]. Similarly, as the lymphoid lineage develops Sca-1 expression is decreased on immature T lymphocytes, and upon the maturation of lymphocytes Sca-1 is then upregulated again [51]. These observations suggest that Sca-1 is an attractive target for a hematopoietic reporter.

The second candidate we wanted to examine was CD34. This marker is a single transmembrane glycoprotein with a mass of 120kDa [123]. This surface antigen is found on a heterogenous cell population that can give rise to clonal aggregates in vitro and reconstituted myelo-lymphopoietic system in vivo [124-126]. The umbilical cord holds sources of hematopoietic stem cells that used for allogeneic bone marrow transplants, and roughly 1-2% of the cells express a high level of CD34 on their surface [127-129]. CD34 is the only known marker for prospectively identifying HSPCs in humans and baboons, and $CD34^+$ bone marrow cells are able to reconstitute long term hematopoiesis upon transplantation into human and primate hosts[130, 131]. Similarly, murine $CD34^+$ HSPCs are able to support short and long-term reconstitution in a mouse transplantation model [132]. These findings suggest CD34 holds promise as a translatable hematopoietic reporter.

In this study, we also examined the transmembrane protein CD45 as the foundation for our reporter. CD45 (also known as Ly-5, B220 and Ptprc) is highly abundant on the surface of all nucleated blood cells, but is absent on other cell types [120, 133]. This receptor is expressed early during hematopoietic development [134]. CD45 expression increases as HSPCs differentiate and transcript expression depends on the developmental stage, lineage specification, and activation state of the cell. Previous work has identified two or more promoters that initiate transcription from one of two alternate starting exons or the first intron, with all transcripts sharing a translational start in exon 2 [134]. Prior studies employing retroviral delivery were unable to document CD45 promoter activity using 0.8kb and 2.8kb promoter fragments [135]. However, Virts *et al.* showed modest expression of a CD45 minigene driven by an 839bp DNA element in stably transfected cytotoxic T cells following drug selection [136]. Additionally, a CD45-YFP reporter mouse was created by Yang *et al.* that facilitated visualization of

hematopoietic cells *in vivo* [137]. This mouse model enabled genetic tracking studies, but since it was a knockin at the translational initiator codon, the identification of minimally sufficient reporting elements required for CD45 marking was precluded. Critically, a virally transducible reporting system that can mark multilineage human hematopoiesis has not yet been published. Based on regulatory elements identified using ChIP-seq and DNase-seq data, we demonstrate that CD45 expression is controlled by a small region that is crucial for driving maximal expression of a GFP transgene. Here, we report that CD45 can serve as a highly specific reporter to efficiently mark multilineage hematopoiesis.

Materials and Methods

Cell lines and culture

Mouse hematopoietic progenitor cell (EML), B cell (CH12.LX), nucleated erythroid progenitor (MEL); and human HSPC (KG1a), B cell (Raji), T cell (Jurkat) and nucleated erythroid (K562) cell lines were used for initial CD45 reporter testing studies.

MEF were isolated from E13.5 embryos from timed pregnant C57BL/6 mice as described previously [36]. Briefly, the head and gut were removed from embryos while the remaining body parts were minced and rinsed in 1X PBS. The tissues were pelleted down and incubated in a 50 ml conical tube containing 0.1% trypsin/0.1 mM EDTA solution (3 ml per embryo) at 37^{o}C for 20 min. Afterwards, to dissociate the tissue, a 10 ml pipette was used to mix the solution, which was then allowed to sit 5 min at room temperature for debris to settle. The supernatant was transferred to a new 50 ml conical tube and spun at 200G for 5 min. The pellet was resuspended with fresh fibroblast media (10% α-MEM medium containing 1% Penicillin/Streptomycin, 1%

glutamine, 1% non-essential amino acid, 1% sodium pyruvate and 10% FBS) and spun again. Cells were then resuspended, counted, and diluted to $1x10^6$ cells/ml. Ten million cells were plated per 10cm dish. The next day the medium was changed to remove the remaining floating cells. MEF were frozen down at passages 2 and 3. J1 ESCs that had been adapted to feeder-free conditions were cultured on gelatin-coated plates in standard 15% D-MEM medium containing 15% FBS, 10 μM 2-mercaptoethanol, 2 mM L-glutamine, 1% of nucleoside mix (100X stock, Sigma), 1000 U/ml leukemia inhibitory factor (LIF) and 1% Penicillin/Streptomycin as previously described [114, 138].

Mouse strains

C57BL/6 mice with the CD45.2/$Ptprc^b$ haplotype and congenic C57BL/6.SJL mice with the CD45.1/ly5.1/$Ptprc^a$ haplotype were obtained from Jackson laboratory. All mouse experiments were in compliance with protocols approved by the Institutional Animal Care and Use Committee of the University of Iowa.

RNA isolation and quantitative real-time PCR (qRT-PCR)

Bone marrow was harvested and treated with 1xBD Pharm Lyse to remove red blood cells. RNA was isolated using the Illustra RNAspin Mini RNA Isolation Kit. cDNAs were synthesized using Promega GoScript Reverse Transcriptase according to manufacturer recommendations. The cDNA products were then diluted 40-fold, and 5 μl was used for each qRT-PCR reaction following amplification with Takara Ex Taq polymerase HS. Primers for CD45, CD34, Vav1, Myb, Gapdh, and Hprt1 were employed (**Table A1**). Detection was performed with a Biorad CFX96 real-time PCR machine.

Gene expression and regulatory element identification

HSPC (c-Kit^{+}, Sca-1^{+}, lineage$^{-/low}$), CMP, GMP, MEP, lymphoid, myeloid and erythroid cell gene expression data were obtained from published data employing Affymetrix high-density oligonucleotide arrays [120, 139]. CD45 gene conservation was obtained from multiple alignments of 30 vertebrate species and measurements of evolutionary conservation using two methods (phastCons and phyloP) derived from the PHAST package [124]. ChIP-Seq datasets were downloaded from GEO (www.ncbi.nlm.nih.gov/geo/, accession numbers: GSE36030, GSE22178, GSE21614, GSE32970, GSE31477), reads were aligned to the mouse (mm9) or human (hg19) reference genomes using Bowtie [140] and peaks were called with MACS [141]. DNAseI hypersensitivity followed by high-throughput sequencing (DNase-seq) [142] datasets were obtained from GEO accession numbers GSE37074 and GSE32970.

Reporter construction

An approximately 1kb enhancer region from the Sca-1 locus was PCR amplified cloned into a GFP plasmid. The 1kb promoter region for Sca-1 was also PCR amplified and cloned into Topo plasmid. The promoter fragment was cut with AscI and BamHI while GFP and enhancer fragment were cut with BamHI and XhoI. These two fragments were ligated together to the Trono lentiviral backbone [51]. For CD34 reporter, the 7 promoter truncations were PCR amplified to contain AttB recombination sequences. The purified promoter fragments were cloned into a Gateway adapted entry vector using a Gateway BP reaction mixtures and subsequently into the lentiviral expression vector via Gateway LR reaction mixtures.

The eight truncations of the CD45 promoter region positioned from -164 to +168 relative to the transcriptional start site were amplified using Ultra Phusion II with primers flanking the

AttB recombination sequences. The PCR products were loaded onto a 2% agarose gel and purified using the Qiagen QIAquick Gel Extraction Kit. Gateway cloning was used to move promoters into the lentiviral vector backbone. The purified promoter fragments were cloned into a Gateway adapted entry vector using a BP reaction. Subsequently, the entry vectors underwent LR reactions with a Gateway adapted donor vector containing EGFP to yield the final lentiviral expression vector. For HM1 embryonic stem (ES) cell transduction, EGFP was replaced with mCherry to enable co-transduction with a HOXB4-GFP expression cassette.

Lentivirus production

The lentiviral vector used in these studies is equipped with a self-inactivating (SIN) LTR, woodchuck hepatitis virus post-transcriptional regulatory element (WPRE) and a central polypurine tract (cPPT) to increase biosafety, full-length RNA production and titer, respectively [50-52, 122]. Three days prior to virus production, 293T cells were split on 2 consecutive days. Approximately 24 hr before transfection, $5.5x10^6$ 293T cells were seeded in 15 cm dish. The next day, 293T cells were co-transfected with 15 μg of packaging plasmid, 5 μg of VSV-G envelope encoding plasmid, and 20 μg vector plasmid per 15 cm plate using a calcium phosphate-based transfection method. At 48 hr, lentivirus supernatants were harvested, passed through a 0.45 μm filter, and concentrated by ultracentrifugation at 26,000 rpm for 90 minutes. The pellets were resuspended in 100 μl OptiMEM and frozen down at -80^{o}C. To ensure quantitative measurement of viral transducing units (TU), real-time quantitative PCR (qRT-PCR) was used for obtaining titers, and this was compared by functional titering using CMV-EGFP transduction of HeLa cells as described elsewhere [51]. Viral RNA was isolated using the QIAamp Viral RNA mini Kit. RNA was then subjected to reverse transcription using the

Promega GoScript Reverse Transcription system. Primers were designed to anneal to the WPRE region.

In vitro B cell differentiation assay

Bone marrow from 3 age and sex matched mice were harvested using B1 buffer (1X PBS containing 0.5% FBS, and 2 mM EDTA). Afterwards, the cell pellet was collected and resuspended in 5 ml 1X BD Pharmlyse for 4.5 min at RT. The cells were rinsed and washed with 30 ml B1 buffer. Subsequently, the pellet was resuspended in 8 ml B1 buffer and incubated with an in house monoclonal antibody cocktail (Ter119, Ly6C, Gr-1, IgM, CD5, and B220). The cell and antibody mixture was incubated on a nutator at 4^{o}C for 30 min. Immediately after, cells were spun down and washed with 10 ml of B1 buffer at 1,500 rpm for 5min. Next, 0.5 ml of washed Dynal Sheep Anti-Rat Beads were added to the cells, and incubated at 4^{o}C for 30 min before magnetic selection. Cells remaining in the supernatant were spun down and resuspended in an HSPC stimulating cocktail (StemSpan SFEM medium with 50 ng/ml mSCF, 100 ng/ml Flt3L, 20 ng/ml IL-3, 50 ng/ml IL-11, and 10 ng/ml TPO) overnight at 37^{o}C. After 20 hr, cells were counted and transduced with an MOI of 30 in the presence of 8 μg/ml polybrene. The next day transduced HSPCs were transferred to 12-well tissue culture plate coated with OP9 stromal cells in Opti-MEM medium containing 5 ng/ml IL-7, 10 ng/ml Flt3, and 10 ng/ml mSCF. Flow cytometry was employed to detect GFP, CD19 (PE-Cy7) and B220 (APC) following 10 and 24 days of culture.

ESC differentiation

HM1 ESCs harboring a HOXB4-GFP expression cassette [143] were transduced with a CD45 reporter driving mCherry for 24hr prior to differentiation as previously described [103,

144]. Briefly, ESCs were subjected to hanging drop formation in culture for 2 days to form embryoid bodies (EB), followed by 4 days in suspension. On day 6, the EBs were dissociated and co-cultured with the OP9 stromal cell line and supplemented with the hematopoietic cytokines mSCF, mVEGF, hTPO and hFLT3L. At day 10, cells were stained with antibodies against c-kit, Sca-1, CD45, and Gr-1/Mac-1. Antibodies used for day 10 staining of ES differentiated hematopoietic cells: c-kit-PerCP-eFluor 710 (eBioscience; cat #: 46117180), Sca-1-Brilliant Violet (BV) 421, (Biolegend; cat #: 108127), CD45-APC (eBioscience; cat #: 17045182), Gr-1/Mac-1-BV 650 (Biolegend; cat#: 108441 and 117339).

Transplantation

Bone marrow from CD45.1 mice was harvested and lineage depleted with biotin conjugated monoclonal antibodies (B220, CD4, Gr-1, Mac-1, NK1.1, CD8). Subsequently, cells were cultured overnight in D-MEM supplemented with 10% FBS, mSCF, hTPO, and hFLT3L. After 24hr, $1x10^6$ cells were transduced with a multiplicity of infection (MOI) of 30 in the presence of 8 μg/ml polybrene and incubated overnight. Next, $2x10^5$ cells were transplanted into congenic CD45.2 mice that had been irradiated with a 950 cGy dose. In transplants receiving supportive bone marrow, $1x10^4$ untransduced cells from the donor were also supplemented. Secondary transplants received $1x10^6$ bone marrow cells from primary donor recipients.

Antibody staining and flow cytometry

For the in vitro B cell differentiation assay, cells were incubated in 1 μg/ml Fc block for 20 min to preclude non-specific antigen binding, followed by two rinses (PBS + 3% FBS) and then incubated for 30 min with antibodies (CD19-PE.Cy7, B220-APC, Mac-1-PE, and Gr-1-PE) and then rinsed twice being processed through a LSR Violet II flow cytometer. For staining of

peripheral blood and bone marrow from transplants, cells were incubated in 2 ml 1X RBC lysis buffer for 5 min at RT. Subsequently the samples were diluted and washed twice in 8 ml of 3% FBS/ PBS. The cells were transferred to a 96 well plate and blocked with 1 μg Fc block for 20 min before washing and incubation with the following conjugated antibodies for 30 min: CD45.1-BV421, Mac-1-PE.Cy7, Gr-1-PE, Ter119-APC, and Cd3ε-Fluor710. FlowJo 8.6 was used for analysis of FACS files.

Results

Sca-1/ly6A is a promiscuous promoter

To find a suitable hematopoietic reporter gene, we examined gene candidates that were highly expressed in bone marrow cells, but excluded in other cell types. By mining publically available expression profiling data, we evaluated the levels of well-established hematopoietic markers, including Sca-1/Ly6A, CD34, CD45, Myb, and Vav1. The expression of these candidates was verified by qRT-PCR, as shown in **Figure 15**. Base on Novartis gene atlas, Sca-1 is highly expressed in blood cells but low or absent in other cell types especially in ESCs (**Figure 7)**. Thus we designed and generated a Sca-1 promoter that contained 1kb promoter region as well as 1kb enhancer region as shown in **Figure 8.** To test for promoter activity we transduced the Sca-1 LV at MOIs of 10, 3 and 1 into 293T, J1 ES, MEF, and SNLP cells. After 48 hrs we used flow cytometry to determine the percentages of GFP^+ cells. It was observed that overall these cells have low reporter activity as evidenced by the lack of GFP positivity. Although, 293T and SNLP cells have higher GFP positivity while MEF and J1 ESCs had lower GFP positivity. Low markings in MEF and J1 ESCs are crucial since these two cell types will be used as platforms for our reprogramming strategies. As the MOI increased, the level of GFP

positivity also increased. To determine the efficiency of Sca-1 as a hematopoietic reporter, we transduced J1 ESCs with Sca-1 LV and coaxed the cells to become HSPCs as described by McKinney et al. [144]. GFP was expressed in embryoid bodies as the ESCs differentiated (**Figure 9**) and the brightness increased after the embryoid bodies were dissociated and grown on top of OP-9 stromal cells and cytokines. There were many cells expressing GFP under the fluorescence microscope. However, when we repeated the experiment to select only for cells that contained the reporter we found that MEF and ESCs were GFP positive after selection for puromycin resistance (**Figure 10**). These results suggested that Sca-1 might not be a useful tool as a reporter since its promoter activity was not exclusive to hematopoietic cells.

CD34 promoter activity is not robust in marking hematopoietic cells

Next, we examined the promoter activity of CD34, a well-known hematopoietic stem cell marker. This marker is expressed in different blood cells but not in ESCs (**Figure 11**). Based on the ChIP-seq and Dnase-seq data, we designed 6 constructs containing the promoter region with the highest enrichment of different transcription factor binding sites (the promoter scheme shown in **Figure 12**). However, upon transduction into the hematopoietic cell line KG1A it appeared that the promoter activity was not robust (**Figure 13** and **Figure 14**). We found that the promoter region from -207 to +128 contained the highest activity as evidenced by GFP expression levels. Conversely, CD34 truncations conferred lower activity, with the shortest promoter spanning -29 to +73 region exhibiting the lowest fluorescence signal. This finding suggested that the upstream promoter region containing PU.1 and Myb sites, as well as the CACCC box motif employed by Klf family transcription factors, are essential for high GFP expression levels. Finally, upon transduction into ES and MEF cells we found a small percentage of GFP positive cells,

indicating that the promoter activity is not restricted toward blood compartments but other cell types as well. To increase the reporter activity and enhance tissue specificity, we can perform additional truncation along the CD34 promoter or we can couple this promoter with another tissue specific promoter.

CD45 promoter is robust and highly tissue specific

CD45 levels were as anticipated significantly higher than any of the other genes, since this protein makes up approximately ten percent of protein on the surface of hematopoietic cells [145]. Importantly, CD45 levels were high in a panel of mouse and human blood cell lineages that included representatives of the lymphoid, myeloid and nucleated erythroid compartments, as well as the multipotent progenitors and HSPCs that specify these cell types **(Figure 15 and Figure 17)**. Levels of the hemato-endothelial marker Vav1 were approximately 10-fold lower than CD45 in all mouse and human hematopoietic lineages **(Figure 15A)**. CD34 is a marker widely used for human HSPC isolation [146]; however, our initial study using reporters based on CD34 and Sca-1 promoter showed that their promoters were not as strong and as tissue specific when compared to CD45 (results not shown) . The data suggested that CD45 was a strong hematopoietic reporter candidate.

CD45 gene regulatory element identification

Next, we analyzed global transcription factor occupancy following ChIP-seq and open chromatin from DNase-seq to discover the regulatory elements that controlled CD45 expression in the mouse **(Figure 16 and Figure 18)** and human CD45 gene **(Figure 19)**. DNase hypersensitivity sites (HS), transcription factor and RNA polymerase II chromatin occupancy, and the histone H3 lysine 4 trimethylation (H3K4me3) mark for gene activation were determined

in mature and progenitor B cell subsets, unstimulated T cells, nucleated erythroid cells (MEL), and a CD34 positive multipotent myeloid progenitor cell (416B) line **(Figure 16)**.

We focused our attention on the strong HS observed at the proximal promoter region encompassing the downstream alternate exon 1 as this is where expression of most CD45 transcription initiates from. We identified HS and chromatin occupancy of essential hematopoietic transcription factors that mark the CD45 core promoter region and overlap exon 1 and intron 1, (**Figure 16)**. PU.1, GATA1, c-Myb, Mafk and Mxi1 are all well-known regulators of multilineage blood cell specification and their high level enrichment can be observed at this region. GATA2, Tal1, and other regulators of HSPC specification were also found enriched within the CD45 gene and proximal promoter region **(Figure 18)**. Significant Pol2 and TATA-binding protein (TBP) occupancy was observed as expected and helped define the critical promoter for analysis. Collectively, the DNase-seq and ChIP-seq data confirmed that the region of interest most crucial for our CD45 reporter is largely confined to exon 1 and the first intron.

CD45 promoters efficiently mark mouse hematopoietic cells

We constructed lentiviral GFP reporters with 8 different CD45 promoter truncations to determine the minimal regulatory element configuration required to mark hematopoietic cells (**Figure 20)**. The regulatory region spanning promoters P1b, intron 1 and P2 (-164 to +168; -164/+168) contained dual PU.1, c-Myb and GATA binding sites, as well as other critical regulatory footprints. We used our same lentiviral transduction system employed previously for long-term (5-7 months) *in vivo* GFP marking and delivery of therapeutic protein [122, 147]. To test CD45 promoter strength we used a hematopoietic progenitor cell (EML), B cell (CH12.LX), and an erythroid progenitor (MEL) cell line. EML and CH12.LX lines are notoriously difficult to

transduce, but we were able to demonstrate high level marking in both lines that could be readily detected by fluorescence microscopy (**Figure 20B**). Following reporter introduction into the multipotent progenitor, B cell and erythroid cell lines, the eight CD45 reporters regions resulted in similar frequencies of GFP positivity (**Figures 21 & 22**), but notable differences in GFP mean fluorescence intensity (MFI) levels emerged that enabled us to better assess promoter activity. Flow cytometric analysis demonstrated that constructs consisting of regions -164 to +99 (-164/+99), -164 to +168 (-164/+168) and -112 to +168 (-112/+168) had the highest GFP intensity (**Figure 20C-E)**. Minimal constructs, including reporters with promoters comprising only -112 to +4 or -99 to +4, were less efficient in promoting high GFP levels. We also stained the murine lymphoid (CH12LX) and HPC (EML) cell lines to demonstrate that these express CD45 at high level; upon the transduction of the promoter CD45 activity comparable to CD45 surface protein level (**Figure 23**). This finding corroborates the requirement for sequences that are bound by GATA1, c-Myb, Mafk, Mxi and PU.1 [148] to enable maximal promoter strength and GFP marking (**see Figure 16**).

CD45 promoter activity is evolutionarily conserved in human hematopoietic cells

Since the CD45 promoter region of study is well-conserved, we tested the efficiency of reporter marking in human cells. The eight CD45 reporters were used to transduce human multipotent progenitor (KG-1), nucleated erythroid (K562), T cell (Jurkat) and B cell (Raji) lines **(Figure 24, Figures 25 and 26)**. Constructs lacking upstream and/or downstream regulatory elements were generally weaker across the different cell lines. Regions between -164 to +99, and -164 to +168 delivered maximal GFP levels, confirming the trend observed in mouse cell lines. The downstream PU.1 site and transcription factor binding sites resident in exon 1b were

essential for maximal GFP expression. To ensure that our hematopoietic reporters would be suitable for human cellular reprogramming strategies, we replaced the mouse promoter sequence (-164 to +168) with the human promoter element following a multi-vertebrate species alignment (data not shown). We used the human CD45 reporter to transduce the human multipotent progenitor and B cell lines (**Figure 26, panels E-F**). GFP levels observed were comparable to or exceeding those documented following transduction with the mouse reporters. We note that reporter strength may be underestimated due to lower CD45 promoter outputs from several of the human cell lines we used here, as documented by their lower CD45 expression levels (see **Figure 16**). Collectively, marking measured by GFP intensity exhibited a similar trend to the mouse cell lines, indicating that the regulatory elements utilized are evolutionarily conserved.

CD45 reporter is active in primary myeloid and lymphoid blood cell lineages

Based on our results from transduction of hematopoietic cell lines, we selected the -164/+99 promoter driven CD45 reporter to assess reporting in a primary B cell culture derived from mouse BM as illustrated in Figure 4A. The enriched HSPCs from BM were transduced and stimulated with cytokines to differentiate them into myeloid and lymphoid lineages [149, 150]. After 10 days, B cells (B220), pro-B cells (CD19), and myeloid cells (Gr-1/Ma-1) were assessed. High-level GFP marking and expression were observed with the CD45 reporters via confocal microscope **(Figure 27B)**. The reporter marked approximately 70% of these cells as seen in the first panel of **Figure 27C**. This population was sub-gated to show 94% of the cells were Gr-1/Mac-1 positive (myeloid markers) in the middle panel. The remaining Gr-1/Mac-1 negative cells were B220 and CD19 positive. The reporter transduced cells exhibited the strong GFP levels that would be required for an effective reporter. To promote robust B cell differentiation,

SCF was removed from the culture medium. At day 24 these cells were assayed again for B cell identity; as expected, complete B cell differentiation was observed. The ratio of GFP positivity remained unchanged with the exception that there was a much better separation between the non-transduced cells (negative), and transduced cells (positive). All of the GFP positive cells were mature B cells, as demonstrated by double positivity for B220 and CD19 **(Figure 27D)**. The data suggest efficient reporting in primary B cells.

CD45 promoter activity is restricted to hematopoietic cells

To validate the tissue specificity of our CD45 reporter activity, we transduced fibroblasts and ESCs; neither should be permissive for CD45 reporting potential based on gene expression analyses **(see Figure 15)**. These cell types were selected because they are the substrate for direct reprogramming [99, 105] and pluripotent cell based differentiation [101, 102] respectively. After a 48 hour transduction period, cells were assayed for GFP expression using FACS. Transduction of mouse embryonic fibroblasts (MEF) with a ubiquitous CMV promoter at low multiplicities of infection (MOI 1) typically used in reprogramming experiments resulted in approximately 40% marking **(Figure 28)**. In contrast, our strongest CD45 reporter did not result in marking above baseline levels even 5 days post transduction (data not shown). Similarly, no evidence of GFP marking was observed by the CD45 reporter in ESCs (**Figure 28C**, compare CMV driven EGFP in the top right and CD45 in the bottom panels). To further demonstrate that CD45 promoter activity is restricted to blood cell lineages, ESCs that were transduced with the CD45 reporter were differentiated into HSPCs via embryoid body formation and subsequent OP9 stromal line co-cultivation **(Figure 28D)**. Following EB formation and 10 days on OP9, cells were assayed for the presence of the cell surface hematopoietic markers Sca-1, c-Kit, CD45, and Gr-1/Mac-1. Sca-1 and c-Kit expression represents enrichment of HSPCs and Gr-1/Mac-1 serves as the

predominant marker of the myeloid lineage. Prior to differentiation, these ESCs were transduced with HOXB4-GFP for increasing hematopoietic specification following differentiation. The HOXB4-GFP marker also facilitates isolation of ESC derived lineages from the OP9 stromal layer. Since the GFP marker was embedded in the HOXB4 expression cassette, we employed an mCherry reporter to detect CD45 positivity in these experiments. HOXB4-GFP transduced cells were then subgated on mCherry and assessed for lineage marker expression. At this early time point approximately 27% of transduced cells were mCherry positive and these cells exhibited significant marking for Sca-1, c-kit, CD45, and Gr-1/Mac-1. These results highlight the multilineage marking capability of the CD45 reporter after ESCs are differentiated.

CD45 reporters efficiently mark blood cells following transplantation

To confirm that our CD45 reporter could effectively mark multilineage hematopoiesis *in vivo*, bone marrow from CD45.1 mice (C57BL/6.SJL) was harvested and enriched for HSPCs . We transduced lineage depleted HSPCs with the three strongest CD45 reporters. As expected, approximately seven percent of our lineage negative population exhibited double positivity for Sca-1 and c-Kit **(Figure 31A,B)**, confirming enrichment of an HSPC population. HSPCs were injected into sublethally irradiated CD45.2 mice, and reconstituted the host blood system by 4 weeks post transplantation **(Figure 29)**. For the first round, we added non-irradiated and non-transduced bone marrow to the host mice. Thus, approximately 50% of host peripheral blood cells were from the donor ($CD45.1^{+}$), and out of this population GFP marking remained high 5 months post transplantation in T lymphocytes, myeloid cells (Gr-1/Mac-1), nucleated erythroid cells (Ter119) and B lymphocytes (B220). This percentage of marking was even higher in reconstitution experiments performed without the addition of supportive bone marrow, resulting

in persistent GFP labeling as high as 98% of cells in the peripheral blood of reconstituted recipients (see the lower left panel of **Figure 29D and Figure 26C)**. The CD45 promoter was also able to support erythroid (Ter119), myeloid (Gr-1/Mac-1), and lymphoid (B220/CD3ε) lineage marking in the bone marrow, spleen and thymus as indicated in **Figure 30**, confirming that CD45 promoter elements contained within our reporter are sufficient to reliably mark hematopoiesis in lymphoid, myeloid and nucleated erythroid blood cells. To show long term marking, CD45 reporter transduced bone marrow from primary recipients was harvested and transplanted into sublethally irradiated secondary recipients. Again, significant reconstitution was achieved with GFP levels remaining high (95-98%) for up to 22 weeks post-transplant (**Figure 30D and Figure 28).** This result demonstrates the utility of the CD45 reporting system to serve as an efficient method for documenting stable long-term hematopoietic specification *in vivo*.

Discussion

All other direct reprogramming platforms have relied on a GFP reporter to assess cellular respecification, clonality, and the efficiency of reprogramming [106, 107, 151], corroborating the utility of this strategy for blood cell derivation. We considered Sca-1 and CD34 gene regulatory elements (unpublished results) as the foundation for our reporter, but they were either not tissue specific, were not expressed in all blood cell lineages [152], or lacked marking strength. In contrast, CD45 was exquisitely restricted to blood cell lineages **(Figure 15B** and reference 24**)**, making it the most viable candidate that we investigated.

Here we refined efforts from previous studies [134, 135] to locate critical CD45 regulatory elements and identified a 300bp region based on binding signatures for TBP and the blood cell

specific transcription factors PU.1, GATA1, c-Myb, MafK and Mxi1. Our results showed that constructs spanning from -164 to +99 of the promoter region drove the highest GFP levels in murine and human cell lines (**Figures 20 - 25**); this confirms regulatory element activity within this gene promoter is well conserved in mammals. PU.1, Myb, CEBPα/β and GATA binding footprints are required for maximal marking of hematopoietic cells *in vivo,* supporting the contribution of these critical transcription factors for blood cell specification. This is the first analysis to employ ChIP-seq and DNase-seq for fine mapping and identifying DNA elements sufficient for a lineage specific reporter. Importantly, our reporter is small enough that it can be delivered by traditional lentiviral gene transfer or other genome editing approaches [122, 153].

Using ubiquitous promoters and a high MOI have been fairly effective (25-86% marking) in transducing blood cells in the past [147, 154]. However, our blood cell restricted system completely marked the hematopoietic compartment when no additional non-transduced supportive bone marrow was given (98%; see **Figure 28C, Figure 29D** and **Figure 32**). We also see high level marking in the spleen, thymus, and bone marrow (**Figure 30**). Our *in vivo* analyses demonstrated that the lentiviral CD45 reporter marked blood cells persistently without silencing, evidenced by the robust staining of hematopoietic cells 22 weeks post-secondary transplantation. This has been accomplished using a methodology that results in a significantly lower viral load [122]. Importantly, the CD45 reporter is not active in ESCs, and is only activated after these cells had become blood cells (**Figure 28**). We acknowledge in bone marrow transplantation there was about lessen 5% of cells that do not express GFP. We think as a result of incomplete transduction of the lentiviral or the promoter is not as permissive in these cells as compared to others. Regardless, our results suggest that the CD45 reporter system can be used monitor the efficiency of blood cell differentiation from pluripotent cells, to exclude undifferentiated cells from

hematopoietic engraftment, and to serve as a simple strategy to follow blood cell transplantation kinetics and persistence via bioluminescence by tethering CD45 regulatory elements to luciferase [155, 156]. The reporter can also facilitate tracking of normal and malignant lymphoid cells to interrogate the effects of a genetic loss or gain of function in mouse models following adoptive transfer [157]. Additionally, TALEN or CRISPR-Cas9 [153] nucleases can be used to introduce this reporter, together with polycistronic reprogramming vectors, into genomic safe harbors [158]. Finally, hematopoietic marking can be coupled with a potent suicide gene we have used previously [159] when driven by a pluripotency gene promoter, to selectively destroy undifferentiated cells and further augment levels of biosafety. With the recent derivation of naïve pluripotent human ES and iPSC lines that exhibit characteristics of ground state pluripotency [160] it should now be more feasible to generate functional adult hematopoietic cells for the treatment of human blood diseases using this CD45-based GFP marking system.

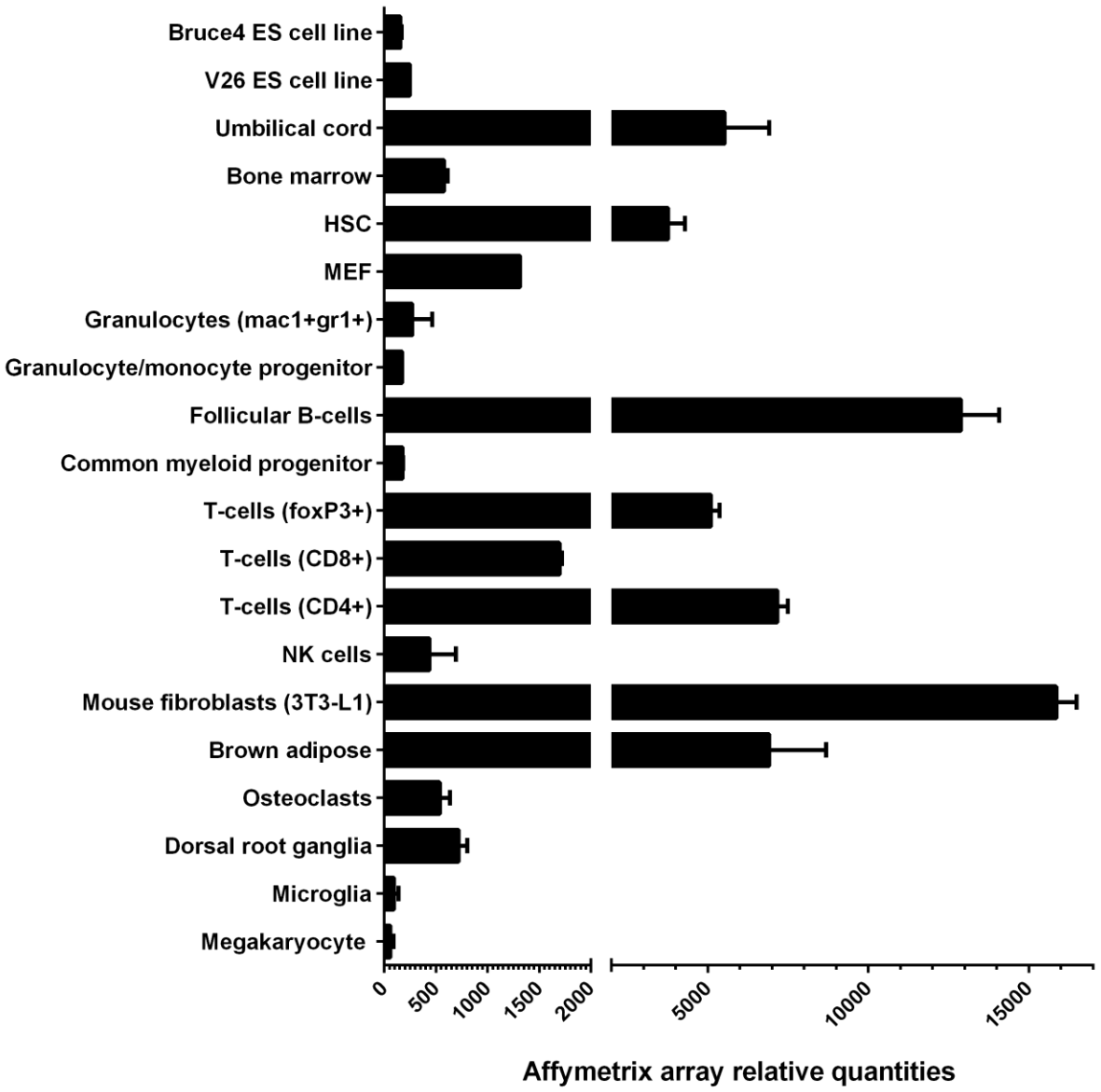


Figure 7. Sca-1/Ly6A are high in multiple blood cell lineages**.** Affymetrix high-density oligonucleotide arrays [120, 139] were used to measure Sca-1/Ly6A levels in non-hematopoietic and hematopoietic cell types. Values shown are from duplicate arrays.

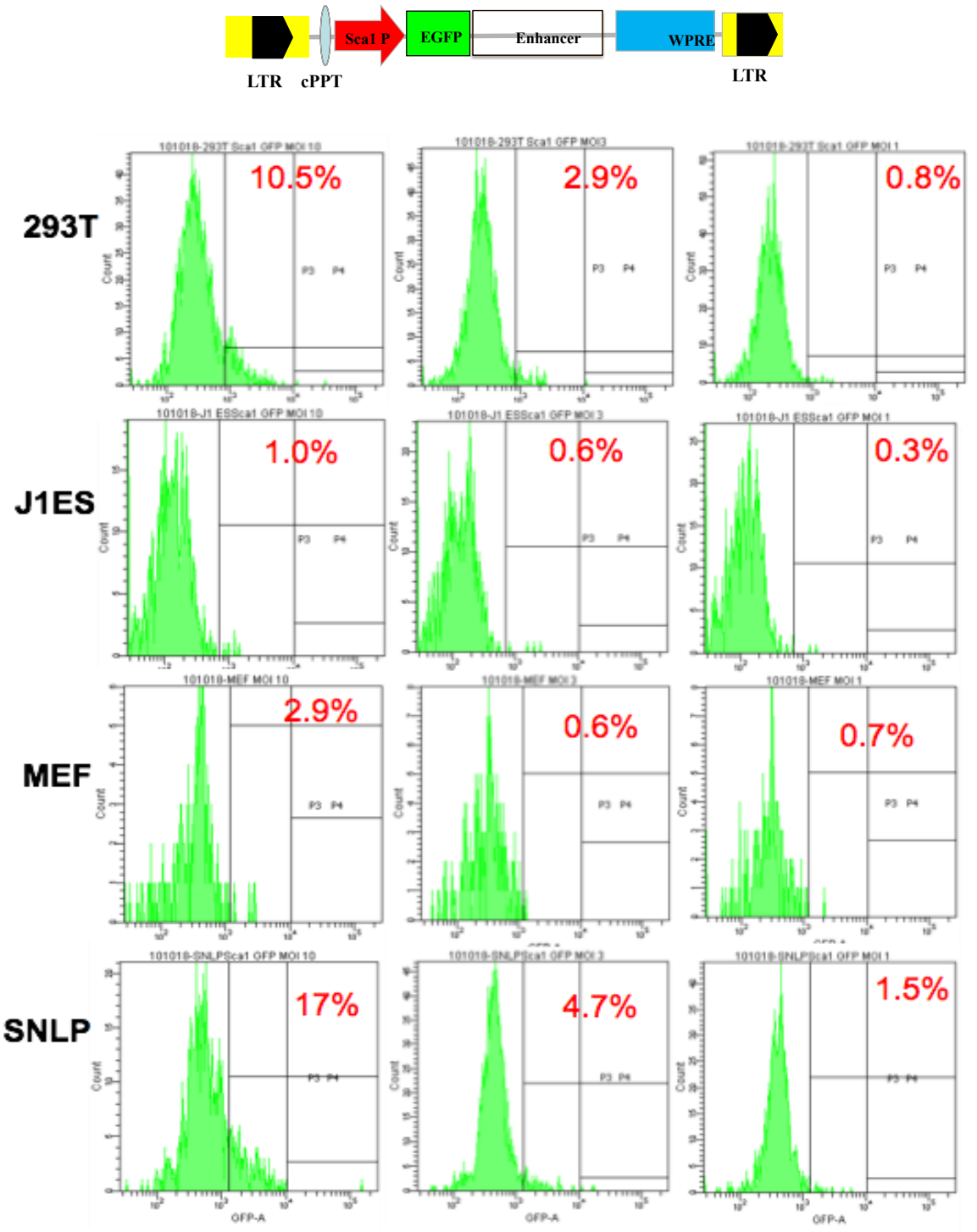


Figure 8. Sca-1 promoter is not active in non-hematopoietic cell lines. Sca-1 promoter design employed in this study (top) to transduce 293T, J1 ES, MEF, and SNLP. Histograms showing GFP positivity after 48 hrs post-transduction with Sca-1 GFP lentivirus at MOI of 10 (left column), MOI of 3 (center column), and MOI of 1 (right column).

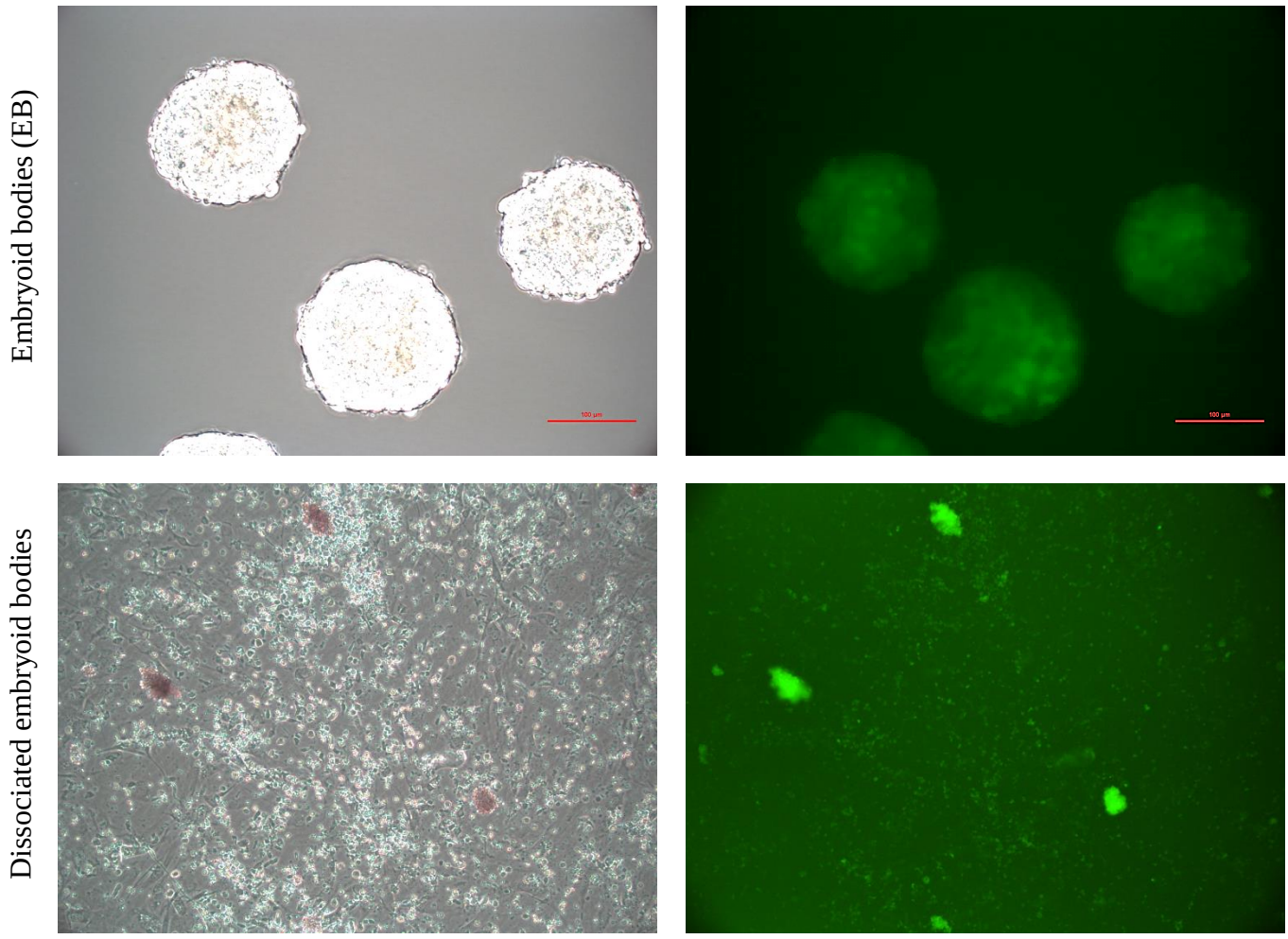


Figure 9. Sca-1 reporter activity in ESCs differentiation. ESCs were transduced with Ly6A/ Sca-1 reporter 48 hours prior to embryoid body formation (top row) and hematopoietic differentiation (bottom row).

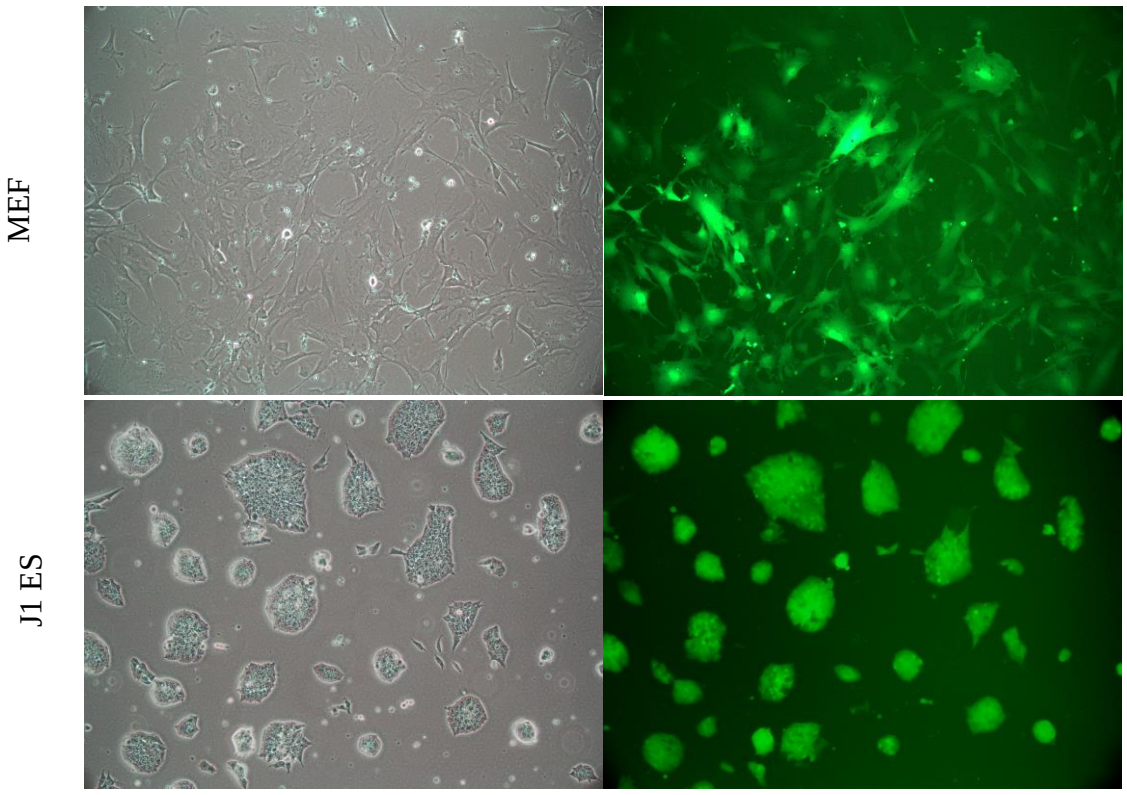


Figure 10. Ly6a/ Sca-1 expression is promiscuous. Bright field and fluorescence images of MEF (top row) and J1 ES (bottom row) after transduction of Ly6A/Sca-1 promoter for 48 hours followed by puromycin selection for 7 days.

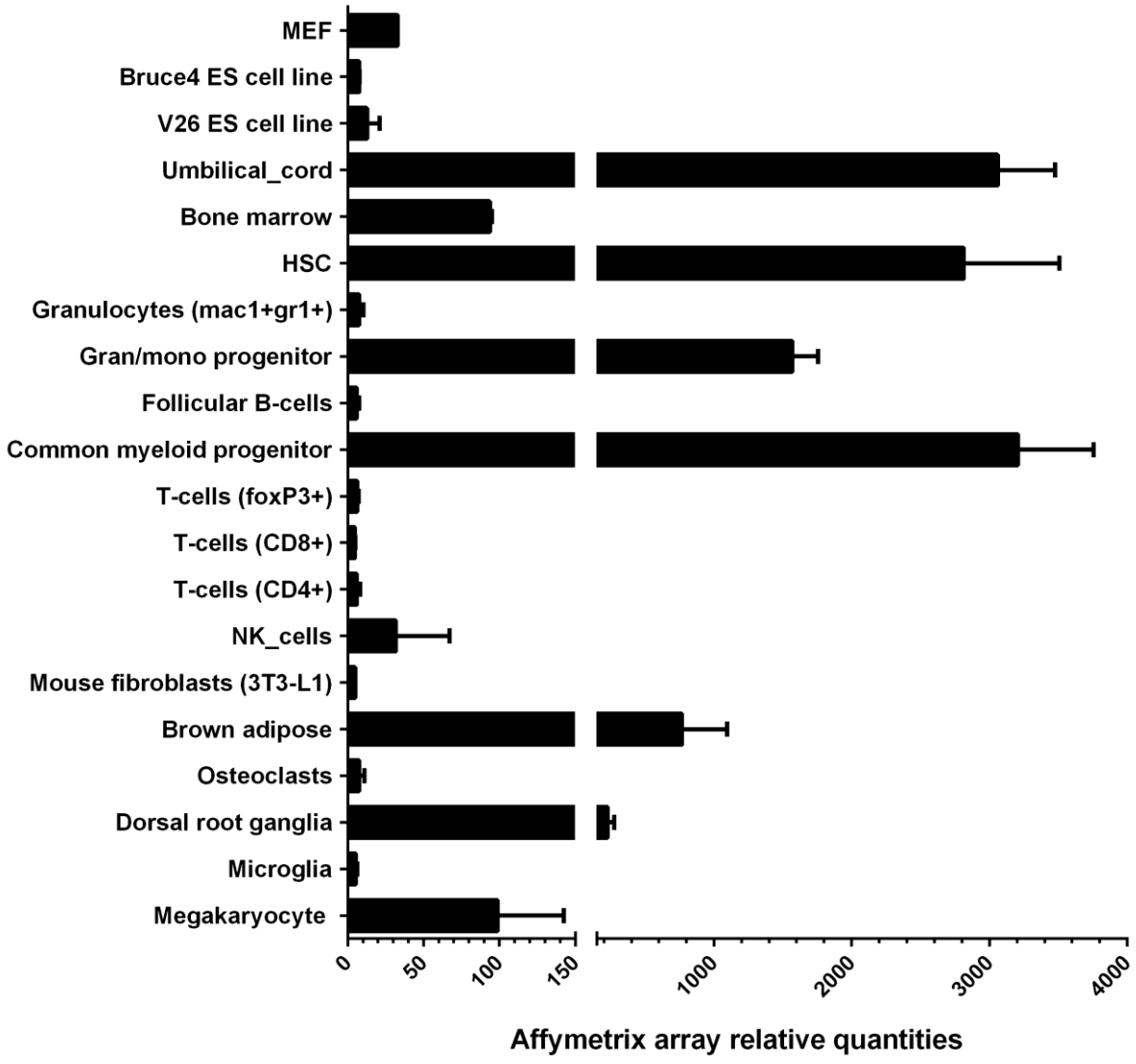


Figure 11. CD34 expressed high in multiple blood cells lineages, but low ESCs and MEF**.** Affymetrix high-density oligonucleotide arrays [120, 139] were used to measure CD34 levels in non-hematopoietic and hematopoietic cell types. Values shown are from duplicate arrays. Values shown are from duplicate arrays.

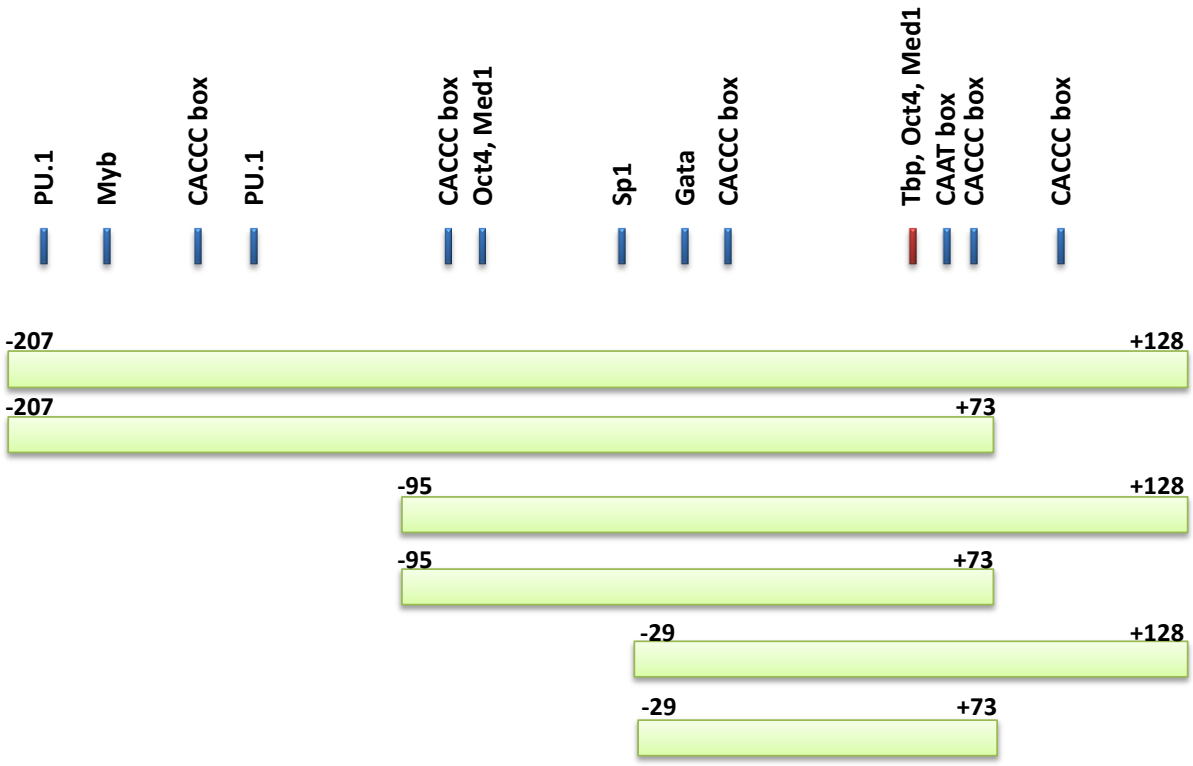


Figure 12. CD34 promoter constructs. Based on ChIP-seq and DNAse-data we have designed the 6 promoter constructs around the region with the highest regulatory elements. The longest construct span from -207 to +128 to the smallest construct spanning from -29 to +73.

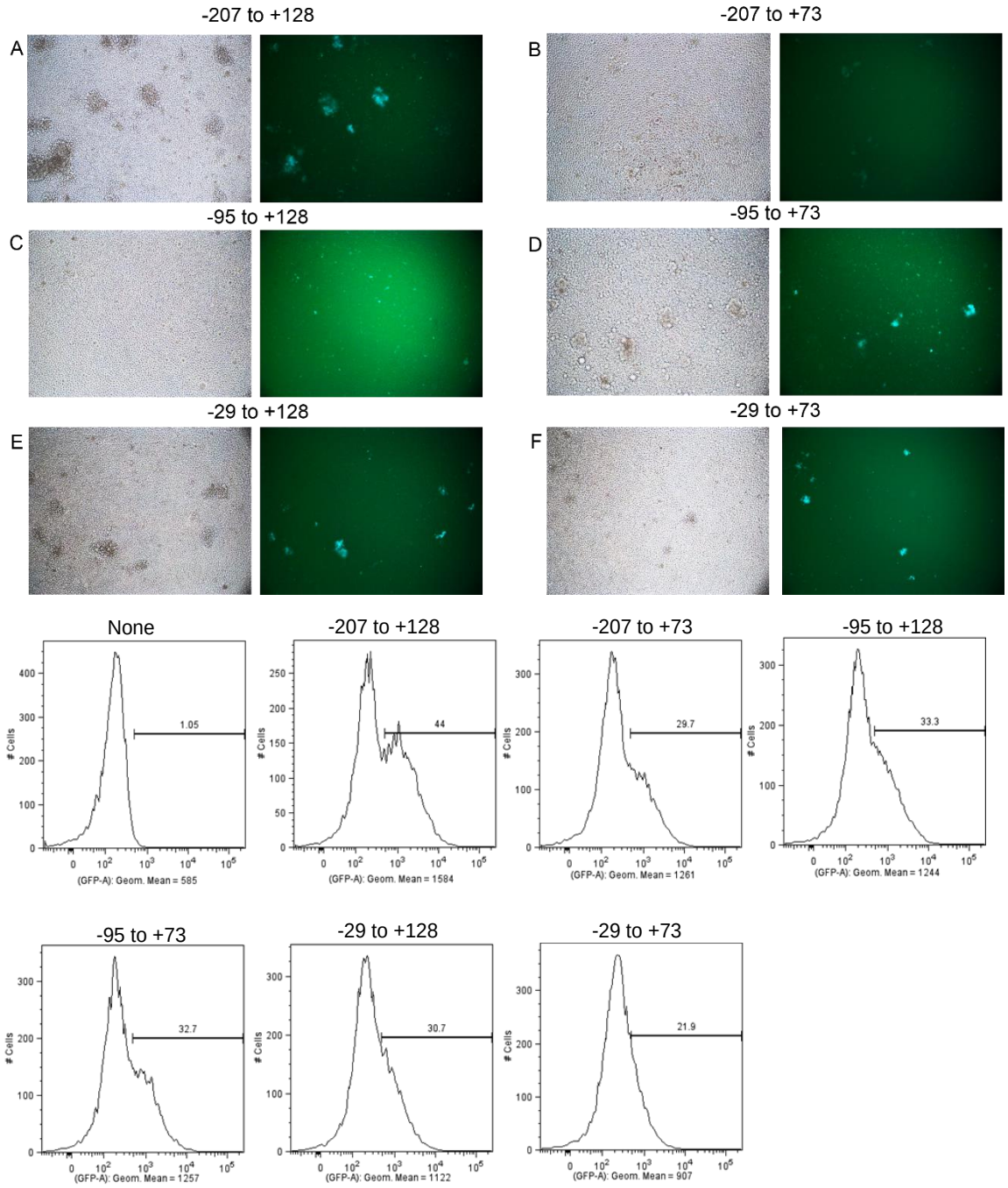


Figure 13. CD34 promoter transduction in KG1a cells**.** Top panels were taken 72 hours post-transduction (A-F). Histograms of cells analyzed via flow cytometry for GFP expression (bottom).

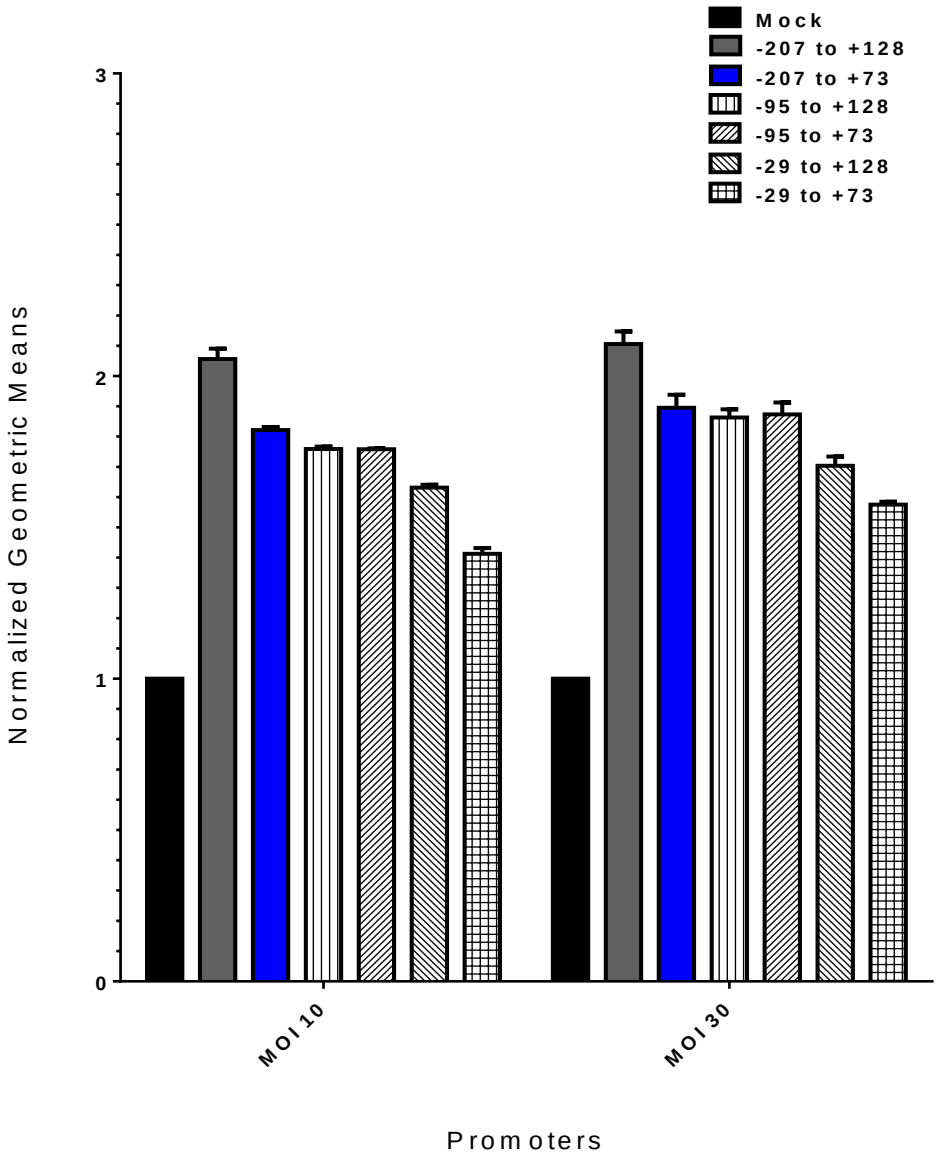
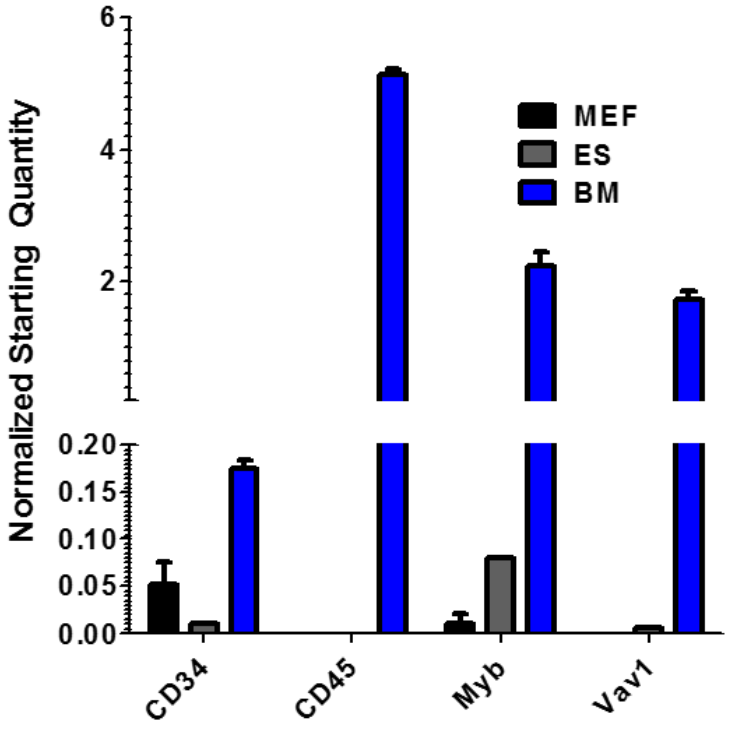


Figure 14. CD34 promoter activity is not as robust. Normalized geometric intensity of the CD34 promoter activity via GFP expression at 72 hrs post transduction at MOI 10 and 30.

A


Normalized Starting Quantity
6
4
2
0.20
0.15
0.10
0.05
0.00
MEF
ES
BM
CD34
CD45
Myb
Vav1

B

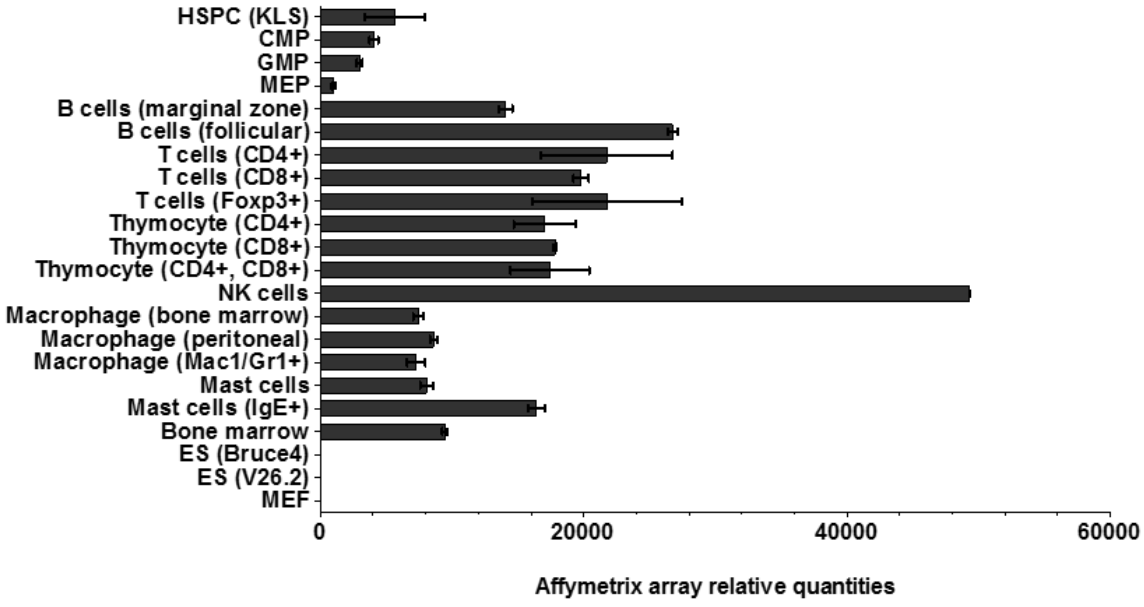

HSPC (KLS)
CMP
GMP
MEP
B cells (marginal zone)
B cells (follicular)
T cells (CD4+)
T cells (CD8+)
T cells (Foxp3+)
Thymocyte (CD4+)
Thymocyte (CD8+)
Thymocyte (CD4+, CD8+)
NK cells
Macrophage (bone marrow)
Macrophage (peritoneal)
Macrophage (Mac1/Gr1+)
Mast cells
Mast cells (IgE+)
Bone marrow
ES (Bruce4)
ES (V26.2)
MEF
0
20000
40000
60000
Affymetrix array relative quantities

Figure 15. Hematopoietic reporter candidates**.** (**A**) Quantitative real-time PCR (qRT-PCR) expression levels of highly hematopoietic or hemato-endothelial (Vav1) enriched genes. Murine embryonic fibroblasts (MEF), embryonic stem (ES) cells and whole bone marrow were analyzed. Levels were normalized to Hprt and Gapdh. (**B**) Affymetrix array measured CD45 expression levels in hematopoietic progenitor and differentiated cell lineages. Kit+, lineage-, Sca-1+ (KLS), Common myeloid progenitor (CMP), Granulocyte-macrophage progenitor (GMP), and Megakaryocyte-erythroid progenitors (MEP) are also indicated. Embryonic stem (ES) and murine embryonic fibroblasts (MEF) are included as non-hematopoietic lineages that do not express CD45.

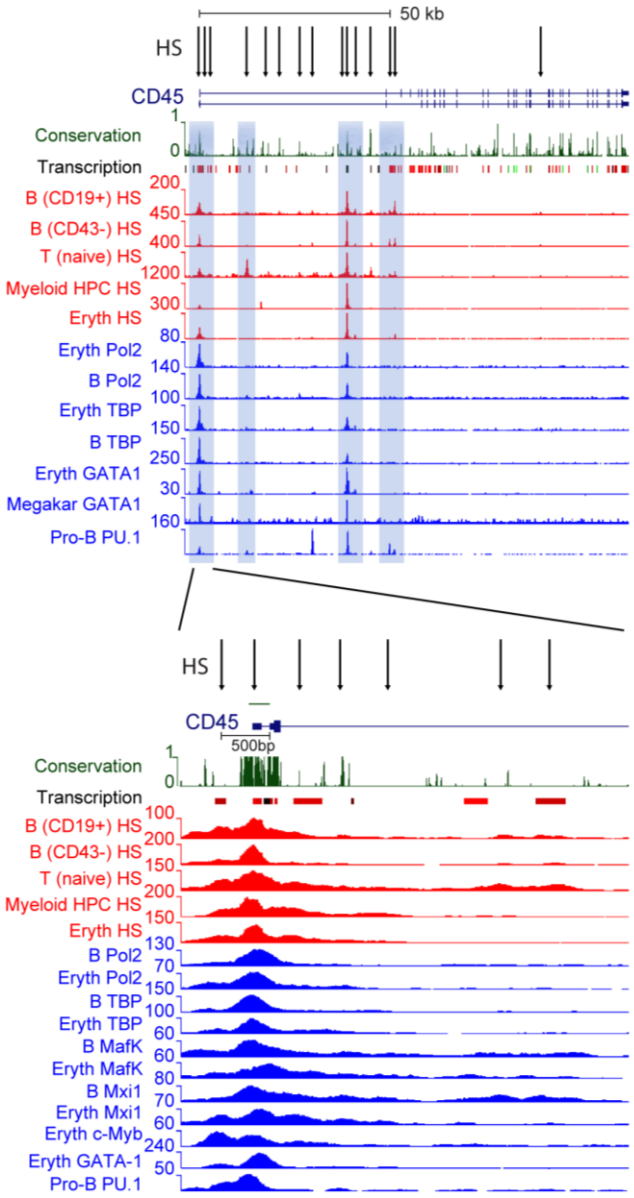


Figure 16. CD45 regulatory element discovery and promoter characterization**.** The full-length CD45 gene with its two most common alternative transcripts is shown in the top panel, and an enlarged view of the proximal promoter region can be seen at the bottom. Exons 1b and 2 are drawn as blue boxes. The region mapped for promoter activity spanning -164 to +168 (-164/+168) is shown with a green line over it. Only the 15 HS that were shared among 3 or more of the 6 cell lines with DNase-seq data are shown. Of these 15 HS, 6 were shared across all 6 lines, and all but 1 were shared by 4 or more cell lines. Vertebrate conservation and Affymetrix exon array enrichment for hematopoietic tissues (labeled Transcription) are indicated immediately below the gene structure, with lines in red depicting high-level expression. Hematopoietic cell regulatory elements marked by DNase hypersensitivity (HS) and transcription factor occupancy were obtained from ChIP-seq data. HS are marked by arrows.

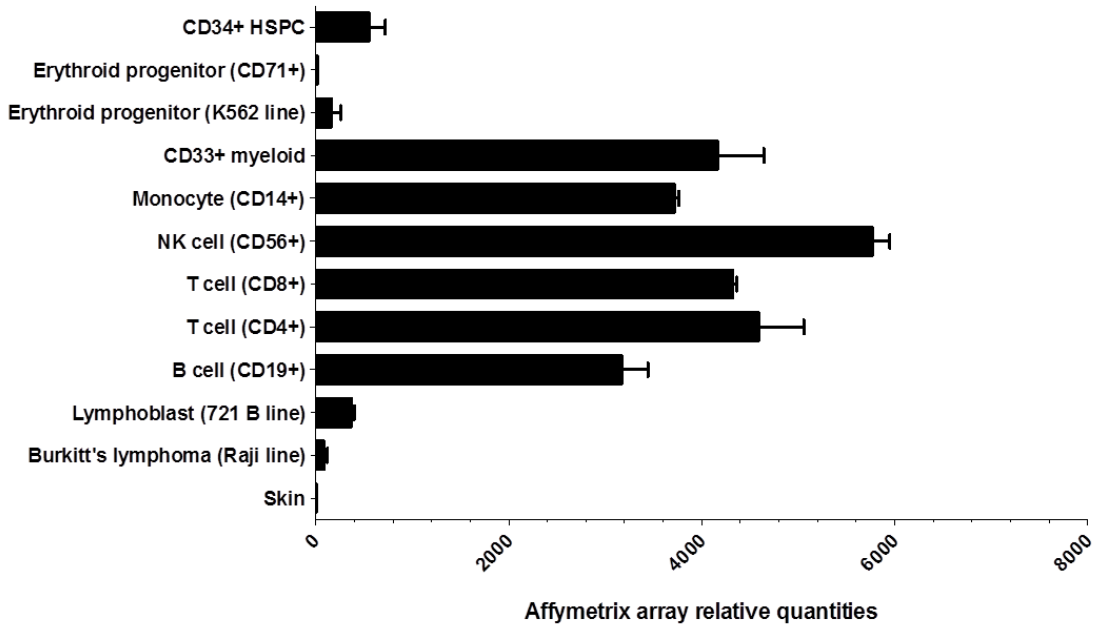


Figure 17. Human CD45 expression levels are high in multiple blood cell lineages. Affymetrix high-density oligonucleotide arrays [120, 139] were used to measure CD45 levels in the human blood cell lineages and cell lines indicated. Values shown are from duplicate arrays.

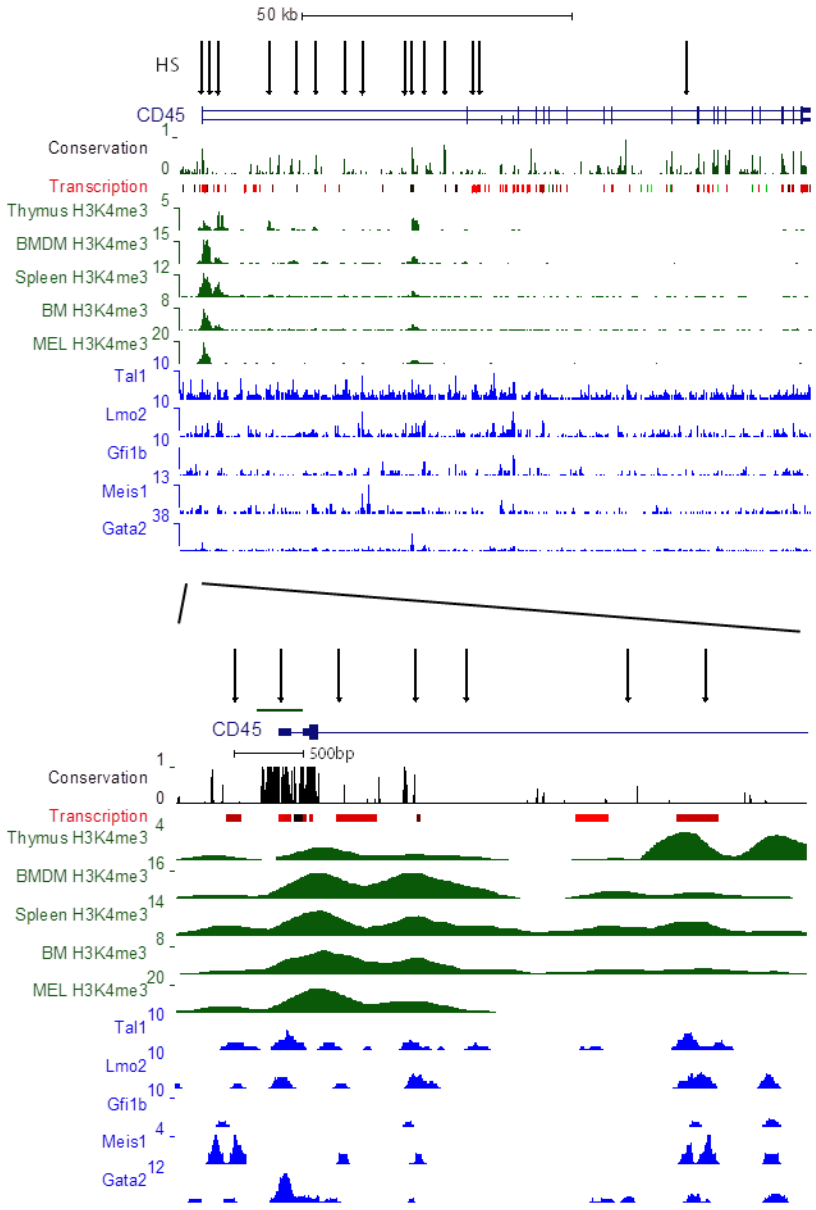


Figure 18. Mouse CD45 regulatory element discovery using transcription factor chromatin occupancy and nuclease sensitivity mapping. Chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) and nuclease accessibility (DNase-seq) analyses were used to identify regulatory elements throughout the murine CD45 gene body (top) and a region confined to the proximal promoter and portion of intron 2 (bottom). Transcription factor ChIP-seq data were obtained from the multipotent hematopoietic progenitor cell line 7 (HPC-7) [161]. H3K4me3 enrichment and associated HSs downstream of exon 2 are observed in the thymus, bone marrow derived macrophages (BMDM), spleen and the erythroid MEL cell line.

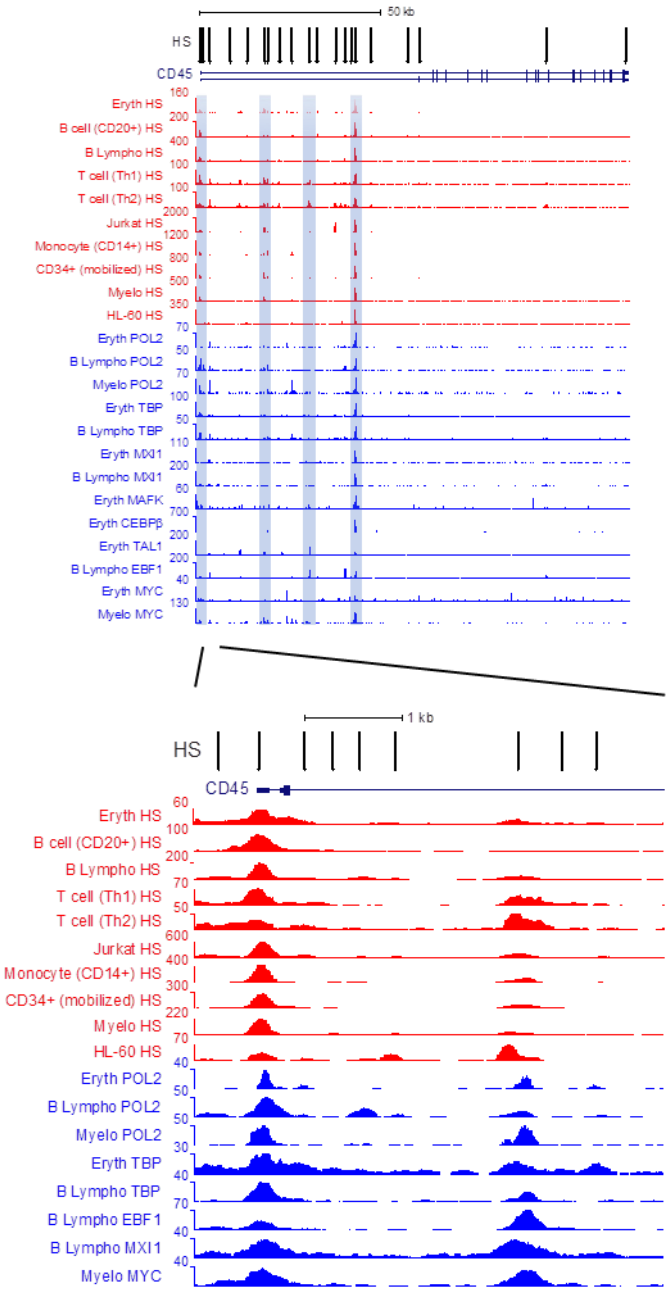


Figure 19. Human CD45 regulatory element discovery using chromatin profiling and nuclease sensitivity mapping. Chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) and nuclease accessibility (DNase-seq) analyses were used to identify regulatory elements throughout the CD45 gene body (top) and a region confined to the proximal promoter and portion of intron 2 (bottom). The human erythroid progenitor cell line K562 (eryth), acute promyelocytic leukemia cell line NB4 (promyelo), EBV transformed B lymphoblastoid cell line GM12864 (B lympho), promyelocytic leukemia cell line HL-60, T lymphocyte line Jurkat and primary human cell lineages queried for ChIP-seq and DNase-seq are indicated.

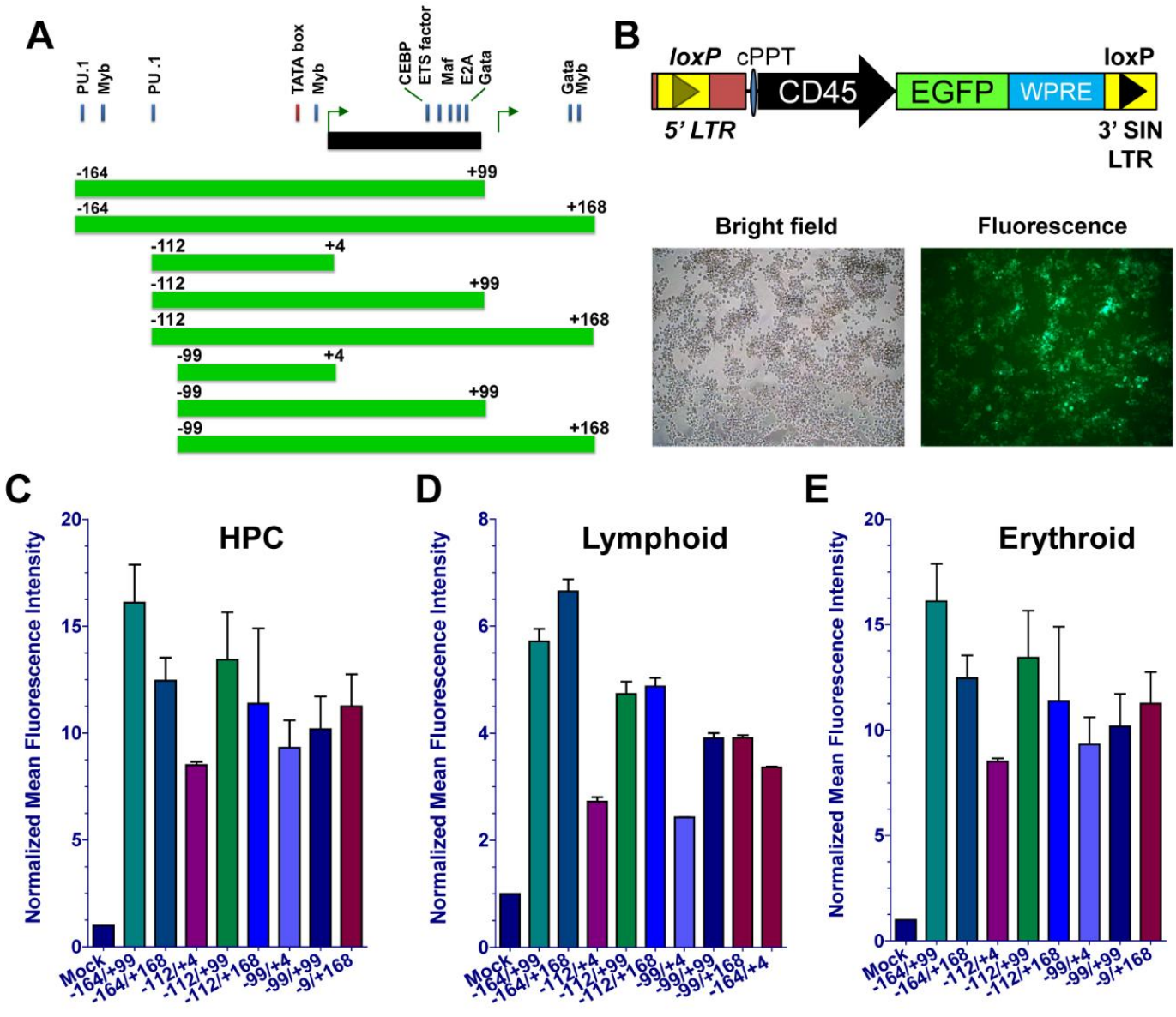


Figure 20. CD45 reporters mark mouse hematopoietic cell lines**. (A)** Promoter design scheme is shown for CD45 reporter truncations based upon critical transcription factor binding sites. **(B)** Lentiviral vector design (top) used to transduce cells is shown. Representative images of a transduced B cell line following bright field and fluorescence microscopy (bottom) are depicted. The geometric mean was used to measure MFI, and GFP levels are normalized to mock as presented. **(C)** lymphoid B cell line (CH12Lx), **(D)** hematopoietic progenitor cell (HPC) line (EML), and **(E)** nucleated erythroid cell line (MEL). Transductions were done in duplicate and repeated with a minimum of 3 independent times. MFI = mean fluorescence intensity; GFP = green fluorescent protein.

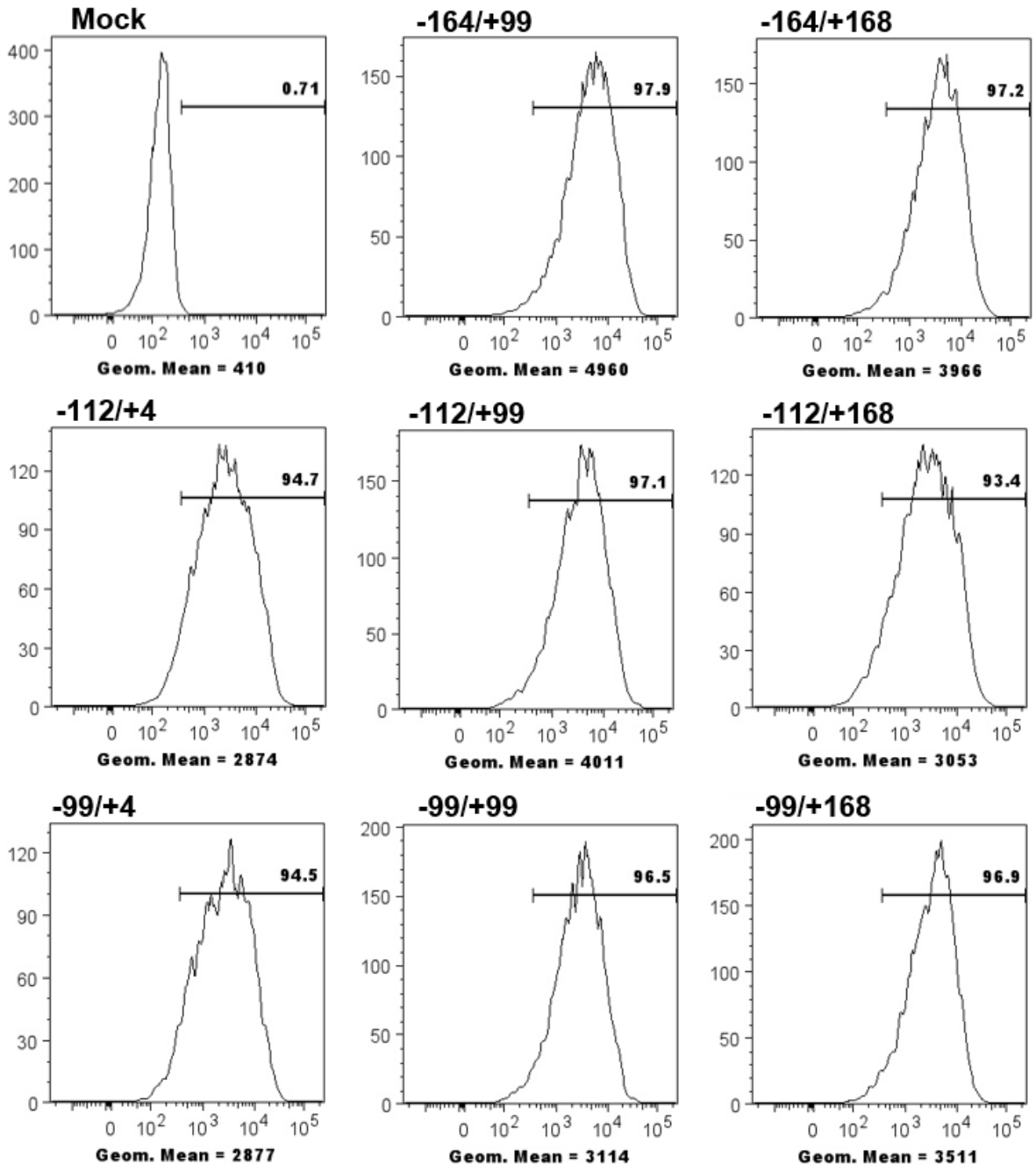


Figure 21. CD45 reporters efficiently mark mouse blood cell line**.** Histograms of murine nucleated erythroid progenitor cell transduced with the CD45 reporter constructs indicated. Green fluorescent protein (GFP) percentage and fluorescent intensity (geometric means) were calculated for each construct.

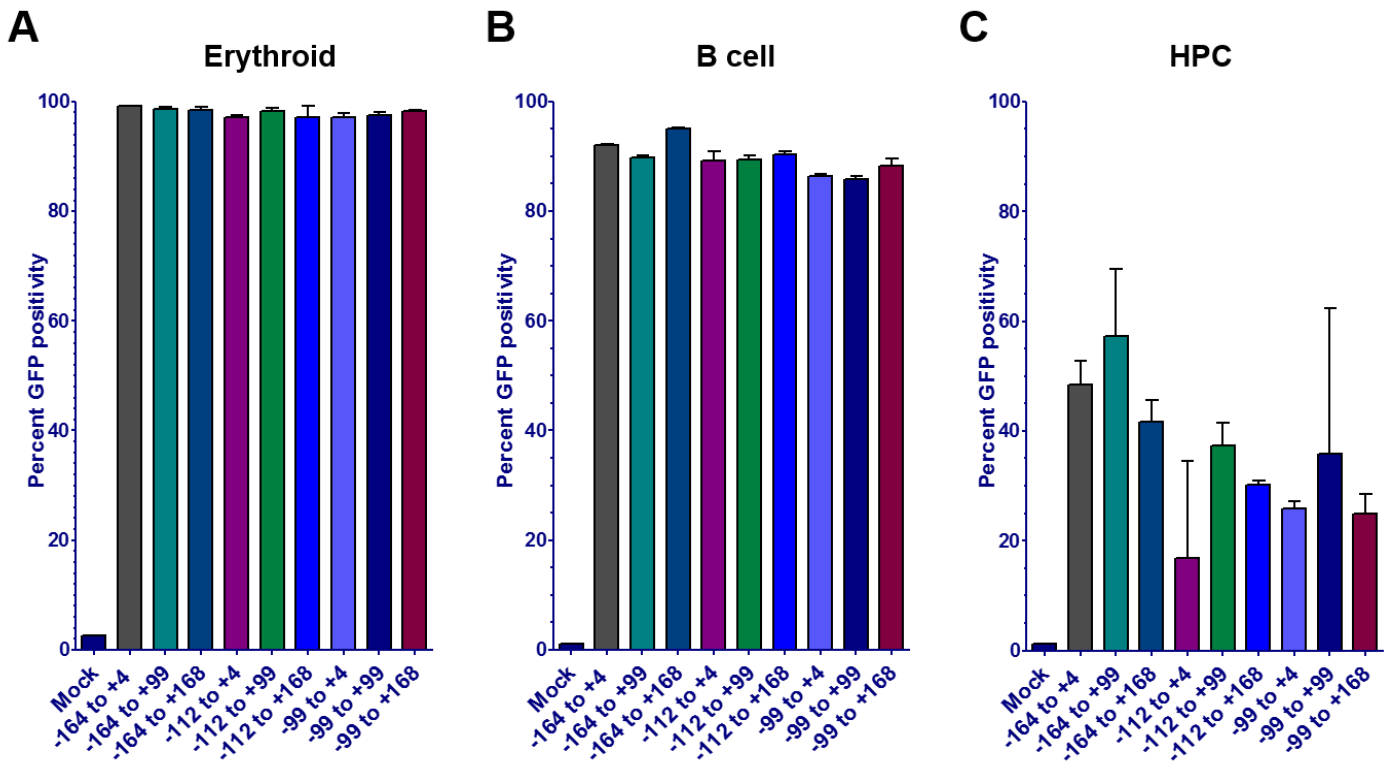


Figure 22. CD45 reporters efficiently mark mouse lymphoid, erythroid, and multipotent progenitor blood cell lines**.** Green fluorescent protein (GFP) percentage was calculated for each construct. (**A-C**) Percent GFP positivity for (**A**) erythroid progenitor (MEL), (**B**) B cell (CH12Lx), and (**C**) a hematopoietic progenitor cell (HPC) line transduced with the indicated CD45 reporters. Fluorescence-activated cell sorting measured analyses were done in duplicate and repeated a minimum of three times.

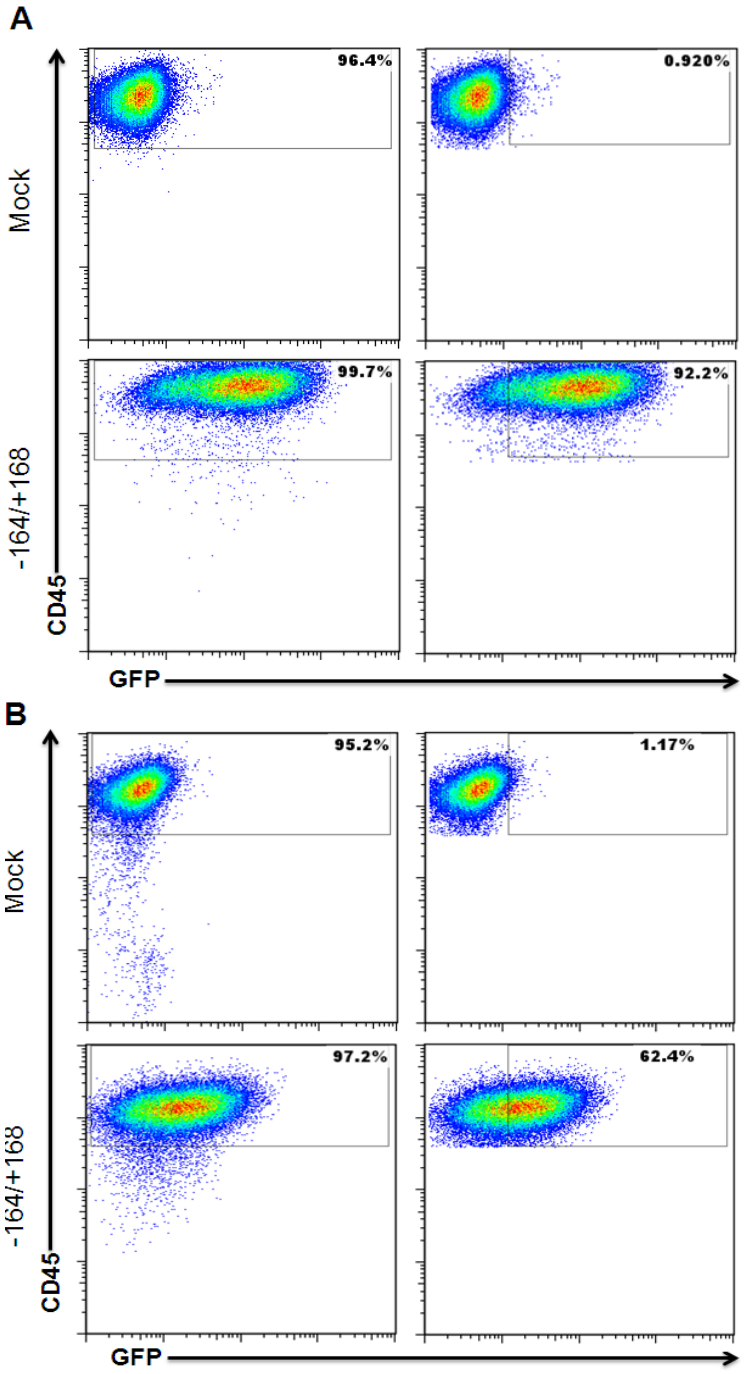


Figure 23. CD45 expression and its promoter activity in murine cell lines. Flow images of murine lymphoid (CH12LX) (**A,** top panel) and HPC (EML) (**B,** bottom panel) cell lines staining for CD45 without or with CD45 promoter transduction.

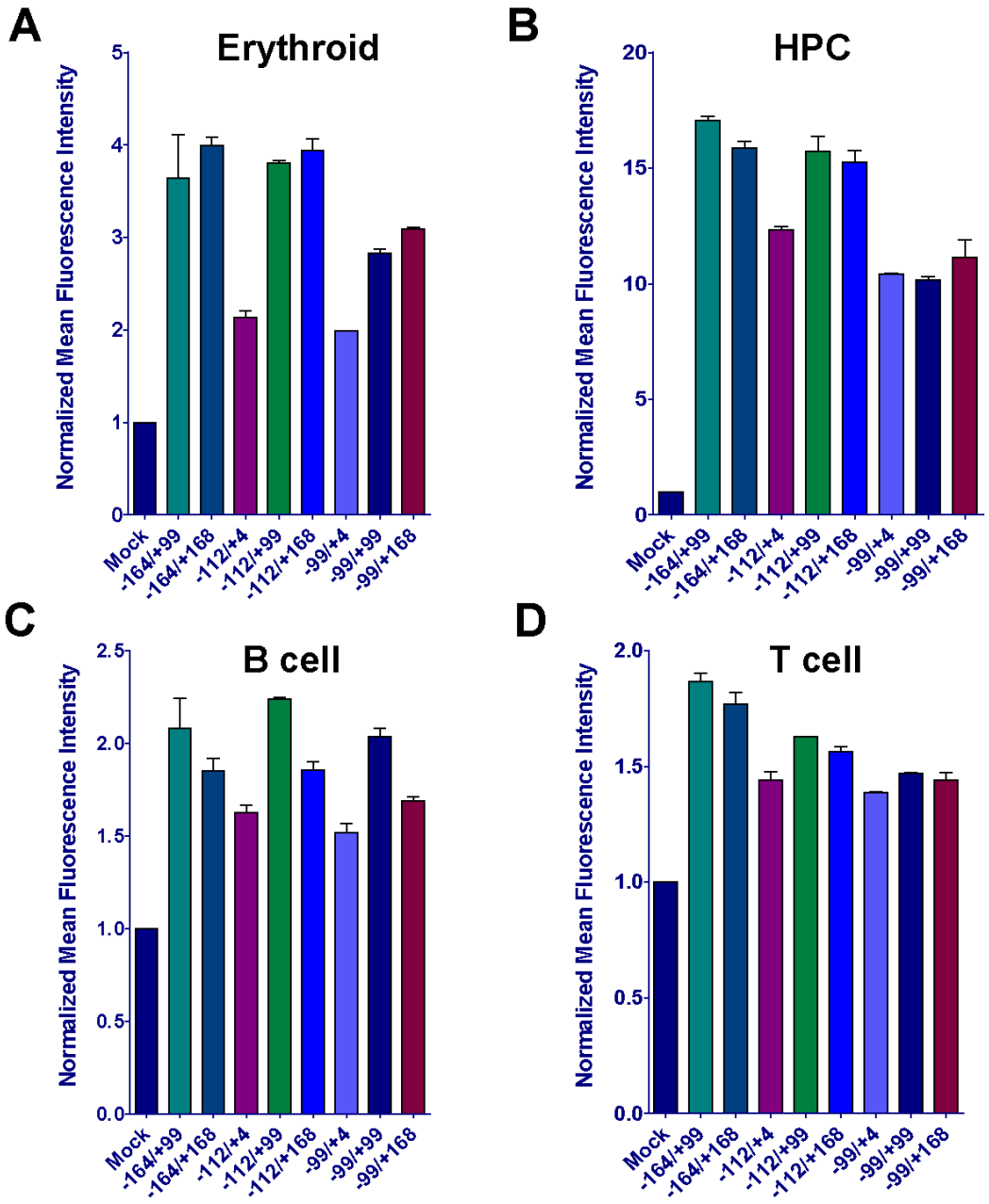


Figure 24. CD45 reporters mark human hematopoietic cell lines. **(A)** Normalized MFI calculated from flow cytometry for erythroid, **(B)** HPC, **(C)** B cell (**C**), and **(D)** T cell lines. Transductions were done in duplicate and repeated a minimum of 3 independent times. MFI = mean fluorescence intensity; HPC = hematopoietic progenitor cell.

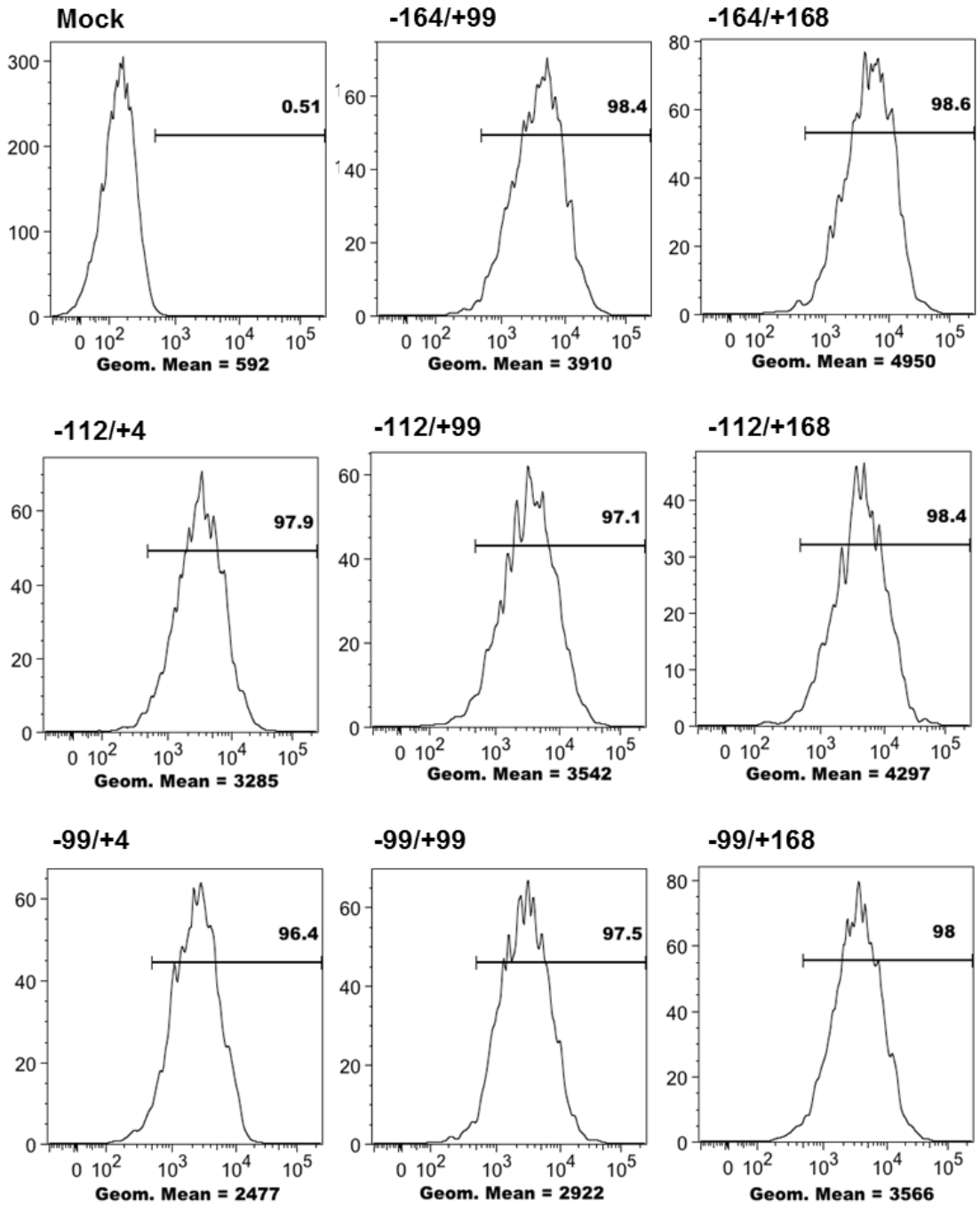


Figure 25. CD45 promoter is active in human hematopoietic cells**.** Histograms of a human T cell line that was transduced with the CD45 reporter constructs indicated. GFP percentage and mean fluorescent intensity (geometric mean) were calculated for each construct. FACS measured analyses were done in duplicate and repeated a minimum of 3 times.

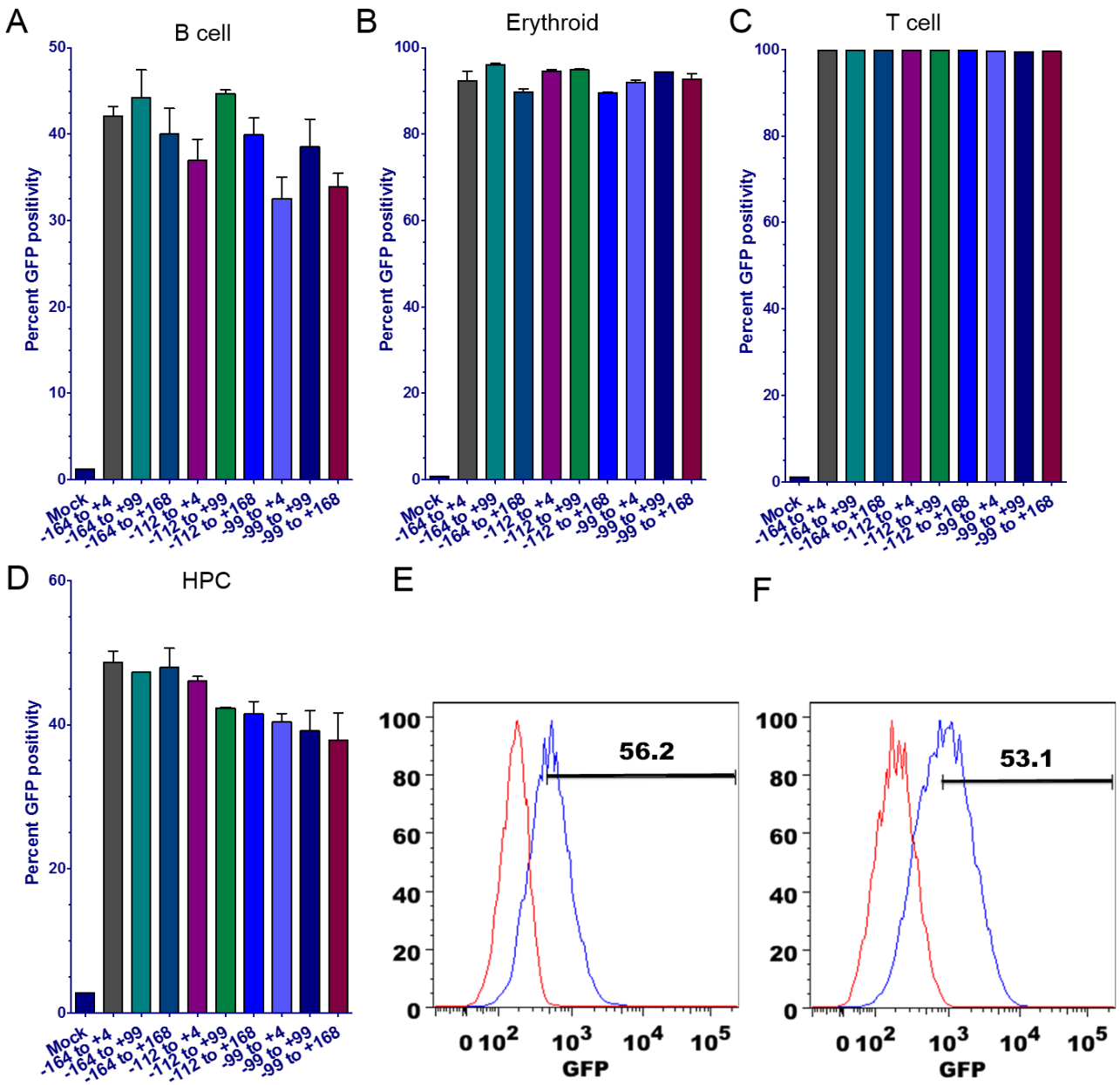


Figure 26. CD45 reporters efficiently mark human lymphoid, erythroid and multipotent progenitor blood cell lines**.** (**A**) Percent GFP positivity for B cell, (**B**) erythroid progenitor, (**C**) T cell and (**D**) a hematopoietic progenitor cell (HPC) line transduced with the indicated mouse CD45 reporters. FACS measured analyses were done in duplicate and repeated a minimum of 3 times. **(E)** Human CD45 reporter was used to transduce the HPC and **(F)** B cell lines. The mock control transduction is the red histogram.

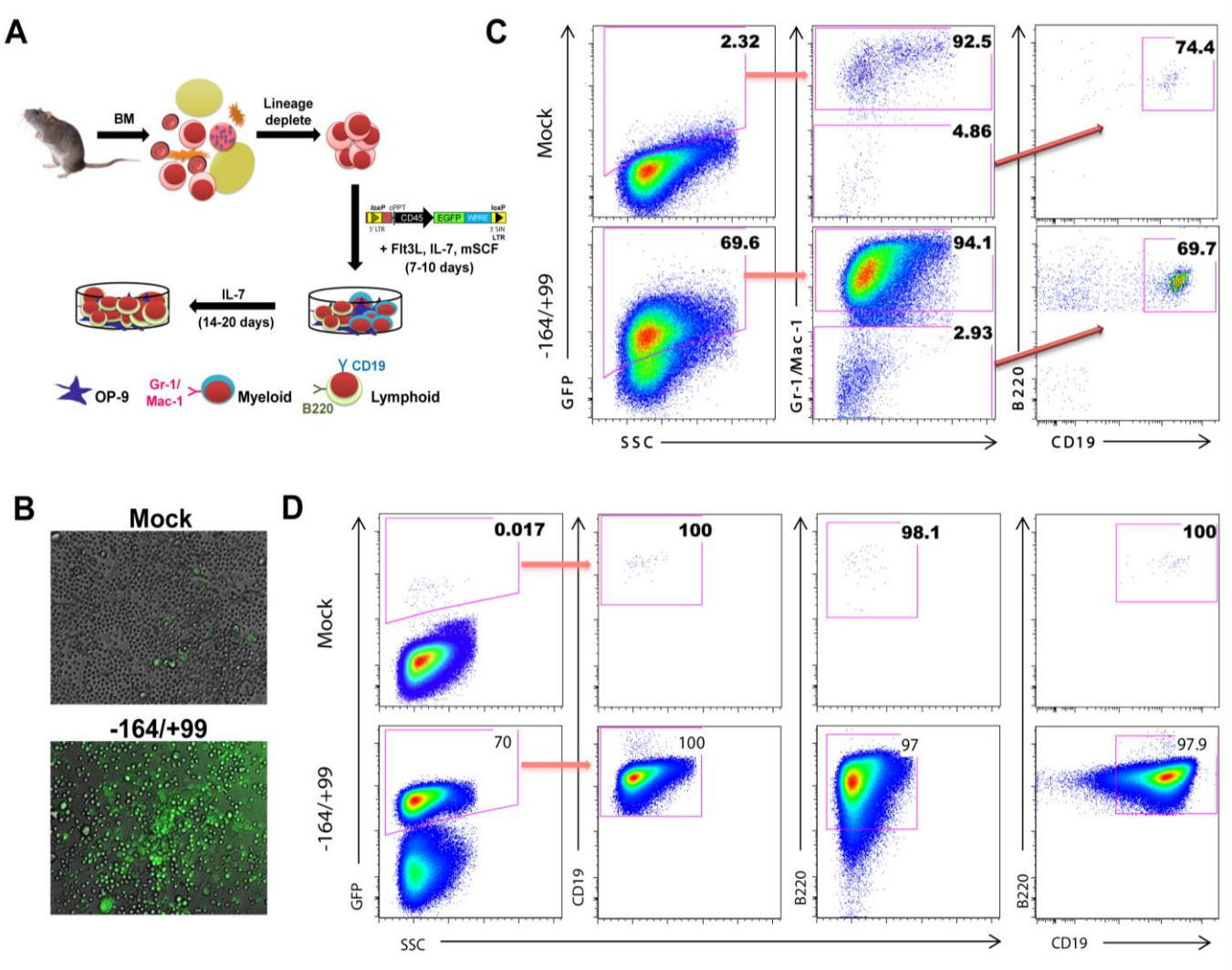


Figure 27. CD45 promoter enables primary myeloid and lymphoid cell reporting. (**A**) Schematic of *in vitro* B cell differentiation assay. Mouse BM was isolated, lineage depleted, and stimulated with cytokines for the generation of myeloid cells (day 10) and B cells (day 20). (**B**) Mock and CD45 reporter fluorescent images were captured at 20X magnification with a confocal microscope. (**C**) Bone marrow derived B cell differentiation assay was used to document Gr-1/Mac-1 (myeloid), and B220/CD19 (lymphoid) surface staining 10 days following transduction. Cells were gated first for GFP positivity and subsequently this population was sub-gated (red arrows) for myeloid (Gr-1/Mac-1) and lymphoid (B220/CD19) lineages. (**D**) By day 24 of B cell differentiation, a uniform population of CD19 and B220 double positive B cells can be observed. BM = bone marrow.

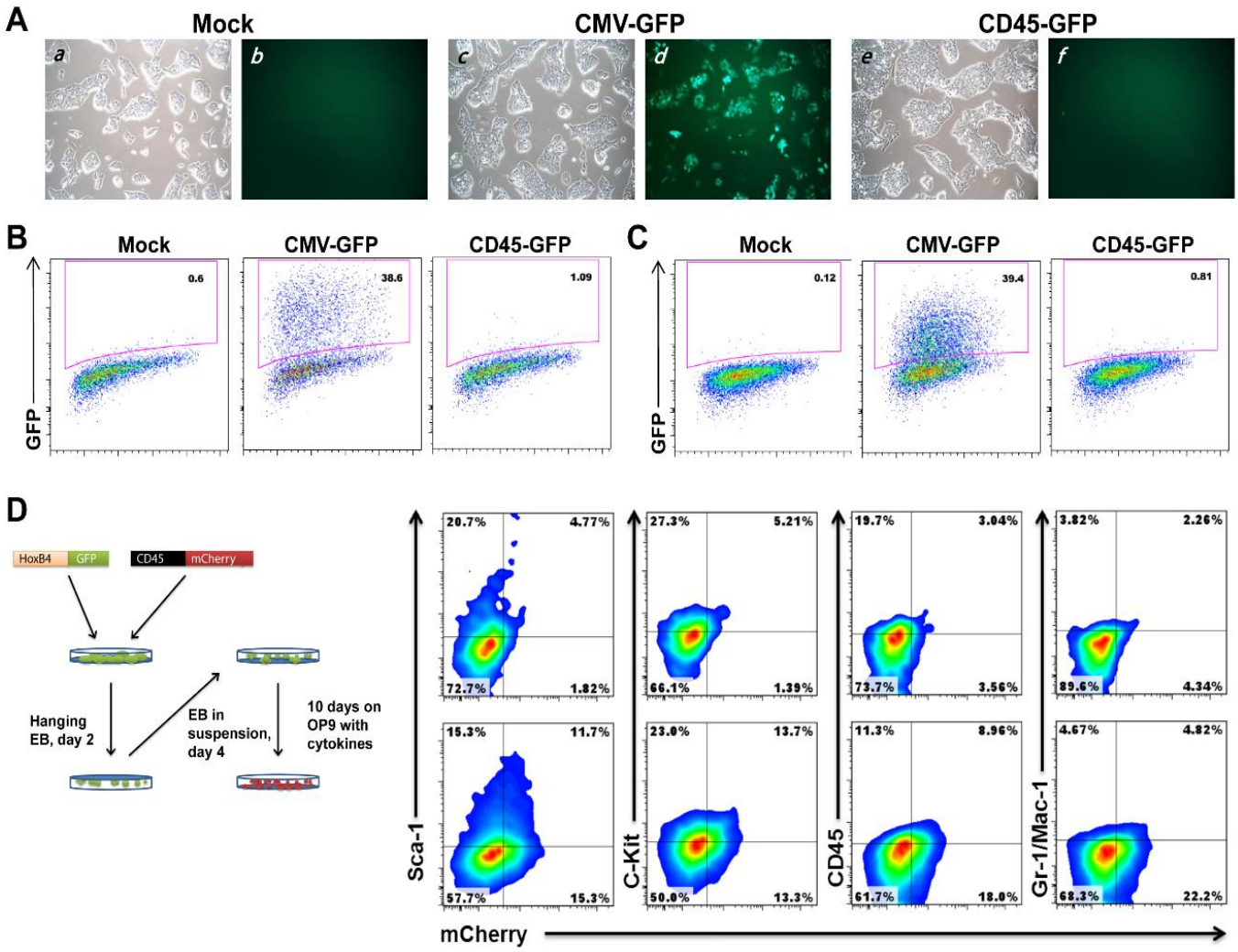


Figure 28. CD45-based hematopoietic reporter is not active in ESCs and fibroblasts. (A) ESCs with mock (panels a, b), ubiquitous (CMV-GFP; panels c, d) or hematopoietic lineage restricted (CD45-GFP (-164/+99 construct); panels e, f) reporter transduction following brightfield (panels a, c, e) and fluorescent (panels b, d, f) imaging. **(B)** Flow cytometric measurement of MEFs and **(C)** ESCs following transduction with the designated lentiviral reporters. Note the lack of ESC GFP fluorescence with the CD45 reporter. **(D)** HPC generation from ESCs. The strategy for ESC differentiation is shown on the left. Sca-1, c-Kit, CD45 and Gr-1/Mac-1 surface staining (right) on HPCs and hematopoietic lineages following CD45 reporter (mCherry) and HOXB4-GFP transduction of ESCs. Mock (top) and -164/+99 CD45 reporters (bottom) are shown. These cells were analyzed at day 10 following co-culture on OP9 stromal cells. GFP positive cells were sub-gated for mCherry positivity. ES = embryonic stem; MEF = murine embryonic fibroblast; HPC = hematopoietic progenitor cell; mCherry = monomeric cherry fluorescent protein.

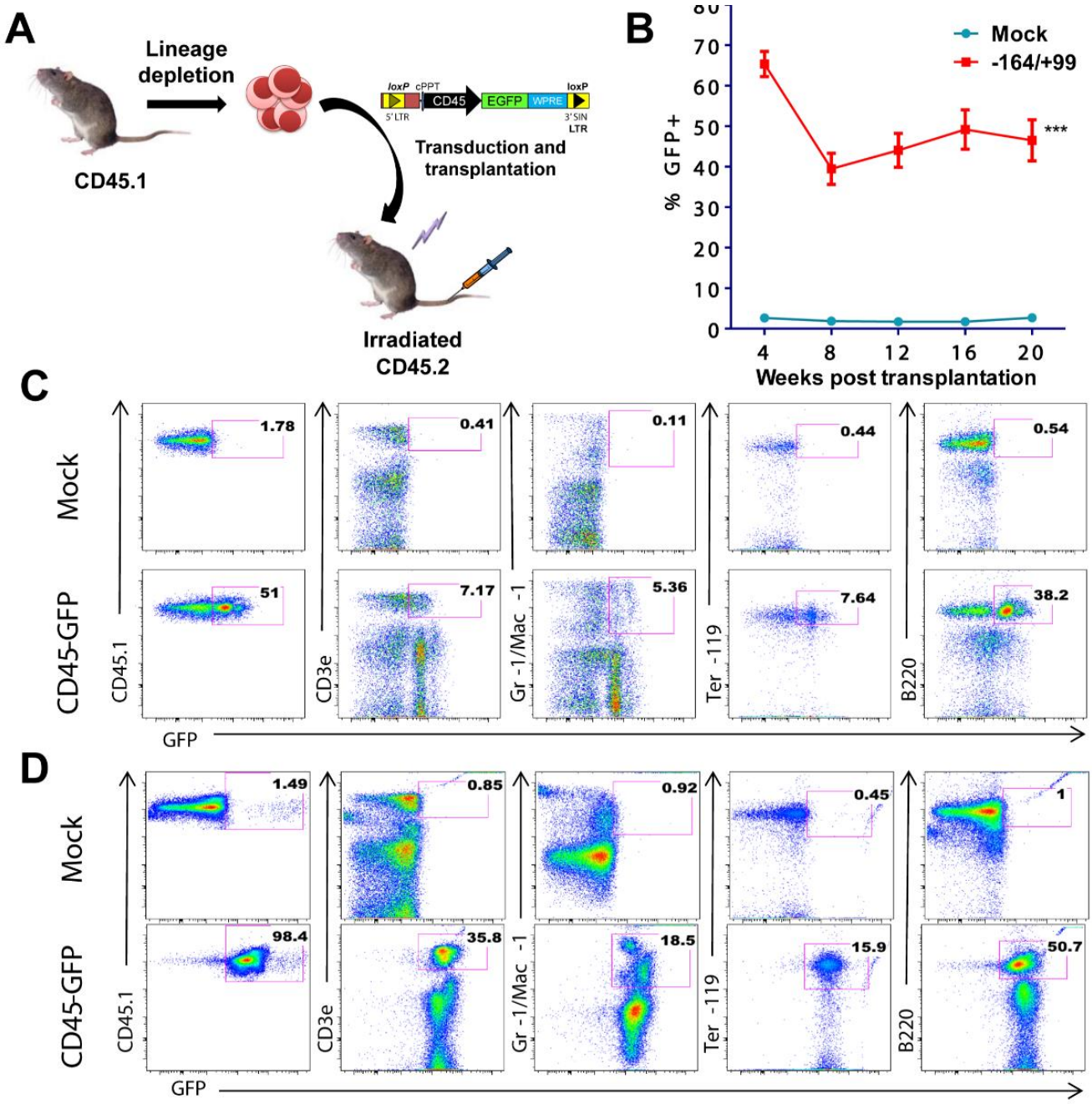

A
Lineage depletion
CD45.1
loxP cPPT
CD45
EGFP WPRE
loxP
5' LTR
3' SIN LTR
Transduction and transplantation
Irradiated CD45.2
B
Mock
-164/+99
% GFP+
***
Weeks post transplantation
C
Mock
CD45-GFP
CD45.1
GFP
CD3e
Gr -1/Mac -1
Ter -119
B220
1.78
0.41
0.11
0.44
0.54
51
7.17
5.36
7.64
38.2
D
1.49
0.85
0.92
0.45
1
98.4
35.8
18.5
15.9
50.7

Figure 29 CD45 reporter marks persistent hematopoiesis after transplantation. **(A)** Transplatation scheme depicting HSPCs derived from CD45.1 strain being transduced with a CD45 reporter, followed by transplantation into irradiated congenic CD45.2 recipients. **(B)** Peripheral blood stained for CD45.1 to determine chimerism in recipient mice that received CD45.1 HSPCs from mock transduced (n=5) and CD45-GFP (n=5) lentiviral reporter constructs. Error bars represent standard deviation. ***$p < 0.001$, statically significant 2 way ANOVA test with $p<0.001$. **(C)** Peripheral blood staining at 20 weeks with antibodies detecting CD3ε (T lymphocytes), Gr-1/Mac-1 (granulocytes and macrophages), Ter119 (erythroid progenitors), and B220 (B lymphocytes). Approximately 50% chimerism resulted following transplantation, and the first panel represents donor (CD45.1) cells. The fraction of GFP positivity for each lineage derived from CD45.1 HSPCs is indicated in the *pink square box*. The positive population for CD45.1 was further sub-gated to determine the percentage of cells representing myeloid and lymphoid lineages. **(D)** Same as in (**C**) except following HSPC transplantation without non-transduced supportive bone marrow. ANOVA = analysis of variance between groups; GFP = green fluorescent protein; HSPCs = hematopoietic stem and progenitor cells.

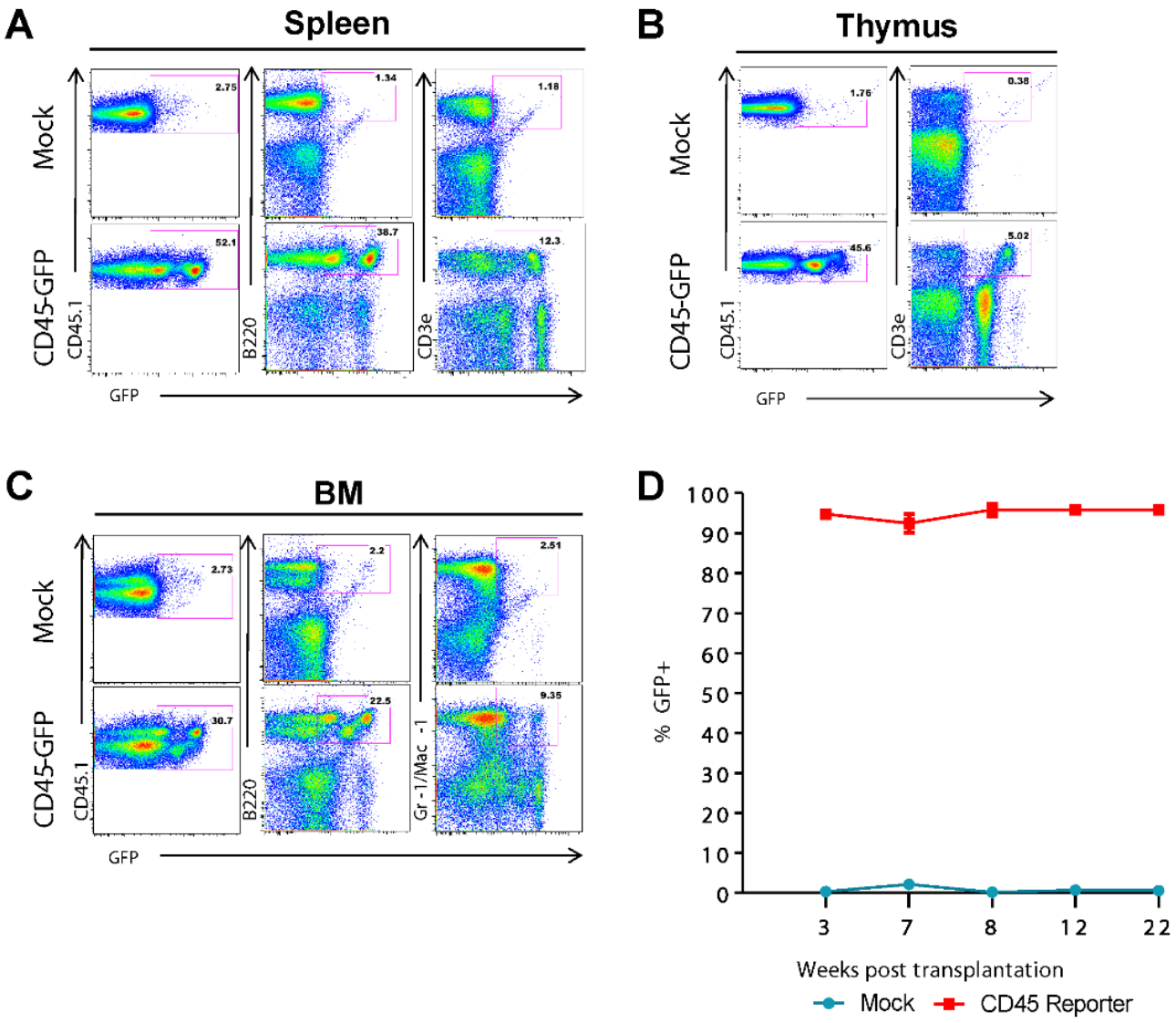


Figure 30. CD45 reporter marks blood cell formation in bone marrow and secondary hematopoietic organs, and persists in secondary transplants. (**A-C**) Surface staining on cells of the (**A**) spleen, (**B**) thymus, and (**C**) bone marrow 22 weeks after secondary transplantation. At this time, more than 80% of cells come from the donor (CD45.1). High-frequency multilineage marking with the CD45 reporter can be observed. (**D**) Bone marrow of primary transplant recipient mice was transplanted into secondary CD45.2 recipients. Peripheral blood staining 3 weeks posttransplant reveals reconstitution, with approximately 95% of CD45.1 cells exhibiting green fluorescent protein positivity 22 weeks postranplantation.

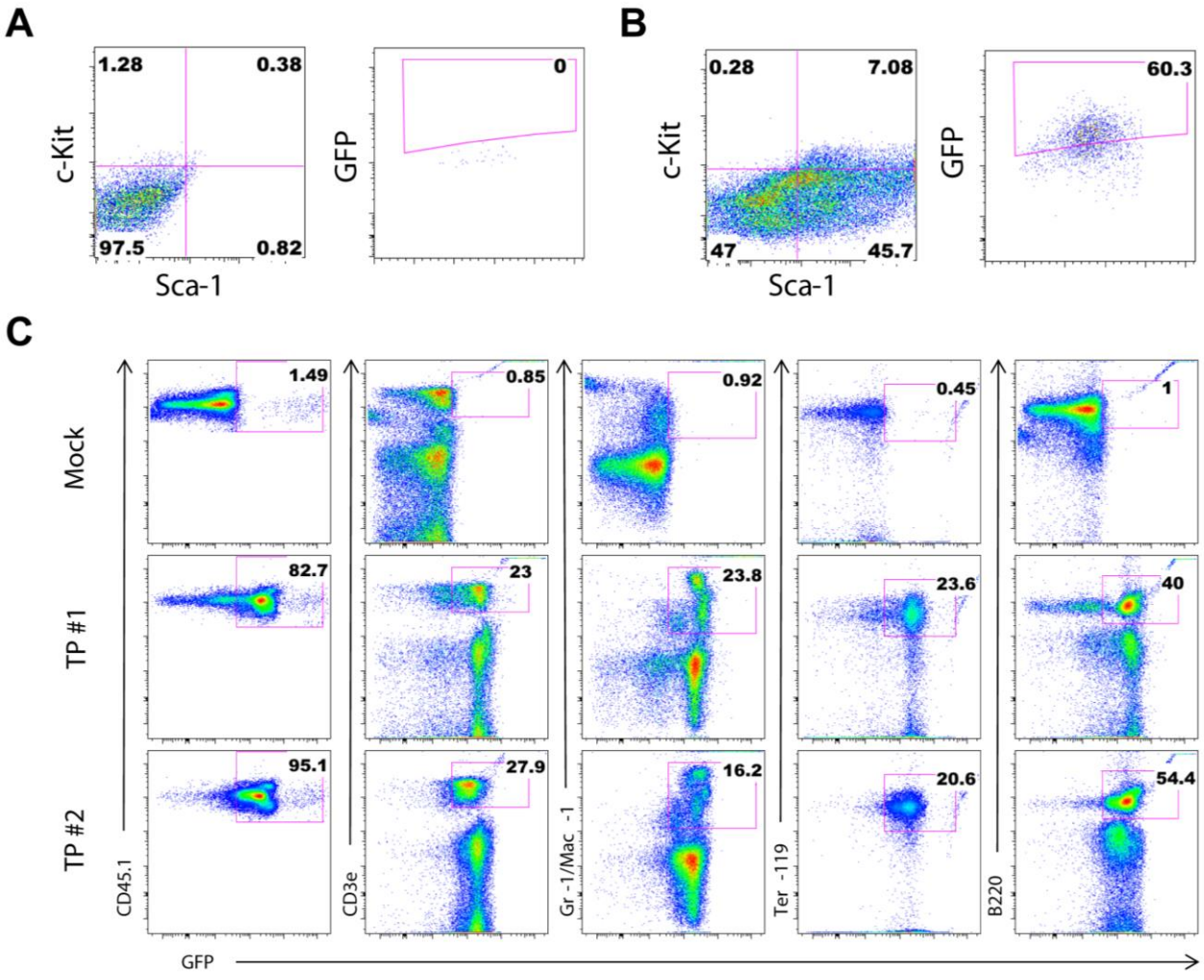


Figure 31. High green fluorescencent protein (GFP) marking following hematopoietic stem and progenitor cell (HSPC) transplantation without supportive bone marrow. (**A**) Lineage-depleted bone marrow was subjected to a mock or CD45 reporter transduction for 24 hours. The percentage of c-Kit, Sca-1 double positive, lineage-depleted (KSL) HSPCs was determined in cells that were unsteained (**A**, left) or stained (**B**, right) with antibodies directed against c-Kit and Sca-1. (**B**) CD45 reporter transduction results in high-level GFP marking efficiency in the KSL fraction (**B**, right). (**C**) Peripheral blood staining at 16 weeks posttransplantation in CD45.2 mice showing two independent transplants (TP 1 and TP 2) that received CD45 reporter-transduced HSPCs without supportive bone marrow. Note the virtually complete marking (95%) of donor-transduced reporter cells in TP 2 (**C**, bottom left).

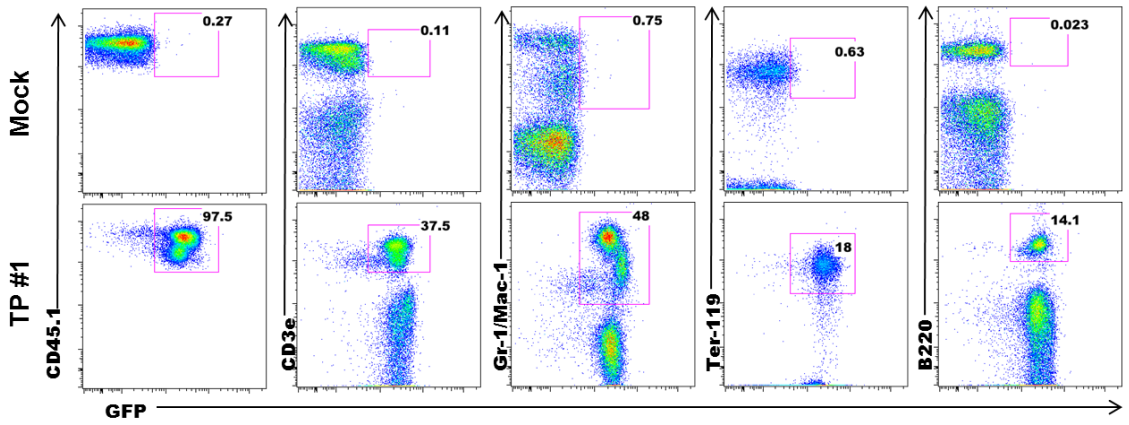


Figure 32. Persistent and robust green fluorescent protein (GFP) marking in secondary transplanted recipients. Peripheral blood staining at 22 weeks posttransplantation in CD45.2 mice that received bone marrow cells from primary transplant mice. Data displayed are representatives of five mice/group with an average of 50% chimerism. Approximately 98% of CD45.1 cells are GFP positive compared to mock controls. The reporter efficiently marks the myeloid (Gr-1/Mac-1), nucleated erythroid (Ter-119) and lymphoid (CD3ε, B220) compartments.

CHAPTER III

INDUCED PLURIPOTENT STEM CELL-BASED THERAPY

Chapter summary

To further our understanding of cellular reprogramming mechanisms we created induced pluripotent stem cells (iPSCs) from mouse fibroblasts using a lentiviral system. The four reprogramming factors (OSMK) were each coexpressed with a specific fluorescence protein linked via a P2A sequence. After transduction into mouse embryonic fibroblasts (MEFs) we selected colonies that were representative of silenced and active transgenes to expand into stable iPSC lines for our study. We found that iPSC lines with silenced transgenes were much more effective in differentiating into neuronal progenitor cells and could form teratomas in immune-compromised mice. However, cell lines with actively expressed transgenes were unable to differentiate into neurons nor form teratomas. Upon excision of the transgenes in these cells, teratomas formed readily and at a faster rate compared to silenced transgenic iPSC lines. Interestingly, in the presence of highly active transgenes, there was a complete loss of Nanog, Esrrb, and Lamin B expression. These results suggest that multiple factors might contribute to the differentiation potential of different iPSC lines and that cell lines with silenced transgenes are better candidates for clinical studies.

Introduction

Cellular plasticity and differentiation have long been areas of interest not only to developmental biologists, but also to the greater healthcare community. In 1952, Briggs and King generated a tadpole by transferring nuclei from differentiated blastula cells into frog eggs [20]. Subsequently, Gurdon in 1962 demonstrated that nuclei from more mature cells in the intestinal epithelium when injected into enucleated frog egg could generate feeding tadpoles [22]. These studies showed that epigenetic marks in differentiated nuclei are reversible when exposed to undifferentiated cells and paved the way for cloning experiments in mice and sheep.

Other groups later found that cells isolated from the inner cell mass of the growing embryo at the blastula stage had a remarkable ability to divide indefinitely and could differentiate into other cell types [33, 162]. These embryonic stem cells (ESCs) had higher pluripotency potential and were more accessible than eggs. While both mouse and human ESCs have been used to study human diseases, ESC usage faced ethical concerns because their isolation requires the destruction of growing embryos. A method to replicate ESC pluripotency was needed.

In 2006 the Yamanaka group indentified four transcription factors, Oct4, Sox2, c-Myc, and Klf4 (OSMK), out of 24 factors studied that convert mouse embryonic fibroblasts (MEFs) into cells that were morphologically and functionally similar to ESCs. These derived cells were called induced pluripotent stem cells (iPSCs); these iPSCs and ESCs had similar stem cell surface markers and gene expression patterns. These same four transcription factors also converted human fibroblasts into iPSCs, suggesting that these four factors are core pluripotency regulators and are conserved across species [94, 163]. Not only did iPSCs enable scientists to

avoid the use of fertilized eggs, iPSC usage minimized the risks of immune rejection in therapeutic applications since the iPS precursor cells could be taken directly from the patient's blood or skin. Many genetic diseases have been successfully modeled using patient cells as a source for iPSCs, including Adenosine Deaminase Deficiency-Related Severe Combined Immunodeficiency, Shwachman-Bodian-Diamon syndrome, Gaucher disease type II, Duchenne and Becker muscular dystrophy, Parkinson's disease, Huntington disease, Juvenile onset type I diabetes mellitus, Down syndrome, Lesch-Nyhan syndrome, LEOPARD syndrome, Fanconi anemia, and β-thalassemia [164].

While iPSCs have great potential in regenerative medicine, a number of issues exist when using iPSCs in the clinic. First, the cell type used to generate iPSCs needs to be carefully considered as different cell types have variable levels of plasticity. Fibroblasts are a commonly used source for iPSCs because they are readily accessible and easily isolated from the host, making them an excellent platform when applied to "personalized" regenerative medicine. However, the iPSCs generated from tail tip fibroblasts (TTFs) had many residual pluripotent cells after three weeks of in vitro differentiation into neural spheres, and these cells later developed into teratomas upon transplantation into immune-deficient brains [165]. This could be due to incomplete reprogramming, epigenetic memory, and/or genomic instability. Other cell types have been transformed with higher efficiency than fibroblasts due to higher level of endogenous pluripotency genes. These include adipose cells, keratinocytes, melanocytes, blood cells, and neural stem cells. Keratinocytes, which have high levels of c-Myc, can be isolated from plucked hairs or foreskin, and are one-hundred times more efficient and two-fold faster than fibroblasts in cellular reprogramming [45, 166]. Neural stem cells (NSCs) only require Oct4 and Klf4 to efficiently convert NSCs into iPSCs since NSCs normally express Sox2 and c-Myc

[41, 42, 90, 167, 168]. Human melanocytes have high levels of endogenous Sox2 and can be reprogramed with only Oct4, Klf4, and c-Myc [169]. Adipose-derived stem cells, obtained from lipoaspiration in patients, transform twenty-fold more efficiently and twice as quickly as fibroblasts [46]. Further study is merited to determine the pluripotent capacity of iPSCs generated from fibroblasts or other cell sources [88]. For example, keratinocytes convert into iPSCs much more efficiently than other cell types [44, 166]. Different germ layers also contain different epigenetic memories, requiring researchers to use various starting materials to generate iPSCs depending on the cell type ultimately desired [42, 43, 61, 170].

Another issue is that these cells must be able to differentiate at one-hundred percent efficiency to prevent teratoma formation. Although teratomas are benign, they are unacceptable in the clinic. Any iPSCs that escape differentiation will give rise to a teratoma upon transplantation. Most studies tried stringent purification measures with specific cell surface markers to ensure only differentiated cells were transplanted, but the mouse lifespan is not long enough to detect tumor development. Another useful alternative is to include a suicide gene cassette as a "fail safe" mechanism to destroy tumorigenic cells [159, 171].

One other consideration when working with iPSCs is the gene expression profile. Gene dosages of Oct4, Sox2, c-Myc, and Klf4 are crucial in the generation and maintaince of iPSCs. Oct4 levels in particular are crucial for iPSCs to remain pluripotent. A three-fold increase in Oct4 levels improved cellular reprogramming while a three-fold decrease nearly eliminated de-differntiation [172]. Intriguingly, excessive amounts of Oct4 perturbed enhanced of iPS generation. Niwa and colleagues showed that a two-fold increase of Oct4 in mouse ESC promoted differentiation into embryonic and extraembryonic cell types, and reducing in levels of Oct4 induced de-differentiation of these cells into trophoblasts. Hay and colleagues showed that

knockdown of Oct4 in mouse and human ESCs led to differentiation [173]. Overexpression of c-Myc, Nanog, and Klf4 does not affect the pluripotentcy of ESCs; however, overexpression of Sox2 results in differentiation of mouse ESCs. It has been established that, as the iPS reprogramming becomes complete, the endogenous gene expression levels will increase while the levels of the transgenes will decrease. This could be a result of general changes in chromatin states and epigenetic changes (methylation, acetylation, etc.). It still not understood what specific players or multifactors lead to this selective and precise shutdown of transgenes.

Identifying a reliable method to generate high fidelity iPSCs is also important. Originally iPSC generation used retroviruses. This method was robust but viral integration is undesirable, not to mention the high chance of viral reactivation [174, 175]. In addition, concerns remain regarding the continued expression of retrovirally introduced transgenes, especially c-Myc [36, 88, 172, 176]. Retroviral integration is biased toward the transcription start sites and open chromatin, this often leads to insertional activation of oncogenes since the viral long terminal repeat promoters are strong enhances [177-179]. This type of virus also has been observed to integrate near oncogenic genes in patients who received retroviral therapy to treat X-linked severe combined immunodeficiency [175, 180-183]. As an alternative, lentiviral vectors can be used. Although, retrovirus and lentivirus integrate into the host genome (which is not ideal in the clinic), unlike retroviral vectors, lentiviral therapy does not lead to tumorigenesis in mouse and human studies. Lentivirus has aroused considerable interest due to less harmful side effects; and the vector has used to treat patients with AIDS or leukemia with good success and no negative effects [184, 185]. Lentiviruses integrate in a random manner with slight bias toward active gene regions, including geneic introns and exons but not promoters [178, 179, 186]. To avoid the oncogenic risks of integration, other methods for generating iPSCs have been devised including

treatment with small molecules, non-integrating viruses, and expression of mRNAs and miRNAs [54, 57, 63, 76, 77, 89, 93, 187, 188]. Unfortunately, these non-integrating methods generally are less efficient and result in significantly lower yields compared to lentiviral therapy.

Since Yamanaka et al. demonstrated that somatic cells can de-differentiate when forced to express Oct4, Sox2, cMyc, and Klf4 [36], many studies (including our own) examined cellular reprogramming at the molecular and cellular level, with the goal of making these de-differentiated cells safe for clinical applications [189]. In our study, we employed a lentiviral system that has the LTRs that lack the usual endogenous promoter activity, thereby eliminating the potential for accidental reactivation of transgenes or any neighboring genes [50-52]. This system also includes the loxP sites embedded within the LTR region, which allows the removal of the transgenes from the iPSCs [49]. Thus, we proposed that this system is an efficient method to generate high fidelity iPSCs for in vitro studies and animal models.

Our goal in this study is to gain an understanding of the kinetics of cellular reprogramming and the mechanisms that can alter reprogramming efficiency. Spatial and temporal aspects of cellular reprogramming have been revealed nicely in studies by the Sadelain group, who tracked expression of the four OSMK factors [172]. In this study, we employed a similar strategy except we used slightly different fluorescent proteins and introduced an adenovirus-expressing Cre recombinase to mediate excision of transgenes. The results from our study showed that when transgenes were greatly overexpressed it interfered with differentiation. Interestingly, Nanog levels do not affect the differentiation potential of iPSC lines.

Materials and methods

Generating OSMK fluorescent protein-tagged plasmids

The coding sequence for murine Oc4, Sox2, c-Myc, Klf4, eGFP, mKOrange2, mTagBFP, and mKate were PCR amplified using Agilent Pfu Ultra II polymerase (Cat. No. 600850) and subsequently cloned into the pCR blunt II TOPO vector (Life Technologies) and confirmed by sequencing. The primers for Oc4, Sox2, c-Myc and Klf4 were engineered with a 5' end Gly-Ser-Gly-P2A linker. Either two-way or three-way ligations were used to ligate the coding sequence of each gene and its respective color into a Gateway adapted entry vector containing two Gateway attL sites. Each fluorescent protein was fused to each gene via a P2A sequence such that: eGFP is linked to Oct4, mKOrange2 was linked to Sox2, mTagBFP was linked to Klf4, and mKate was linked to c-Myc. The lentiviral vector (reference: Levasseur-D *et al*, Blood 2003) was converted into a Gateway destination vector containing attR sites. To obtain the final expression vector, each entry vector was recombined with the lentiviral destination vector via a BP-LR clonase II (Life Technologies, Cat. No. 11789-020 and 11791-019) reaction to generate the final lentiviral expression vectors.

Lentivirus production and titer

Concentrated lentiviral supernatant was produced and titered as described previously [190]. In brief, 293T cells were split for 3 consecutive days and used to package recombinant sequence into lentiviral particles using the protocol described previously [190]. The viral supernatant containing recombinant vectors expressing the reprogramming factors was harvested and concentrated via ultracentrifugation at 26,000 rpm for 90 min. The pellet was dissolved in

Opti-Mem medium and stored at -80^{o}C. The viral titer was determined by two methods: Hela cell transduction and qRT-PCR. .

Transduction of MEFs and generation of iPS

200,000 murine embryonic fibroblasts (MEFs) in each well of a 6-well plate were transduced with OSMK viruses at multiplicities of infection (MOI) of 0.3, 1, and 3. The medium was replaced after 24hr. At 48 hours, the cells were trypsinized (0.05% Trypsin) and seeded (in 10 cm dishes) onto a fibroblasts feeder layer such as irradiated SNLP and irradiated MEFs. At this time point we switched the medium to ES medium with or without valproic acid (VPA), and medium was changed every 2 days until the end of the experiment. Colonies were observed at day 7, post co-culture on SNLP for MEFs that were transduced with OSMK at MOI of 3 with the presence of valproic acid (VPA). Fluorescent images were captured with a Nikon microscope to track reprogramming.

Flow cytometry and reagents

Intercellular staining of OSMK factors was performed by using the BD Pluripotent Stem Cell transcription factor analysis kit (Cat.# 560689) . In brief, iPSCs were rinsed once in 1X PBS, trypsinized, and pelleted by centrifugation. The cell pellets were washed in 1X PBS and $1x10^{7}$ were fixed in 4% paraformaldehyde for 20 min at room temperature. Afterwards, cells were washed twice in 1 ml 1X BD Perm/BD perm/Wash buffer (prepared from 10X concentration supplied by the kit) and incubated in this buffer for 10 min to permeabilize the cells. Subsequently, cells were stained with Nanog Alexa 647, Sox2 Alexa 647 and Oct4 Percp.Cy5.5 antibodies and analyzed via LSR II Violet at the University of Iowa flow cytometry core facility.

Western blotting

Cells were rinsed twice with 1X PBS, scraped from the plate and transferred into 1.5 ml Eppendorf tubes. Cell pellets were lysed with lysis buffer (10 mM NaF, 1 mM Na3VO4, 1:100 protease inhibitor cocktail (Sigma), and 1 mM PMSF). The protein concentration was determined by Bradford assa; and 5 μg of each sample was resolved by 12 % SDS-PAGE ran at 30mA and transferred to an EMD Millipore PVDF membrane. After transferring, the membrane was blocked with 5% milk for 1 hr and then incubated in primary antibodies recognizing either c-Myc (CST, kind gift from Frank's lab), anti-rabbit Nanog (Bethyl), anti-mouse Sox2 (R&D system), anti-rabbit Klf4 (Santa Cruz, kind gift from Dr. Budd Tucker), anti-rabbit-Oct4 (Santa Cruz), anti-goat Lamin B (Santa Cruz), or anti-rabit GAPDH (Santa Cruz, kind gift from Dr. John Colgan).

Quantitative realtime PCR (qRT-PCR)

Fresh RNA was isolated from J1 ES, iPSC lines, and iPS-derived neuronal cells using the GE Healthcare RNA Isolation Kit. cDNAs were synthesized using Promega GoScript Reverse Transcriptase according to manufacturer recommendations. The cDNA products were then diluted 40-fold, and 5 μl was used for each qRT-PCR reaction following amplification with Takara Ex Taq polymerase HS. Primer sets for each gene are mentioned in **Table A2,** below. Detection was performed with a Biorad CFX96 real-time PCR machine.

Teratoma assay

In this assay we used seven-week old female NOD SCID mice (Jackson Laboratory, stock #005557). Two mice were used per iPS clone and approximately $1x10^6$ cells suspended in

200 μl of PBS were injected into two subcutaneous hind leg sites on each mouse. After 21 days, tumors were detected and mice were sacrificed to remove the tumors. The tumors were fixed in a 20:1 formalin solution and sent to the University of Iowa histopathology core to be embedded in paraffin and HE stained.

Neuronal differentiation assay

Pluripotent cells were coaxed to differentiate into neuronal cells as reported by Bibel et al., with a few modifications [191]. N-2 (Life technologies, #17502048) and B-27 (Life technologies, #17504044) supplements were used instead of individual components as mentioned by the author. In brief, $4x10^6$ cells were plated on Grieger bacterial plates in 15ml of ES medium without LIF (DMEM supplemented with 15% FBS, 2% pen/strep, 1% nucleoside mix, 1% L-glut, 10^{-4} M Beta ME). The medium was changed on day 2, day 4, and day 6, at which point retinoic acid was added to a final concentration of 5 μM. On day 10, the EB cultures were removed, washed with 1X PBS, and dissociated in 500 μl of 0.05% trypsin in a 37°C water bath. Finally, about $2x10^5$ cells per cm^2 were plated onto 6-well plates or 12-well plates. Glass cover slips were added into each well of these plates before coating with Poly-DL-ornithine (PORN) hydrobromide (Sigma, #P-8638) 48 hrs prior to seeding the cells. After 48 hrs PORN was suctioned off and rinsed twice with sterile ddH_2O, and laminin (Sigma, #L2020) was added to the wells and incubated at 37°C overnight. The plates were rinsed prior to coating with dissociated EB cells and N2 medium was added for the first 48hrs. N-2 medium is made up of F-12 DMEM medium with a final concentration of 1X N-2 supplement, 1% P/S, 1% glut, 50 μg/ml BSA, and 25 μg/ml Insulin. After 48hr, neuron complete medium was used in place of N-2

medium. This medium made up of Alpha MEM media supplemented with 1X B-27 supplement, 1% pen/strep, and 1% L-glut.

Results

Generation of iPSCs and tracking iPSCs kinetics

We tracked reprogramming of fibroblasts as they de-differentiated into iPSCs by monitoring expression of the four Yamanaka factors, Oct4, Sox2, Klf4 and cMyc (OSMK) that were co-expressed with fluorescence proteins. Oct4 was linked to GFP, Sox2 to mOrange, Klf4 to BFP, and cMyc to mKate, as seen in **Figure 33**. The sequences encoding the fluorescence protein and those encoding the OSMK genes were separated by the picornal ribosome skipping P2A sequence, which ensured equivalent expression of the two coding sequences [192]. The Trono lentiviral system that we employed here contained a loxP site within the 3' UTR that, upon integration, is duplicated at the 5'-UTR, so loxP sites flank the transgene. This will enable us to excise the transgenes via Cre recombinase.

The fibroblasts were transduced at low MOIs (0.3, 1, and 30) to improve the likelihood of only one integration event per cell and the cells were incubated in MEF medium for 48 hrs before being transferred to irradiated SNLP plates or MEF plates with ES medium. Previous studies showed that valproic acid (VPA), a histone deacetylase inhibitor, enhances reprogramming kinetics by maintaining the chromatin in a relaxed state [86-88], so we added VPA to one set of samples to examine its effect on transgene silencing. The plate with the earliest visible iPS colonies (ES-like morphology) was the plate containing MEFs that were transduced with the four reprogramming factors at MOI 3 and added VPA. In general, GFP

(Oct4) and mOrange (Sox2) were the most visible, while mTagBFP (Klf4) and mKate (cMyc) were less so. Based on our observations, it appeared that Oct4 (GFP) expression was detectable first, followed by Sox2 (mOrange), with c-Myc (mKate) and Klf4 (TagBFP) being expressed last. The majority of colonies arose by day 14 and large colonies with or without observable fluorescence color were picked, isolated, and retrieved around day 18 and expanded first on 96-well plates, then on 12-well and 6-well plates. As expected, plates containing VPA gave raise to iPS colonies more rapidly, but most of these colonies did not get silenced as evidenced by bright fluorescence. It is also of note that an MOI of 3 was better at generating iPS colonies than the MOIs of 0.3 and 1. Finally, we found that more colonies arose when the initial transduced MEFs were seeded on SNLP feeders rather than irradiated MEFs feeders.

We observed transduced cells undergoing the mesenchymal to epithelial transition (**Figure 33B**), a hallmark of cellular reprogramming. Transitioning cells fluoresced brightly as they became round (epithelial) and formed colonies (**Figure 34**). As reprogramming became complete, many iPS colonies assumed an ESC morphology (round and clump) and some stopped fluorescing entirely (**Figure 35**). Some colonies selected after few passages showed no fluorescence, single or double color fluorescence, and all four color fluorescence.

For further analysis, we chose colonies representative of two extreme types (either having all four transgenes "off "or all four "on"). Cell line 1 represents colonies with no detectable fluorescence (silent transgenes). Cell line 2 represents colonies with detectable fluorescence (active transgenes). These representative colonies were isolated, expanded, and stored for further analysis.

iPS clones exhibit similar pluripotent gene profiles to ESCs

Next, we wanted to determine whether these iPS lines were similar to ESCs at the molecular level. First, we assayed each iPS line for mRNA transcripts of several pluripotency genes including Nanog, Esrrb, Oct4, cMyc, Klf4, and Sox2. mRNA was isolated from both lines before and after infection by an adenovirus expressing Cre recombinase and gene expression was analyzed by quantitative real-time PCR. As depicted in **Figure 37,** we expected the transgene-specific primers not to detect any transcripts in the J1 ESCs. Similarly, for cell line 1 the transgenes were not amplified. The opposite was expected from cell line 2 with active transgenes, when we expected the transgenes to be highly expressed consistent with the intense fluorescence signal. Additionally, Oct4 transgene was expressed at a significantly higher level than other cell lines. After introduction of Cre recombinase, as expected exogenous Oct4 level diminished after Adenovirus was added to excise the transgenes.

Next, we measured total expression of Oct4, cMyc, and Sox2 mRNAs using PCR primers that recognized both endogenous and exogenous forms. We found that before and after introduction of the Cre recombinase, the gene expression profile of cell line 1 (transgenes are "off") was comparable to that of the J1 ESC line. The expression levels for Esrrb, Nanog, Oct4, cMyc, and Klf4 were similar between the two cell lines. The level for Sox2 mRNA seemed to be higher in cell line 1 compared to J1 ESCs. In cell line 2, the expression of the pluripotency genes Nanog and Essrb was undetectable in the presence of active transgenes (transgenes "on"). However, after the introduction of Cre and the excision of transgenes, mRNA levels of Nanog and Essrb dramatically increased (**Figure 38)** (the data in **Figure 39** is a complete experimental design extrapolated from **Figure 38**). To ensure complete excision of the transgenes, we tested multiple MOIs including 30 and 100 of the adenoviral cre recombinase virus. In general, a MOI

of 30 restored the Nanog expression to normal ESC levels. Although, MOI of 100 was more effective in removing the transgenes, as detected by fluorescence, a significant toxicity effect was observed. Cell line 2 had Nanog levels close to ESC levels at MOI of 100. The higher MOI effectively diminished fluorescence signal, as there was very little detectable amount after transduction (**Figure 39)**.

Flow cytometry was used to monitor expression of pluripotency factors at the protein level. Both lines 1 and 2 expressed levels of Oct4 and Sox2 protein similar to that of the J1 ESC line **(Figures 41 & 42)**. Nanog expression, however, while high in the other cells, was undetectable in cell line 2 prior to infection by the Adeno Cre virus; the Nanog level was restored to the same level as seen in the other cells after Cre recombinase was introduced. Oct4 and Sox2 were similarly expressed at the protein level as compared to the other two pluripotent cell lines. This finding was confirmed with Western blots as shown in **Figue 40**. Surprisingly, Lamin B (one of our loading controls) was undetectable in cell line 2 until after the excision of the transgenes (**Figure 40**). This result was repeated three times with different protein lysates. Overall, Nanog, Essrb, and Lamin B levels were significantly altered depending on the presence of the transgenes. Our Sox2 antibody was not specific and required increase exposure time. The Western blot was repeated as in **Figure 40A** with the exception that we loaded iPSC line two that had underwent two rounds of adenviral cre transfection (each time MOI of 100 was used). To enhance Sox2 signal, we expose the film at various time points, but overnight exposure gave the best signal (**Figure 40B**). A band representing Sox2 was detected for each of the samples. We noted that Sox2 antibody binding was not specific as there was a significant background observed. The result suggested the activity of the transgenes might downregulates the Nanog, Esrrb, and Lamin B expression. When the transgenes are active, no detectable mRNA and

proteins for Nanog, Esrrb, and Lamin B were observed. However, excision of the transgenes restored expression of these genes.

Quantitative PCR revealed that Oct4 was the most highly expressed of the transgenes in the active iPSC line 2. This doasage imbalance could be responsible for the downregulation of Nanog, since Oct4 can act as a negative regulator of Nanog. To test this hypothesis, we overexpressed Oct4 in J1 ESCs using lenitviruses at various concentrations. RNA was isolated 72 hr post transduction for qRT-PCR analysis. The change in Nanog level for this particular experiment was not enough to validate our hypothesis. This could be due to several reasons. First, the lentivirus used to transduce the cells was not as robust as it could be; using three times the amount of the virus resulted in very few GFP positive cells. Second, the CMV promoter might have been silenced by the J1 ESCs. After observing these results, we repeated this experiment using the same lentiviral design, but with a fresh batch of viruses, and replaced the CMV promoter with the PGK promoter instead. When overexpressed, we found that many J1 ESCs differentiated in the plates using either condition and they all fluoresced from GFP.

Pluripotency characterization

One of the gold-standard tests for true pluripotency in iPSC lines is the teratoma assay. In the teratoma assay, high quality iPSC lines should form teratomas as readily as ESCs when injected into SCID mice. Teratomas are benign tumors that consist of all three germ layers: ectoderm, mesoderm, and endoderm. To test our cell lines, roughly 2 million cells were subcutaneously injected into SCID mice. SCID mice that received iPSC line 2 with the transgenes excised, by pre-infection with Andeno-Cre, formed large teratoma masses quickly, while the tumors in mice injected with iPSC line 1 took longer to develop in size, but these

formed nicely differentiated three germ layers. The tumors in mice injected with iPS unaltered cell line 2 took the longest to form and grow. These tumors were composed of undifferentiated cells only (**Figure 43**). This is in contrast with SCID mice that received transgene-excised iPSC line 2. **Figure 43** shows the tumors from transgene-excised cell line 2 were well differentiated into the three germ layers as exhibited by muscles (mesoderm), nerve fibers, and airway epithelial. Thus, iPSC lines either silenced or excised transgenes, form teratomas with all three germ layers.

Neuronal differentiation

To determine whether our iPSC lines can efficiently differentiate into neuronal cells, we adapted the differentiation assay of Bibel et al. [191], using N2 and B27 supplements as our differentiation medium instead of the master mix recommended by the authors. The silenced iPSC line 1 showed robust differentiation capacity, before and after the excision of the transgenes. After 24hr incubation in N2 medium, cells formed neuronal extensions, and did so even more efficiently than J1 ESCs. In contrast, iPSC line 2 (which expressed the OMSK transgenes) failed to differentiate during the first 24 hours, with very few cells adopting the neuronal morphology. These cells then died quickly even after the transgenes were excised; the cells also failed to form neuronal cells as they either failed to proliferate or divide rapidly to form a sheet of cells that resembled retinal epithelium cells. To confirm that neuronal differentiation neuronal occurred, we used immunohistochemistry to detect the presence of two neuronal markers: MapII (microtubule associated protein 2) and GFAP (an astrocyte's specific marker). **Figure 44** shows representative images of the ES and iPSC lines after neuronal differentiation. J1 ESCs stained positive for MapII (red) as did cell line 1 before and after introducing Cre

recombinase. iPSC line 1 also stained positive for GFAP. There were not enough adherent cells for cell line 2 for this experiments. Overall, the result showed that cell line 1 can differentiate effectively into neuronal cells, and stain positive for neuronal markers.

To further confirm the neuronal identity of the differentiated line 1 cells, we analyzed Pax6 transcripts by qRT-PCR, as performed by Bibel et al., at three different time points: before differentiation, during differentiation embryoid body stage (CA) and after differentiation. In J1 ESCs, used here as a positive control, as expected Pax6 was absent in pluripotent cells, highly expressed during neuronal differentiation at the embryoid body CA stage, and remained at reduced levels after differentiation was completed (**Figure 45**). The black bar represents the pluripotent cell line prior to differentiation, the blue bar represents the embryoid stage (CA) during differentiation, and the red bar is the final stage of differentiation. Similar results were observed using iPSC line 1 before Cre, iPSC line 1 after Cre recombinase, and iPSC line 2 after Cre recombinase, although induction of Pax6 was less robust in iPSC line 2 after Cre. In constrast iPSC line 2 without Cre had elevated levels of Pax6 before differentiation and the level of this marker appeared to be the same before, during and after neuronal differentiation.

Next, we looked the expression of Trk2b, as this protein should be exclusively expressed in mature neurons. As expected, in J1 ESCs and cell line 1 (before and after Cre recombinase was expressed), the level of Trk2b increased significantly at the final differentiation time point (red bar **Figure 46**). Interestingly, in cell line 2 (before and after Cre recombinase), the Trk2b mRNA level also increased at the final differentiation time point; however, the expression level is lower than in the other cell lines. These results are similar to those reported by Bibel et al. and suggest that our adapted differentiation protocol is working [191].

Neuronal differentiation was further characterized using primers that would amplify transcripts encoding neuronal markers including Nestin, Synaptophysin, and Blbp. Nestin, a type IV intermediate filament, is transiently expressed during nerve cell development (as well as other cells) [193]; synaptophysin [194], a protein found in neuroendocrine and synaptic cells; and Blbp a protein that is essential for the developing nervous system [195]. We expected Nestin mRNA to be expressed in a pattern similar to that of Pax6; however, in the J1 ESC line it was detected both during embryoid stage CA and in differentiated cells. In cell line 1 (and to some extent line 2 after Cre exposure), Nestin expression showed a similar pattern. Cell line 2 was not exposed to the Cre recombinase, however, Nestin mRNA was most abundant at the final time point for differentiation (**Figure 47**). As for Synaptophysin, we anticipated our neuronal cells derived from iPSCs should express this neuronal-specific mRNA at high levels. Interestingly, this gene was expressed at the highest levels in lines 1 and 2 after exposure to Cre recombinase (**Figure 48**); these cells expressed significantly more Synaptophysin than J1 cells. Finally, we anticipated that at the differentiated time point, our cells would overexpress Blbp. There was a significant increase in Blbp mRNA level at the final neuronal differentiation stage compared to the two earlier stages for J1 ES, iPSC line 1, and iPSC line 2 before adenoviral cre (**Figure 49)**. There was slight increase in iPSC line 2 after transgene excision. Taken together, these expression profiles suggest that although our various iPSC lines could differentiate into neuronal-like cells, some conditions elicited variation in the expression profiles of the neuronal markers.

Discussion

Since the development of pluripotent stem cells by the Yamanaka group [95], a great number of studies sought mechanisms and clinical applications of cellular reprogramming.

Identifying a reliable method to generate high-fidelity iPSCs is critical in order to implement these cells in patient-specific cell therapy in the clinic.

To track the reprograming kinetics and detect gene silencing in this study, we co-expressed each of the four reprograming factors with a specific fluorescence protein. After transduction of the four factors, we found that Oct4 (GFP) and Sox2 (mOrange) fluorescence became visible early on compared to Klf4 (mTagBFP) and c-Myc (mKate). Most of the transduced fibroblasts changed their mesenchymal morphology to resemble epithelial cells, and this MET transition is the hallmark of iPS reprogramming, showing that de-differentiation was proceeding as predicted [111, 196].

As cellular reprogramming comes to completion, fluorescence from tagged transgene expression should diminish (transgenes silenced/transgenes "off"), while the levels of endogenous OSMK should increase significantly. We observed random silencing. In our culture conditions, we found that silencing of transgenes occurred randomly. Some colonies had all four transgenes silenced ("off") while others had an active combination of one, two, three, or all four transgenes. Additionally, in a side-by-side comparison of reprogramming kinetics, plates treated with valproic acid, an HDAC inhibitor, had colonies that emerged earlier and in greater numbers. Most of these colonies however were highly fluorescent even after reprogramming was complete suggesting the transgenes were not silenced effectively (transgenes "on" or active). VPA causes chromosome structure to relax, allowing more general epigenetic changes, which explained that plates treated with VPA had higher number of colonies with high fluorescent signals. The phenomenon of transgene silencing is another way for the cells to maintain appropriate gene dosage of pluripotent genes, and it is potentially mediated by changes in chromatin structure and epigenetic states (methylation and deacetylation). Silencing appeared bias toward transgenes and

tissue specific genes. However, the exact mechanisms to bring about these changes are not well defined, nor are the methods cells use to detect the relative abundance of pluripotency gene products. It could be possible that there is a selective advantage for the cells that had optimum gene expression to survive and proliferate. However, this still does not explain why some iPS colonies still thrive and proliferate when the transgenes were highly active.

While it is not clear how cells regulate pluripotency, relationships between OSMK genes have been observed. First, Oct4 is a key player in pluripotency and Sox2 maintains pluripotency by regulating Oct4 levels [55, 197]. The published literature indicates that Oct4 and Nanog also positively regulate each other. Takahaski and Yamanaka et al. in 2006 suggested that c-Myc proteins relax chromatin structure of somatic cells which facilitates Oct4 and Sox2 binding on target genes and Klf4 assists in upregulating ES cell-specific genes. However, other studies showed that c-Myc is dispensible but significantly increases the frequency and kinetics of reprogramming. Absence of Oct4 abolished cellular reprogramming. How do cells identify transgenic OSMK for regulation or silencing, especially since integration is mostly random in active gene regions. Those colonies could have the machinery in place needed to methylate and condense these gene regions to keep the cell in a pluripotent state while keeping the endogenous pluripotency genes, especially Oct4, active. Additional study on the differences in promoter elements and epigenetic marks which make Oct4 resistant to silencing would be informative. It could be that cells with the most favorable dosage are selected for. In our study, VPA enhanced the kinetics and efficiency of cellular reprograming, but was not effective in silencing transgenes. This result suggests transgenes silencing be a global epigenetic change.

As alluded to earlier, gene dosage is critical for cellular reprogramming. In cell line 2, which contained the constitutively active OSMK transgenes, qRT-PCR data showed Oct4

transgene was expressed remarkably higher than the Sox2, Klf4, and c-Myc transgenes. However, the protein level of Oct4 by Western blot was not strikingly different from the other proteins (**Figure 40**). Sox2 mRNA levels, but not protein levels, seemed higher in cell line 1 prior to excision of the transgenes which could possibly be a result of post-translational modifications, alternatively, be from lentiviral integration near enhancers which increase Sox2 levels pre-excision. Regardless, this change in Sox2 mRNA levels did not interfere with the ability of cell line 1 to differentiate into neuronal cells or form teratomas in SCID mice. Post-transcriptional or post-translational modification likely contributed to the differences observed between mRNA levels and protein levels. Interestingly, we found that in cell line 2 prior to the excision of the transgenes had no expression for Nanog, Esrrb, or Lamin B (**Figures 39-42**). However, these three proteins were expressed after transgenes removal. This phenomenon might be a direct effect of Oct4 overexpression. Further experimentations of Oct4 overexpression or knockdown in these iPSC lines are needed to validate this theory.

The above results suggested that overabundance of Oct4 downregulated pluripotency. To test this possibility, we overexpressed of Oct4-GFP in J1 ES cells and observed no striking change in Nanog mRNA levels. However, all GFP-positive cells that efficiently expressed Oct4 were spread out like fibroblasts or the “cooked-egg” in appearance suggestive of differentiation. This observation agreed with studies that showed that two-fold increased dosage of Oct4 triggers differentiation of iPSCs. The lack of change in Nanog expression could be because Oct4 does not directly inhibit Nanog and instead inhibits Esrrb. The nuclear receptor Esrrb has been used together with Oct4 and Sox2 to convert MEFs into iPSCs. Esrrb mediates reprogramming by upregulating ES-cell specific genes, especially Nanog [195]. A decrease in Esrrb should downregulate Nanog, so the changes in Nanog levels might a delayed in response. Another

possibility is that another factor besides Oct4 negatively regulate Nanog instead. Thus, when Oct4 was being overexpressed, we observed no changes in Nanog levels. Additionally, the active transgenes might act as a sponge that bind to other proteins that are required for the activation of Nanog, Esrrb, and Lamin B genes. Finally, our Oct4-GFP lentivirus might not be functional, thus this does not lead to inhibition of Nanog. In the future, FACS can be used to sort for positively transduced cells.

Lamin B1 expression was absent when high levels of transgenes were expressed. Unlike lamin A and C, Lamin B type proteins are expressed in most cells and are cell-cycle independent. Studies showed that Lamin B might have important roles in cellular processes including regulation of DNA replication, cellular differentiation, cell proliferation, gene expression, and developmental processes [198, 199]. The mechanisms for these processes are not well known and vary between cell types. For example lack of expression of Lamin B in human diploid fibroblasts (HDFs) caused an increase in ROS level, while silencing of Lamin B in mouse fibroblasts caused a decrease in ROS level [198-202]. Mouse fibroblasts with Lamin B1 mutations suffered from aberrant differentiation, increased polyploidy, and premature senescence [200]. Lamin B is essential for the development of the brain and lung as the deletion of Lamin B1 and/or B2 gene prevents nuclear elongation in neurons, neuronal migration and organization of the brain, and proper mitotic spindle orientation in neuron progenitors [201-204]. Silencing of Lamin B in HeLa cells caused apoptosis. Conditional knockout of Lmnb1 in mouse skin keratinocytes did not affect the cell's formation in chimeric mice, but complete knockout of Lamin B is lethal.

The active transgenes could be the reason for the lack of Lamin B expression in our cells. Although the regulation and mechanism of Lamin B still needs further study, it has been shown

that Lamin B is repressed by p53 and p21. In iPS generation, c-Myc activates p53 and p21, potentailly inhibiting Lamin B. This repression might be lifted when the transgenes are excised, restoring Lamin B expression. This theory could be tested by generating iPSCs without using c-Myc as a reprogramming factor. It has not been shown that pluripotency factors bind to Lamin B promoter and regulate its expression. Experiments using chromatin immunoprecipitation and sequencing (ChIP-Seq) would be informative in identifying OSMK targets in our model. Alternative, the cytoskeletal network has been demonstrated to be important for cellular reprogramming. Oct3/4 and Sox2 been shown to inhibit Snail, a key factor in the epithelial-to-mesynchemal transition (EMT). This benefits iPS reprogramming as it is the reverse process, a mesynchemal-to-epithelial transition (MET). C-Myc also enhances MET by inhibiting TGF-β, while Klf4 upregulates genes associated with epithelial cells including E-Cadherin. Thus, Lamin B could potentially be another cytoskeleton gene needing careful regulation to form proper iPSCs.

It has been established that Nanog is essential for iPSC self-renewal and pluripotency [205-207]. A recent publication suggested that Nanog is not required for iPSC line culture *in vivo* [208]. The authors found that these Nanog-deficient iPSC lines are not only transciptionally similar to ESC lines but they also can support teratoma generation. Our neuronal differentiation result supported these findings since when the Nanog level was restored in cell line 2 after the transgenes excision, these cells did not have the same differentiation and proliferation potential as iPSC line 1. These cells exhibited neuron morphology during the first 48 hours but failed to proliferate. However, our *in vivo* teratoma assay showed after the excision of the transgenes iPSC line 2 give rise to teratomas in SCID mice (**Figure 43**). The differences in the *in vitro* and *in vivo* studies in which iPSC line 2 after transgene excision failed to differentiate effectively

into neurons but was able to differentiate and formed teratomas in SCID mice might suggest differences in the environment the cells were in. Our in vitro condition might not provide all the necessary components for robust neuronal differentiation. On the other hand, not all iPSC lines generated are equal; thus, some might be more efficient in differenting into different cell types than others. Finally, there might be multiple factors that are co-regulated during differentiation, this regulation might be pertubed in presence of active transgenes.

Our assays of neuronal differentiation were also intriguing. The expression patterns for Pax6 and Trk2b in cell line 1, before and after transgene excision, were as expected; these were the molecular markers that Bibel et al used as readouts for successful neuronal differentiation. The level of Pax6 was initially high in early hours of iPSCs differentiation, but decreased as time progressed (**Figure 45**). The reverse held true for Trk2b (**Figure 46**). Interestingly, iPSC line 2 [before and after transgene excision] had the morphology of early neurons with protruding dendrites; the majority of iPSCs without transgenes formed neurons more readily than iPSCs still expressing high levels of transgenes. However, under both conditions the differentiated cells failed to proliferate in culture as differentiation progressed. These neurons were much more flimsy and do not adhere well to plates. Our substitution in culture ingredients might also not support robust neuronal growth or the overabundance of the transgenes might interfere with the expression of neuronal related genes. Interestingly, to compensate for detaching cells, we doubled the cell concentration seeded for cell line 2. We found that cell line 2 [before transgene excision] formed a sheet of epithelium cells. Thus we decided to look at expression of other neuronal markers including Nestin, Synaptophysin, and Blp. Nestin is primarily expressed during early CNS development so perhaps the chaotic transgene expression in line two “freezes” the cells in a premature state—hence, a high level of Nestin. Second, although Nestin is expressed

primarily in the CNS, it is also expressed in keratinocytes and endothelial cells. The qRT-PCR results showed a significant upregulation of Nestin in iPSC line 2 not exposed to Cre (**Figure 48**). We observed significant amplification of Synaptophysin in neurons derived from iPSC line 1 after Cre addition (**Figure 49** – third red bar). These results suggest that different iPSC lines have different differentiation potential and fates.

Although our design was effective in generating iPSCs, the use of viruses remains a concern. Despite the fact that the lentiviral vector used in this system have truncated LTR promoters to abolish any promoter/enhancer activity and Cre recombinase is used to remove the majority of the vector from the host genome, removal of the vector is not complete. The LTR sequences flanked outside the LoxP site will remain permantly in the host genome. While lentiviral usage did not lead to cancer [179, 186], leaving these non-coding sequences might interfere with the regulation or expression of genes nearby. While adenovirus or other non-viral methods can be used to deliver the exogenous genes, these methods often do not generate stable expression of OSMK, are costly, and the integrity of the resulting iPSCs is questionable. We propose to use the CRISPR/CAS9 or TALEN system to directly insert OSMK cassettes to targeted homologous recombination sites at the Rosa locus or other gene desert regions. Extending our findings in this direction should develop safer and more reliable methods for controlling the overexpression of ectopic genes.

The goal was to demonstrate the fluorescent proteins as a way to characterize reprogrammed colonies with or without transgenes and to show that the latter are better reprogrammed. The iPSC lines with lack of fluorescent signals should be selected for patient cell-based therapy. Indeed, we found silenced cells were more effective and efficient at differentiating to neuronal cells. There was a loss in expression of Nanog, Lamin B and Esrrb in

iPSC line with active transgenes. The presence of high OSMK transgenes inhibits teratoma development while silenced transgenes promote it. When comes to generating iPSCs we think it is best to avoid using small molecules that have global effects as VPA usage interfered with gene silencing. Finally, when it came to selecting iPSC lines for study, rigorous differentiation studies should be done to determine which cell line is most applicable.

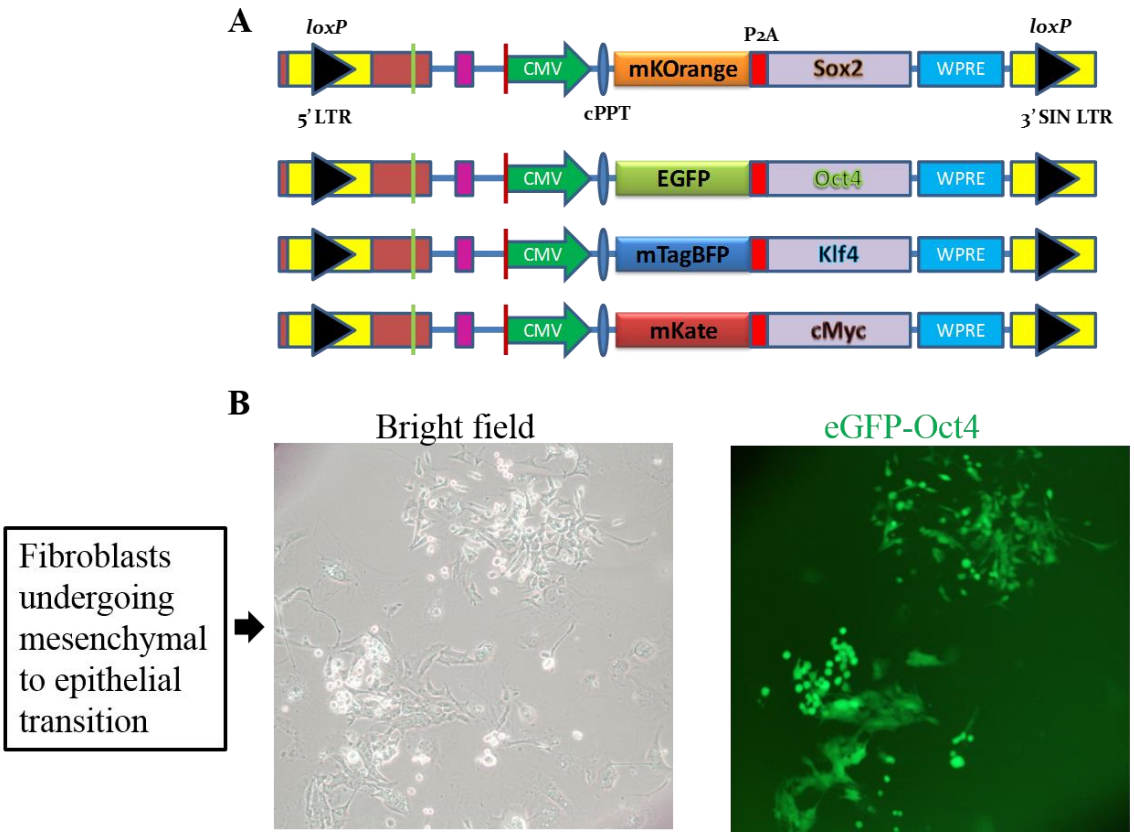


Figure 33. Vector design and strategy used for the reporter candidate. A) Schematic of reprogramming strategy, O/S/M/K tagged with fluorescence protein in P2A. The lentiviral backbone embbed a loxP site within the 3' SIN LTR. B) MEFs undergoing mesenchymal transition to become epithelial cells exhibited high fluorescence

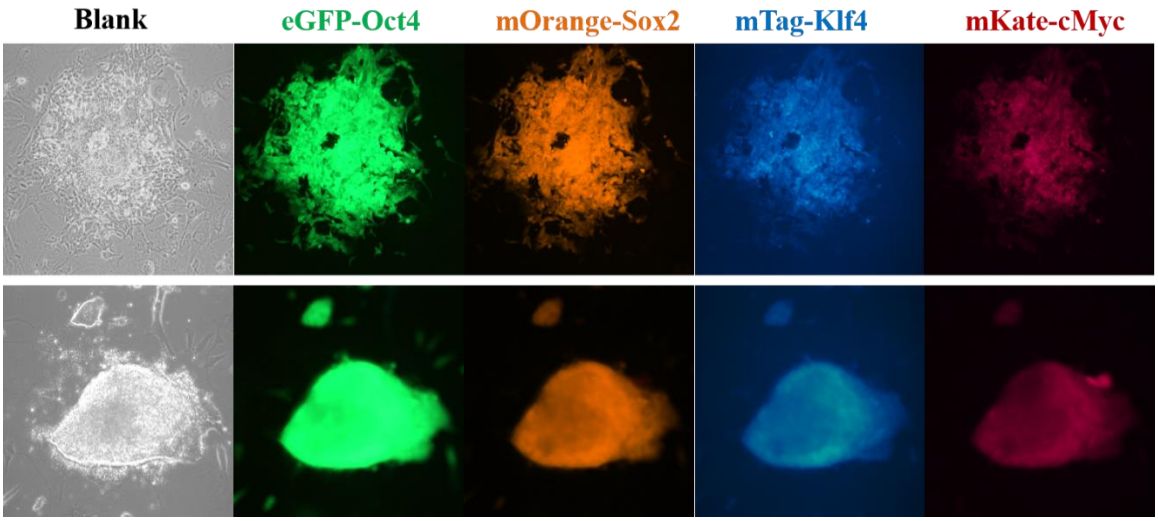


Figure 34. Expression of fluorescence proteins in early stages of iPS formation. As cellular reprogramming occurs the fluorescence proteins are brightly expressed. Bright field and fluorescence images of MEF after transduction with OSMK and incubation in ESC medium. Top panel represented early reprogramming event when cells are undergoing early morphological changes. The bottom panel represented the later stages of reprogramming where cells formed rounded clusters. The four reprogramming factors (OSMK) were highly expressed as evidenced by fluorescence signals emitted: Oct-GFP, Sox2-mOrange, Klf4-mTagBFP, and cMyc-mKate.

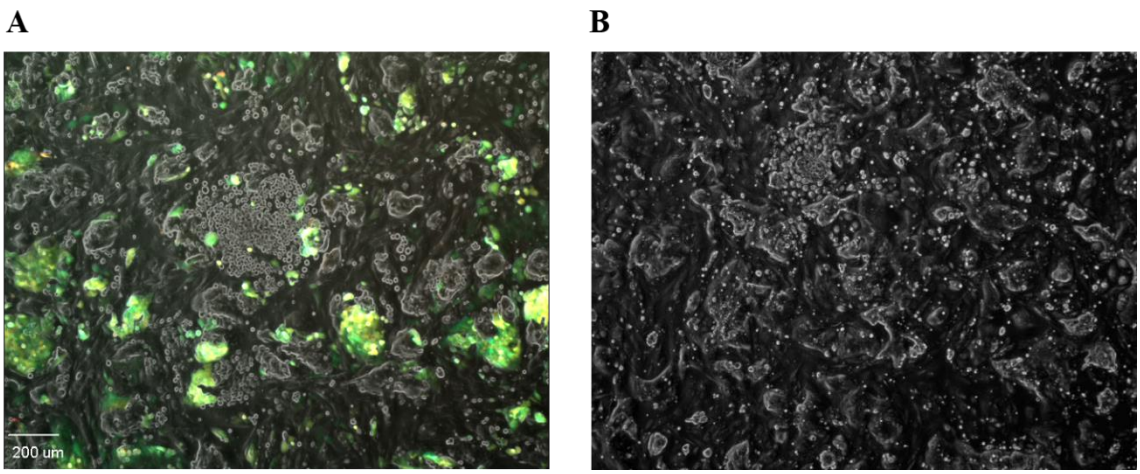


Figure 35. Expression of transgenes at the end of reprogramming. Confocal microscopy of established iPSC lines expanded from ES-like iPS colonies. Some colonies still exhibited high fluorescence – transgenes are active (A), while others did not – transgenes are silenced (B).

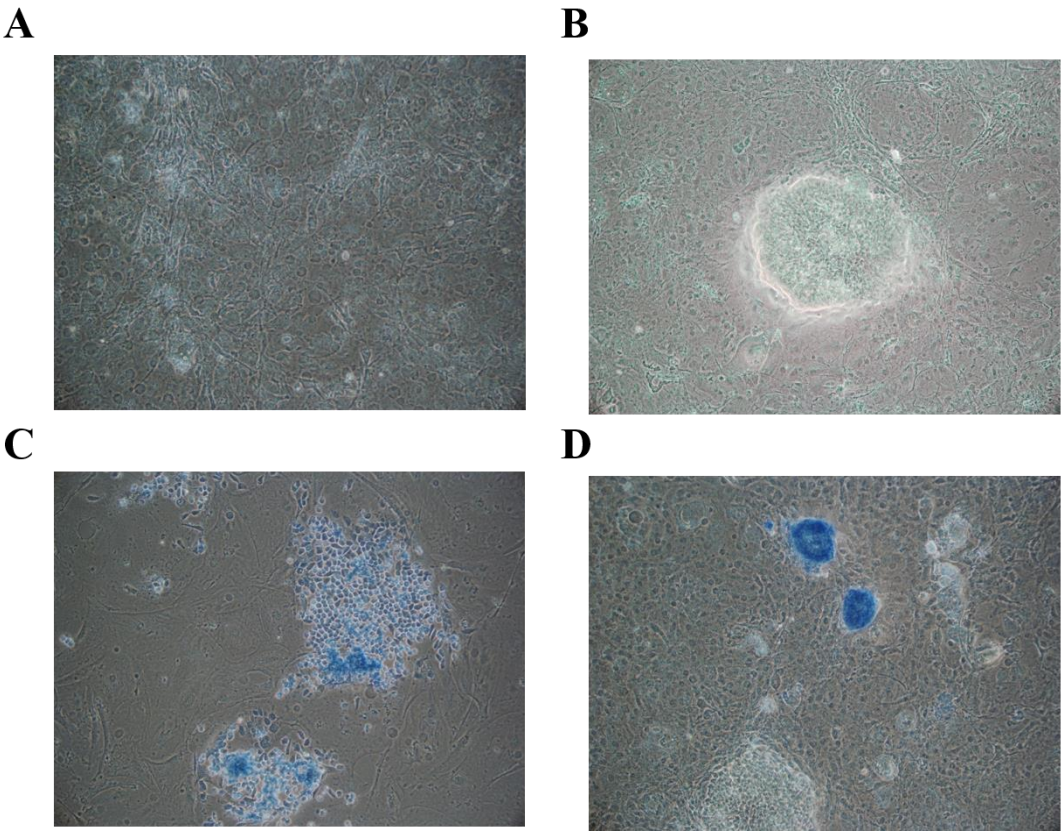


Figure 36. iPS colonies stained positive for alkaline phosphatase (AP). At the end of reprogramming each plate was stained with AP, a stem cell marker. Plates that contained no iPS colonies or only MEF did not stain with AP (A). Some ES-like colonies did not stain with AP (B). Representative iPS colonies did stain positive with AP (C, D).

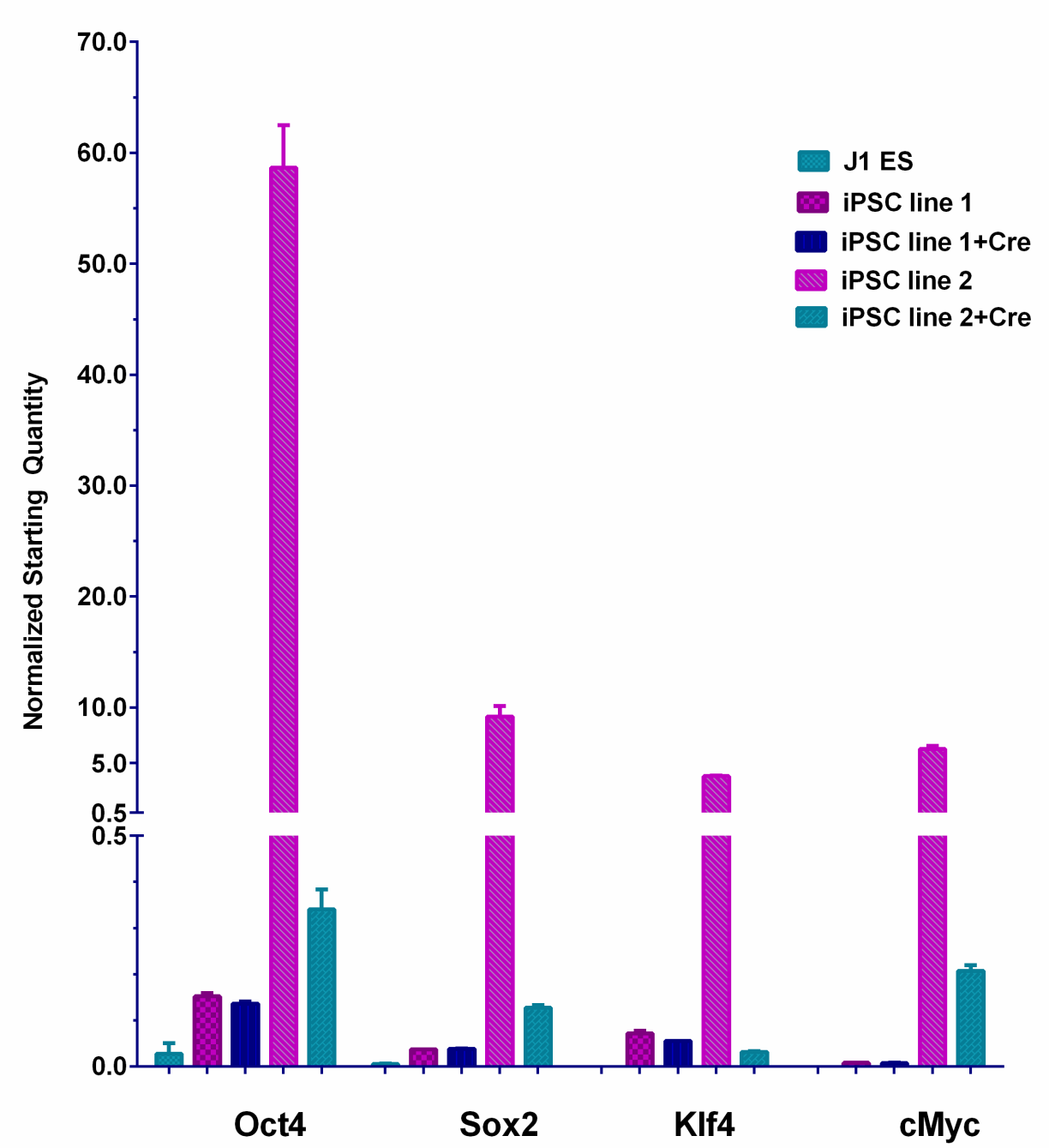
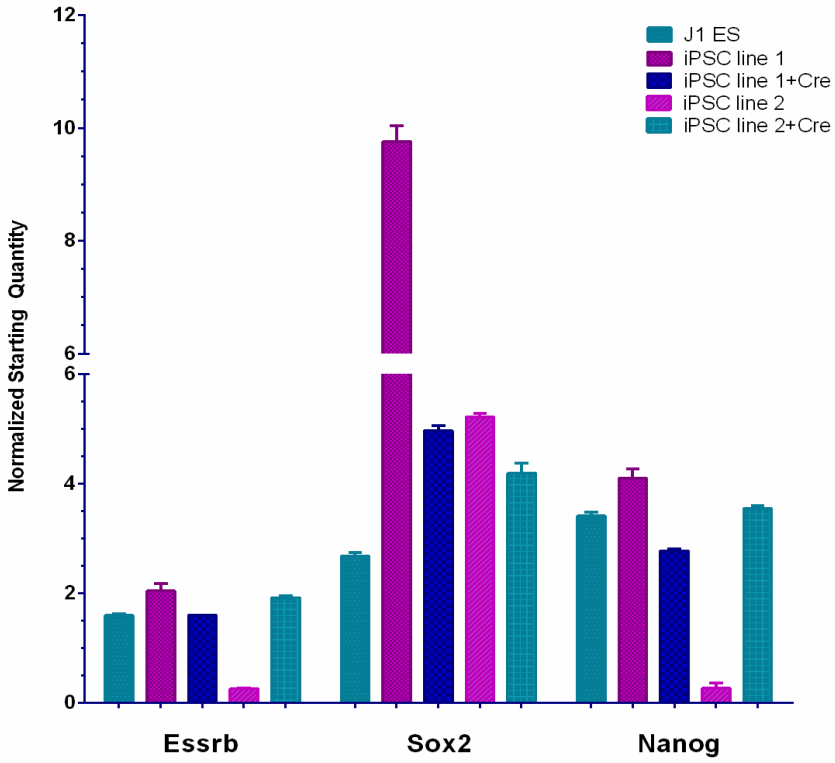


Figure 37. Low transgenes expression in silenced iPSC lines. Realtime PCR detected changes in mRNA levels of transgenes Oct4, Sox2, Klf4, and cMyc in J1 ESCs (first bar), iPSC line 1 (second square bar), iPSC line 1 + Cre (third bar), iPSC line 2 (fourth bar), and iPSC line 2 + Cre (fifth bar). The amount shown was normalized to the housekeeping gene E1A binding protein p300. Samples were prepared in duplicates.


J1 ES
iPSC line 1
iPSC line 1+Cre
iPSC line 2
iPSC line 2+Cre
Normalized Starting Quantity
12
10
8
6
6
4
2
0
Essrb
Sox2
Nanog

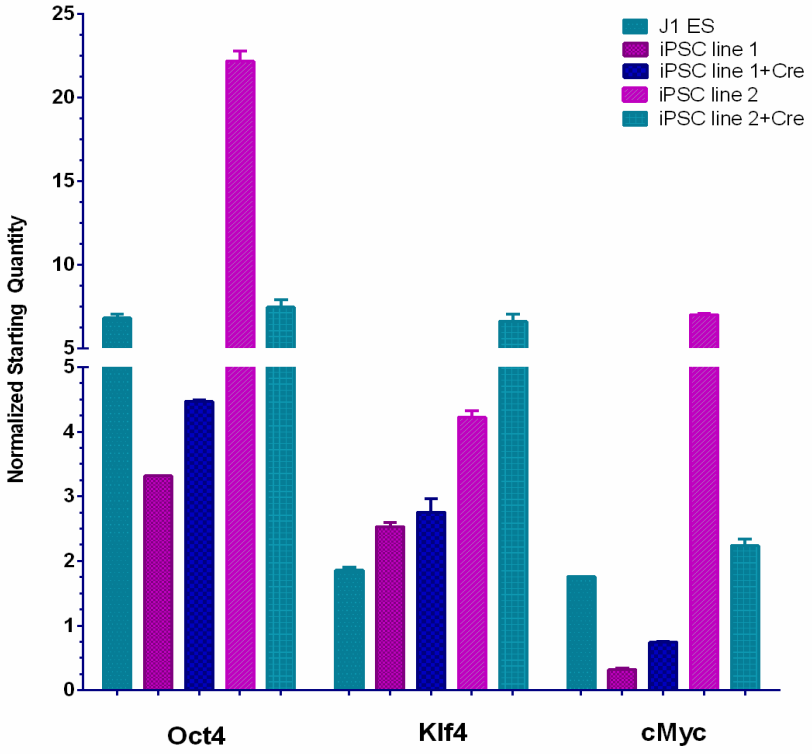
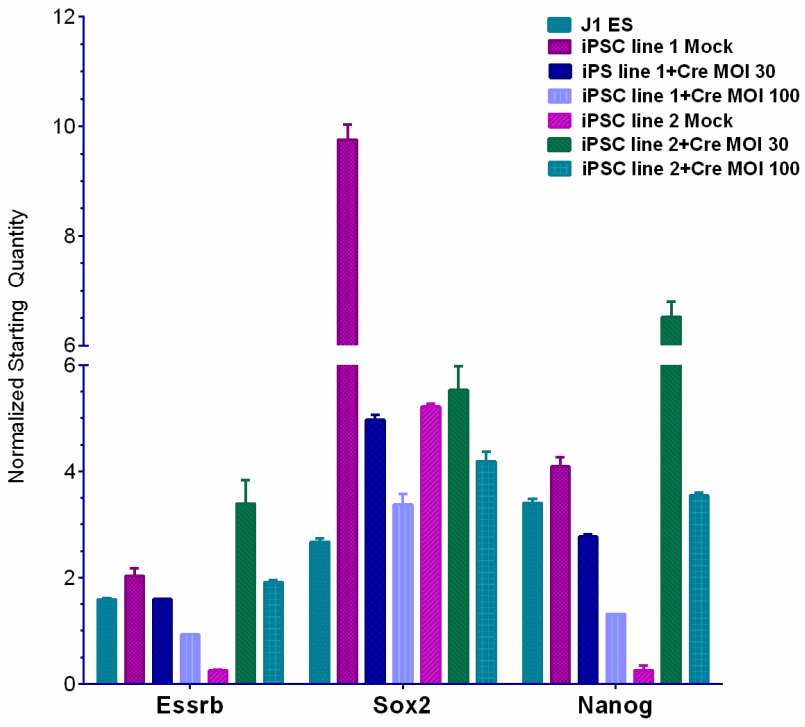

J1 ES
iPSC line 1
iPSC line 1+Cre
iPSC line 2
iPSC line 2+Cre
Normalized Starting Quantity
25
20
15
10
5
5
4
3
2
1
0
Oct4
Klf4
cMyc

Figure 38. Silenced iPSCs and ESCs share an expression profile for key markers of differentiation. Realtime PCR of mRNA showing Esrrb, Sox2, and Nanog (top graph), Klf4, cMyc, and Oct4 (bottom graph) in J1 ESC line (first bar), iPSC line 1 (second square bar), iPSC line 1 + Cre (third bar), iPSC line 2 (fourth bar), and iPSC line 2 + Cre (fifth bar). The amount shown was normalized to the housekeeping gene E1A binding protein p300. Samples were prepared in duplicates.


J1 ES
iPSC line 1 Mock
iPS line 1+Cre MOI 30
iPSC line 1+Cre MOI 100
iPSC line 2 Mock
iPSC line 2+Cre MOI 30
iPSC line 2+Cre MOI 100
Normalized Starting Quantity
12
10
8
6
6
4
2
0
Essrb
Sox2
Nanog

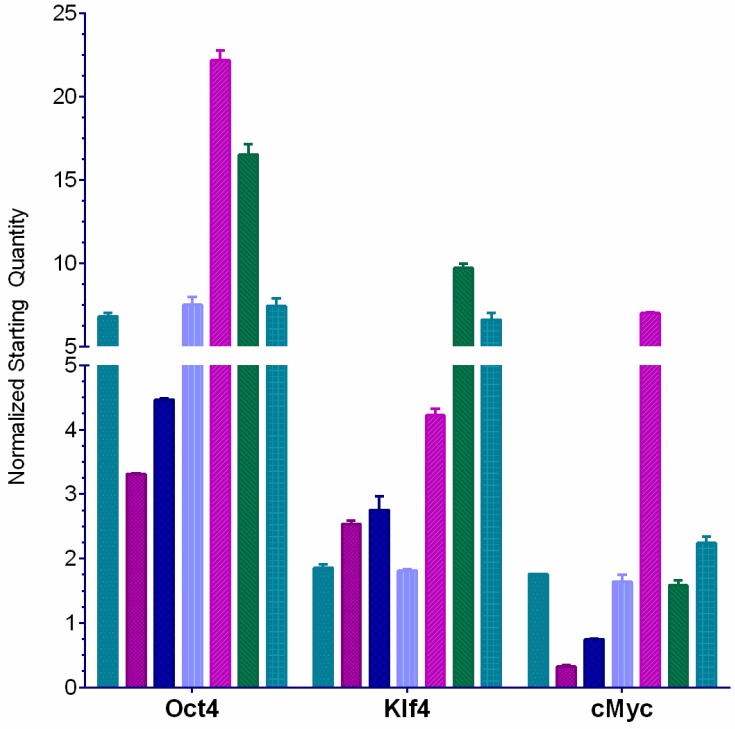

Normalized Starting Quantity
25
20
15
10
5
5
4
3
2
1
0
Oct4
Klf4
cMyc

Figure 39. Complete data for total gene expression levels at different adenviral cre multiplicity of infection (MOI). This is the expanded data from Figure 38. Silenced iPSCs and ESCs share an expression profile for key markers of differentiation. Realtime PCR of mRNA showing Esrrb, Sox2, and Nanog (top graph), Klf4, cMyc, and Oct4 (bottom graph) in J1 ESC line (first bar, teal), iPSC line 1 (second square bar, dark purple), iPSC line 1 + Cre at MOI 30 (third bar, dark blue), iPSC line 1+Cre at MOI 100 (fourth bar, light blue), iPSC line 2 Mock (fifth bar, purple), iPSC line 2+Cre at MOI 30 (sixth bar, green), and iPSC line 2+Cre at MOI 100 (seventh bar, teal stripes) . The amount shown was normalized to the housekeeping gene E1A binding protein p300. Samples were prepared in duplicates.

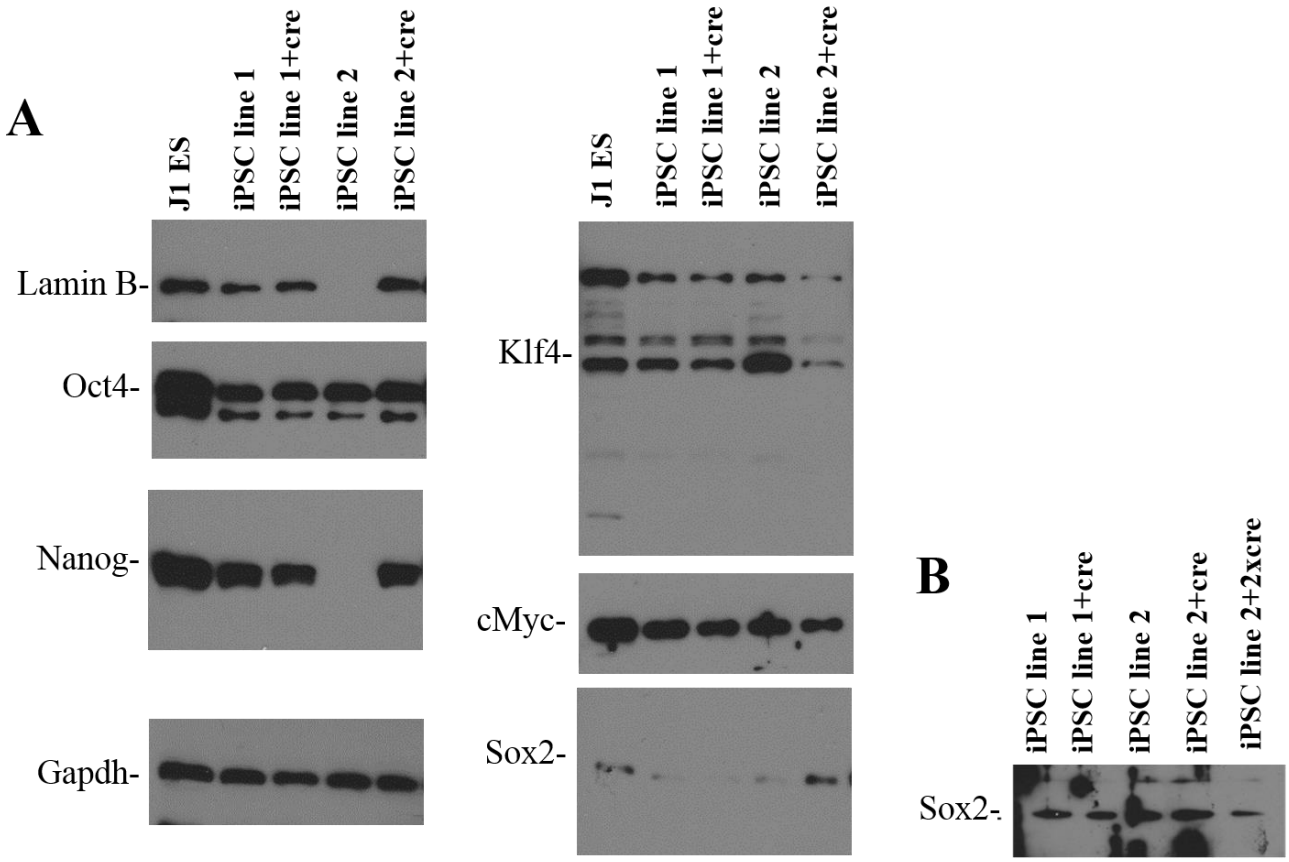


Figure 40. Nanog level is perturbed in iPSC line with high transgene expression**.** (**A**) Whole-cell lysates obtained from J1 ES, iPSC line 1, iPSC line 1+Cre, iPSC line 2, and iPSC line 2+Cre. Approximately 5 μg of protein lysate were loaded into 12% SDS-Page gel. The membrane was blotted for Oct4, cMyc, Sox2, Klf4, Nanog, Lamin B and Gapdh. Sox2 antibody had significant background and require extra exposure time. It was consistent that Oct4, Sox2, cMyc, Klf4 and Gapdh were expressed in similar amount across different cell lines. There was no detectable band for Nanog in iPSC line 2 prior to the removal of the ectopic genes. (**B**) Repeated experiment showing only membrane blotted for Sox2. The film was exposed overnight to enhance Sox2 signal. There was significant background with the Sox2 antibody. However, it showed that Sox2 protein is present in our tested iPSC cell lines. The Western blot was repeated three times with different whole-cell lysates. Cre = adenoviral cre transfection, 2xCre = the cell line was infected with double the amount of adenovirus original used (to ensure complete excision of transgenes, and determine whether this affect endogenous gene levels).

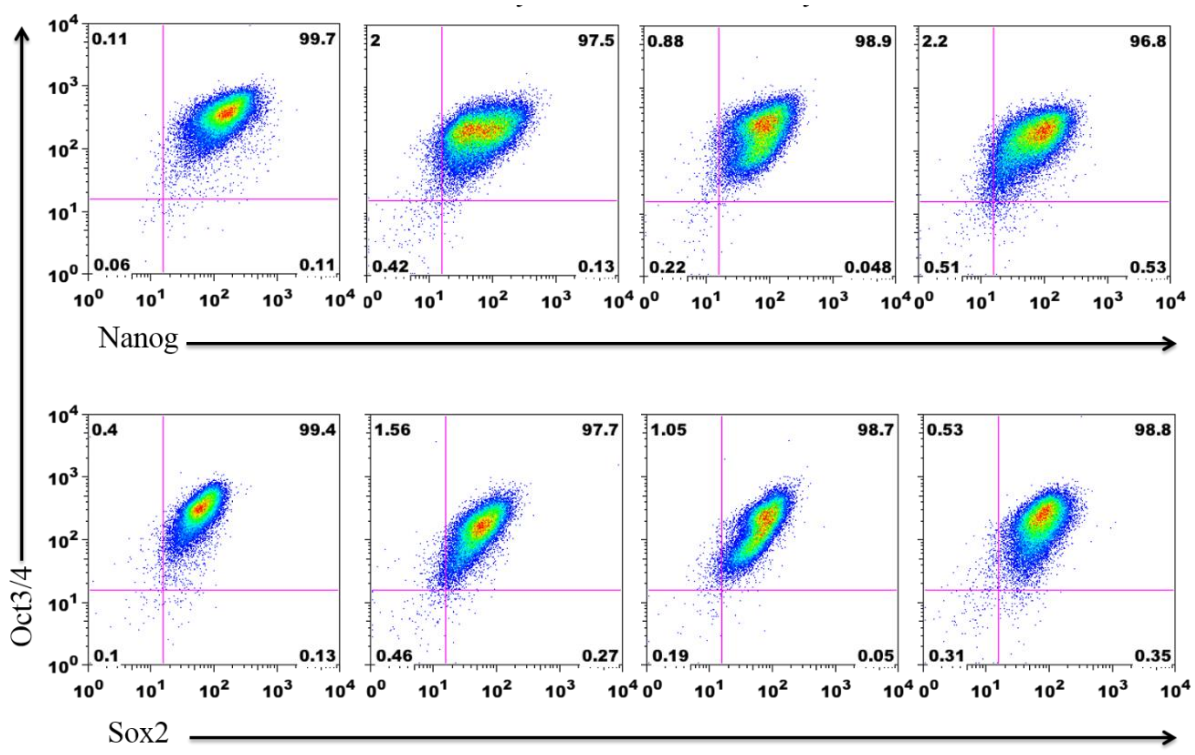


Figure 41. iPSC lines share similar gene expression to ESC lines. Fluorescence activated cell sorting (FACS) plots shown intracellular staining of pluripotency factors in J1 ES (far left), iPSC line 1, iPSC line 2 after adenoviral cre recombinase addition, and HM 1. The top row showed cells population that are double positive for Oct3/4 (Y-axis) and Nanog (X-axis). The bottom row showed cells population that are double positive for Oct3/4 (y-axis) and Sox2 (X-axis).

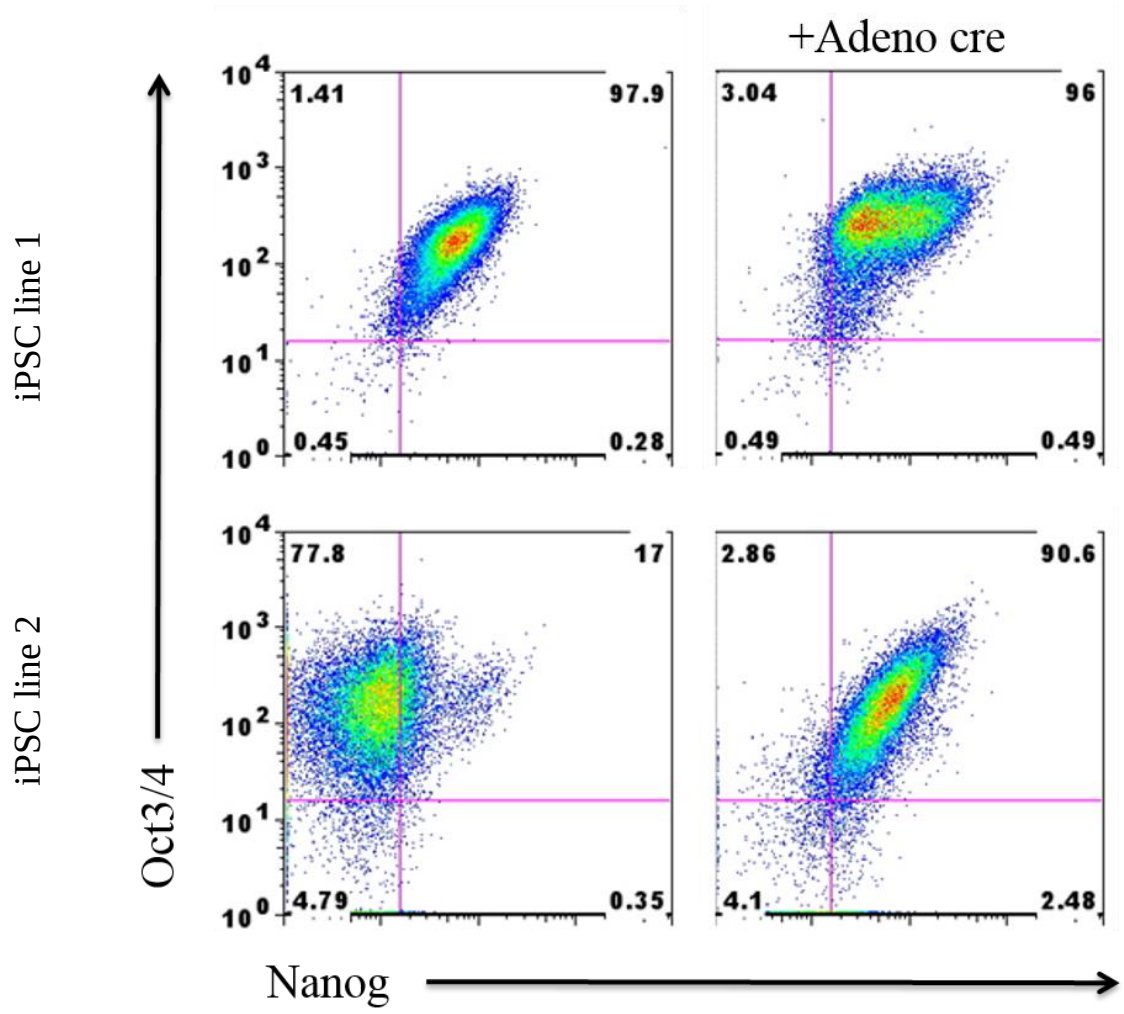


Figure 42. Nanog expression is disturbed in active iPSC line. Fluorescence activated cell sorting (FACS) plots shown intracellular staining of Nanog (X-axis) and Oct3/4 (Y-axis) in iPSC line 1 (top row, transgenes "off") and iPSC line 2 (bottom row, transgenes "on") before (left column) and after excision of transgenes (right column) with adenoviral cre recombinase. Adeno cre = after adenoviral cre infection.

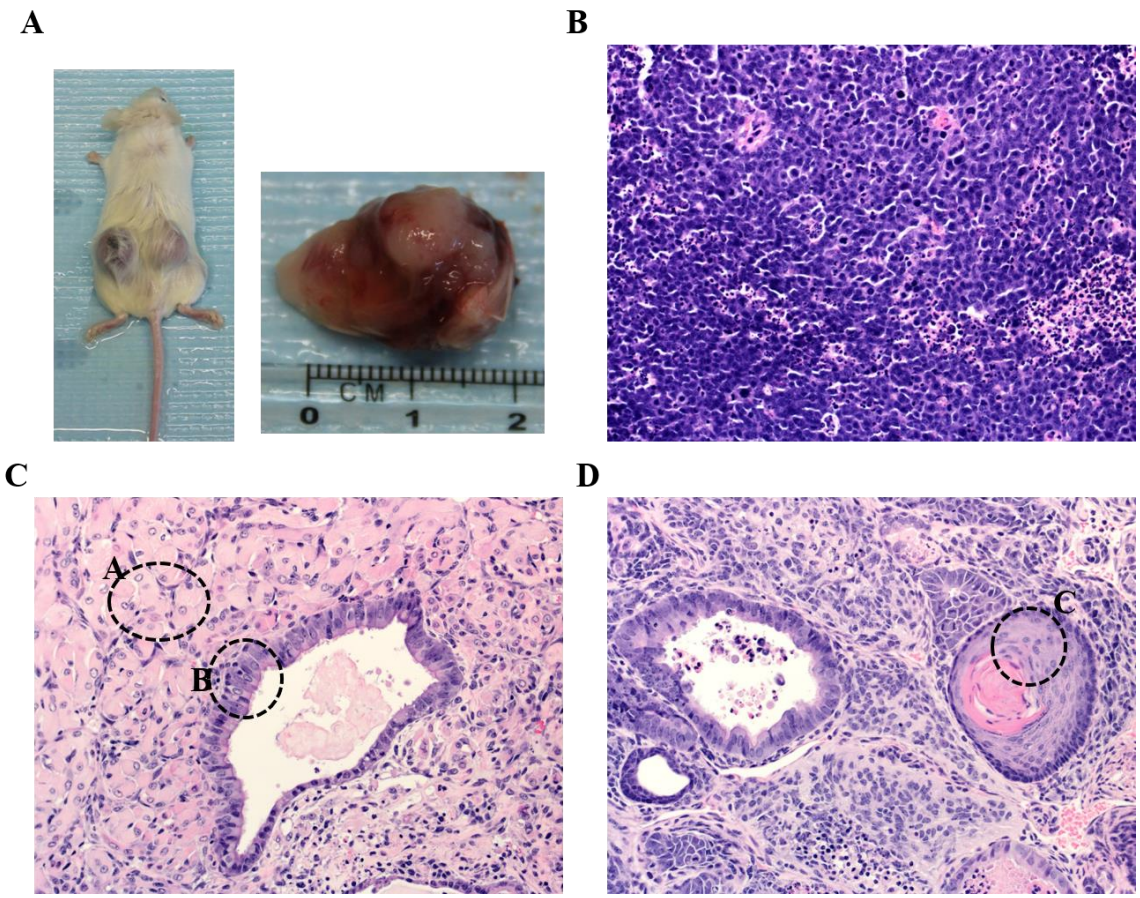


Figure 43. Teratoma assay: absence of transgene expression promotes efficient differentiation. **(A)** IPs clones injected into SCID mice with tumors observed **(B)** IPS clone with high transgene expression **(C&D)** representative images of teratoma from mice injected with iPS clones with transgenes silenced or excised. Images take at 200x resolution under confocal microscope. A-Mesoderm (muscle), B- Endoderm (epithelial in airway lining), C-Ectoderm (stratified squamous epithelium with keratin in the middle)

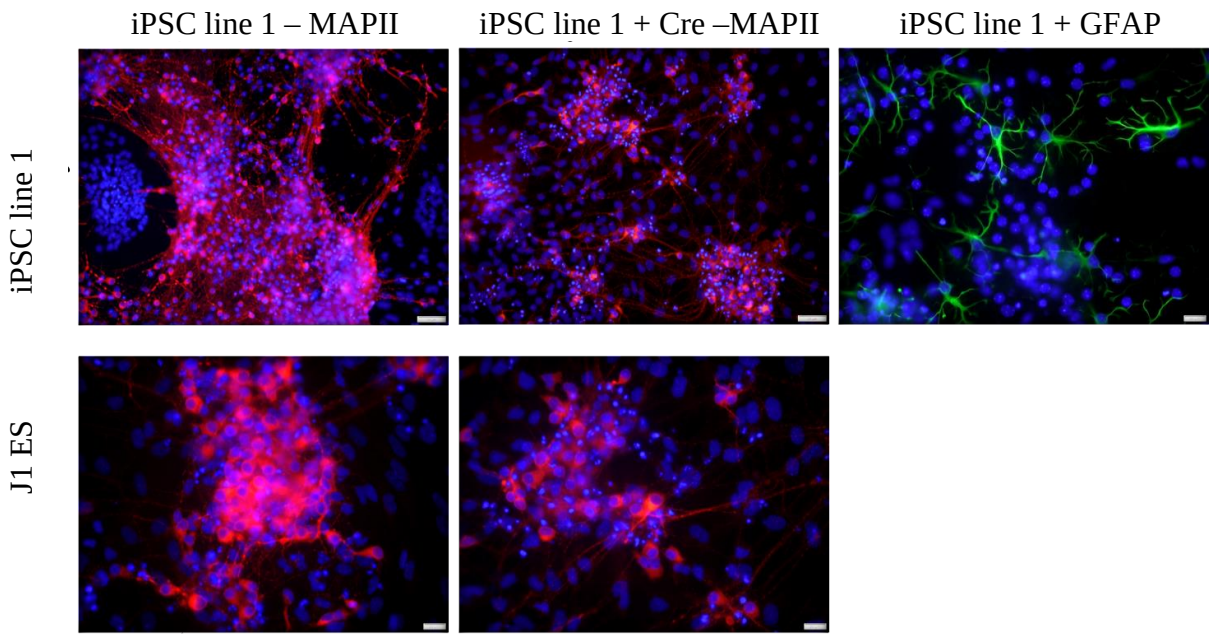


Figure 44. Silenced iPSC line efficiently differentiated into neuronal cells. These iPSCs were coaxed to differentiate into neuronal cells by incubating in N2 and B27 supplements, and images were taken at day 7 in B27 medium. Immunofluorescence images taken of iPS colony (top) stained positive for MAP (dendritics neuronal cells), GFAP (astrocytes) and Dapi (nuclei). Images were captured at 20X. MAP = microtubules associated protein, GFAP = glial fibrillary acidic protein, Dapi = 4',6-diamidino-2-phenylindole.

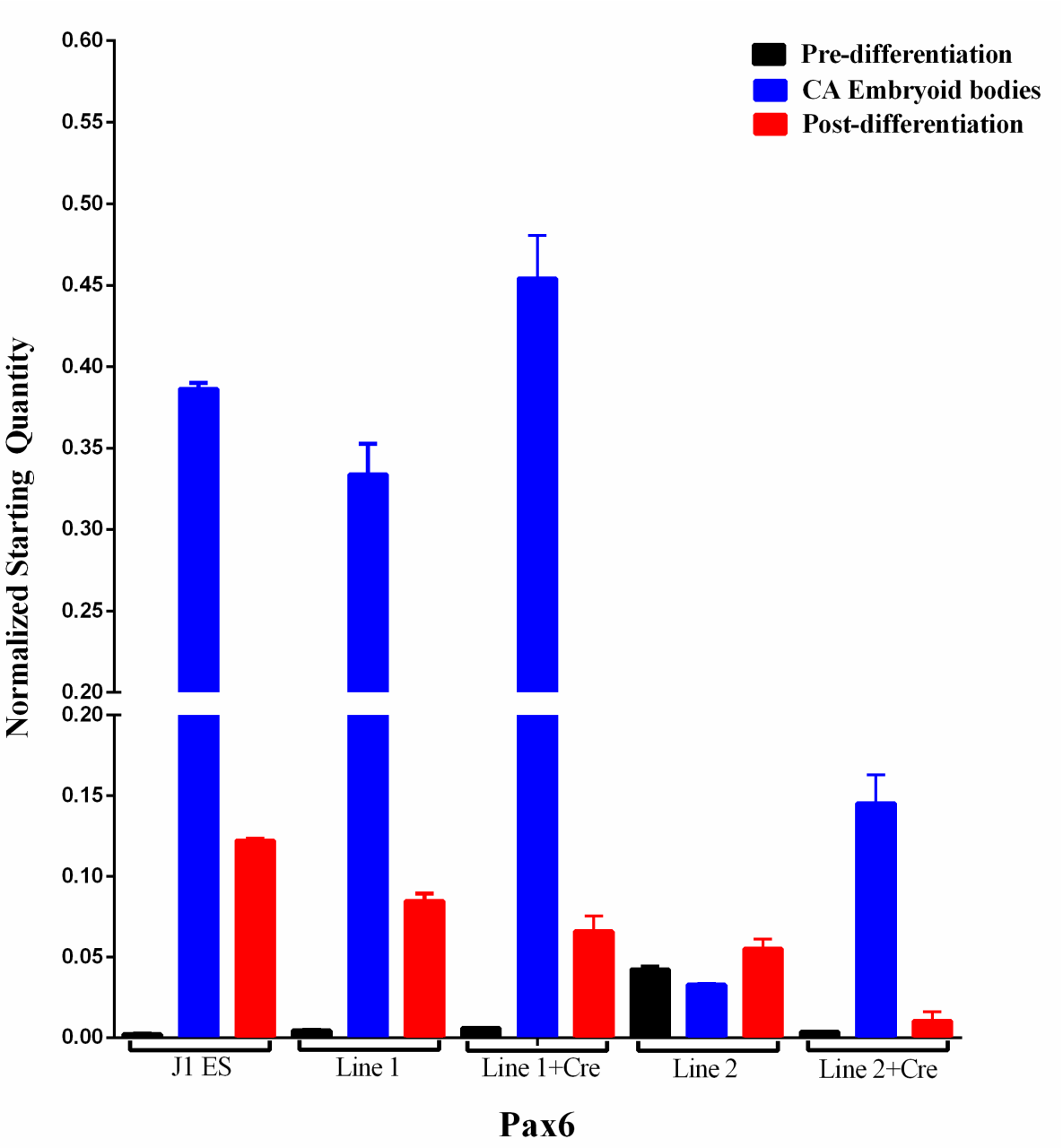


Figure 45. Neuronal precursors (CA) derived from iPSC lines with absence transgenes are positive for Pax6. Quantitative real-time (qRT-PCR) expression levels of Pax6 in iPSCs subjected to neuronal differentiation. RNA was harvested at day 0 (pre-differntiation), CA (embryoid stage) and neuronal stage (post-differentiation). Pax6 expression was examined in J1 ES, J1 ES CA, and J1 ES neuron compared to iPSC line 1 and 2 before and after Cre recombinase. J1 ES served as positive control. Starting quantity amount was normalized to E1A binding protein p300. Samples were prepared in duplicates. Line 1 = iPSC line 1, Line 1+Cre = iPSC line 1 after adenoviral cre transfection, Line 2 = iPSC line 2, Line 2+Cre = iPSC line 2 after adenoviral cre transfection.

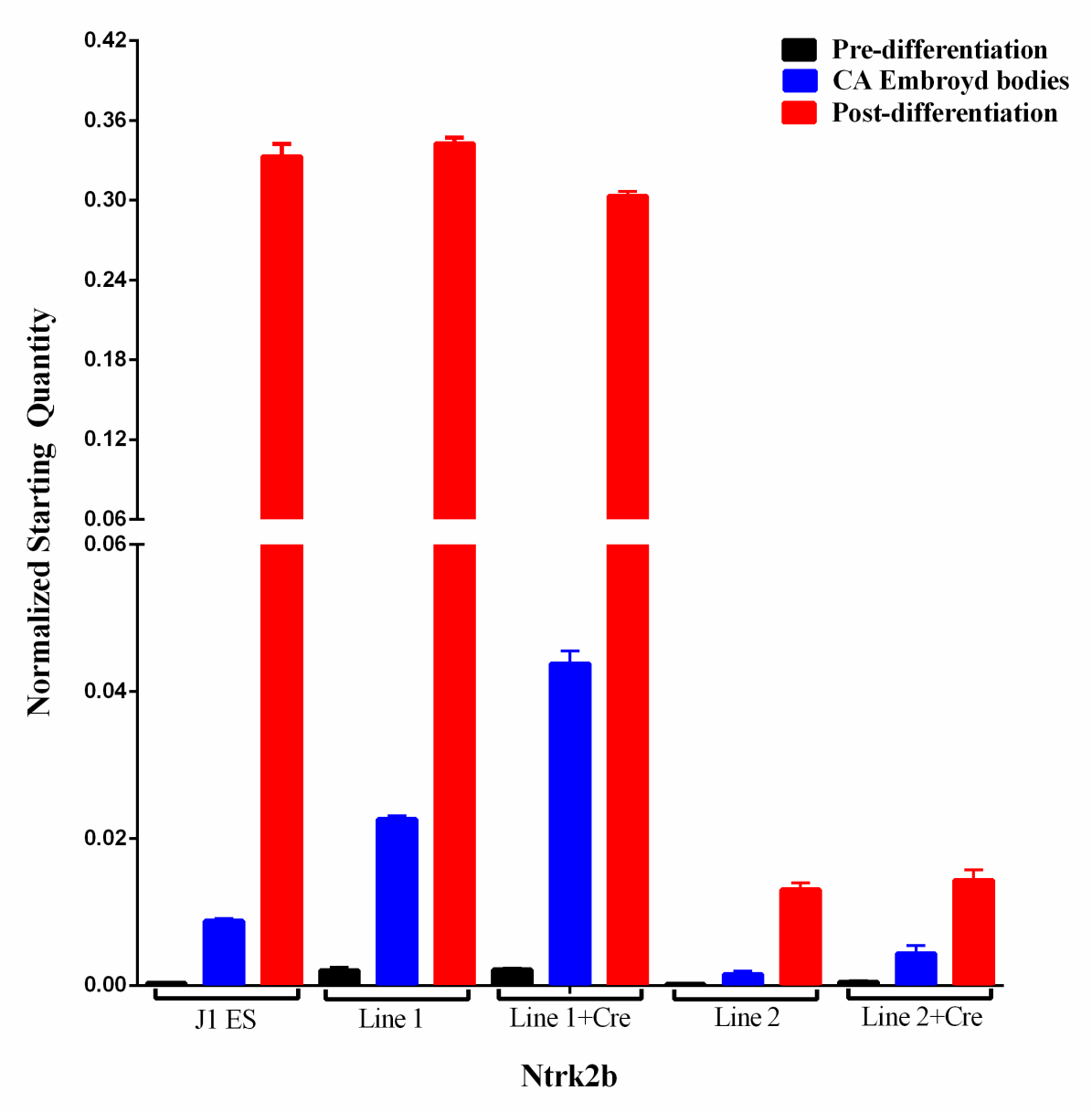


Figure 46. Neuronal cells derived from iPSC lines expressed Ntrk2b. Quantitative real-time (qRT-PCR) expression levels of Ntrk2b in iPSCs subjected to neuronal differentiation. RNA was harvested at day 0 (pre-differntiation), CA (embryoid stage) and neuronal stage (post-differentiation). Ntrk2b expression was examined in J1 ES, J1 ES CA, and J1 ES neuron compared to iPSC line 1 and 2 before and after Cre recombinase. J1 ES served as positive control. Starting quantity amount was normalized to E1A binding protein p300. Samples were prepared in duplicates. Line 1 = iPSC line 1, Line 1+Cre = iPSC line 1 after adenoviral cre transfection, Line 2 = iPSC line 2, Line 2+Cre = iPSC line 2 after adenoviral cre transfection. Ntrk2 = neutrophic tyrosine kinase receptor type 2.

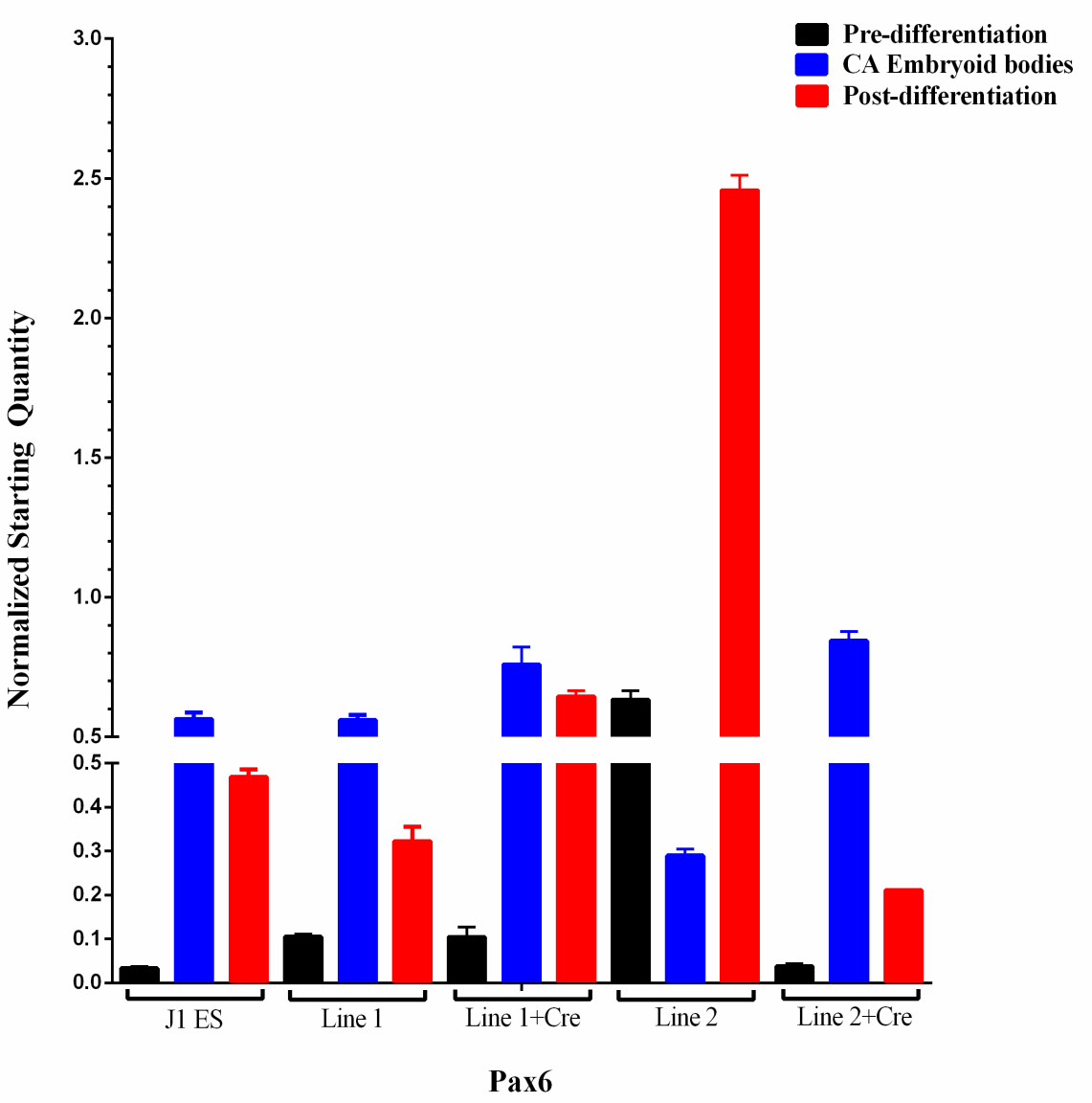


Figure 47. Neuronal precursors derived from iPSC line 2 (transgenes "on") expressed significant amount of Nestin. Quantitative real-time (qRT-PCR) expression levels of Pax6 in iPSCs subjected to neuronal differentiation. RNA was harvested at day 0 (pre-differntiation), CA (embryoid stage) and neuronal stage (post-differentiation). Nestin expression was examined in J1 ES, J1 ES CA, and J1 ES neuron compared to iPSC line 1 and 2 before and after Cre recombinase. J1 ES served as a positive control. Starting quantity amount was normalized to E1A binding protein p300. Samples were prepared in duplicates. Line 1 = iPSC line 1, Line 1+Cre = iPSC line 1 after adenoviral cre transfection, Line 2 = iPSC line 2, Line 2+Cre = iPSC line 2 after adenoviral cre transfection. Nestin = type IV intermediate filament.

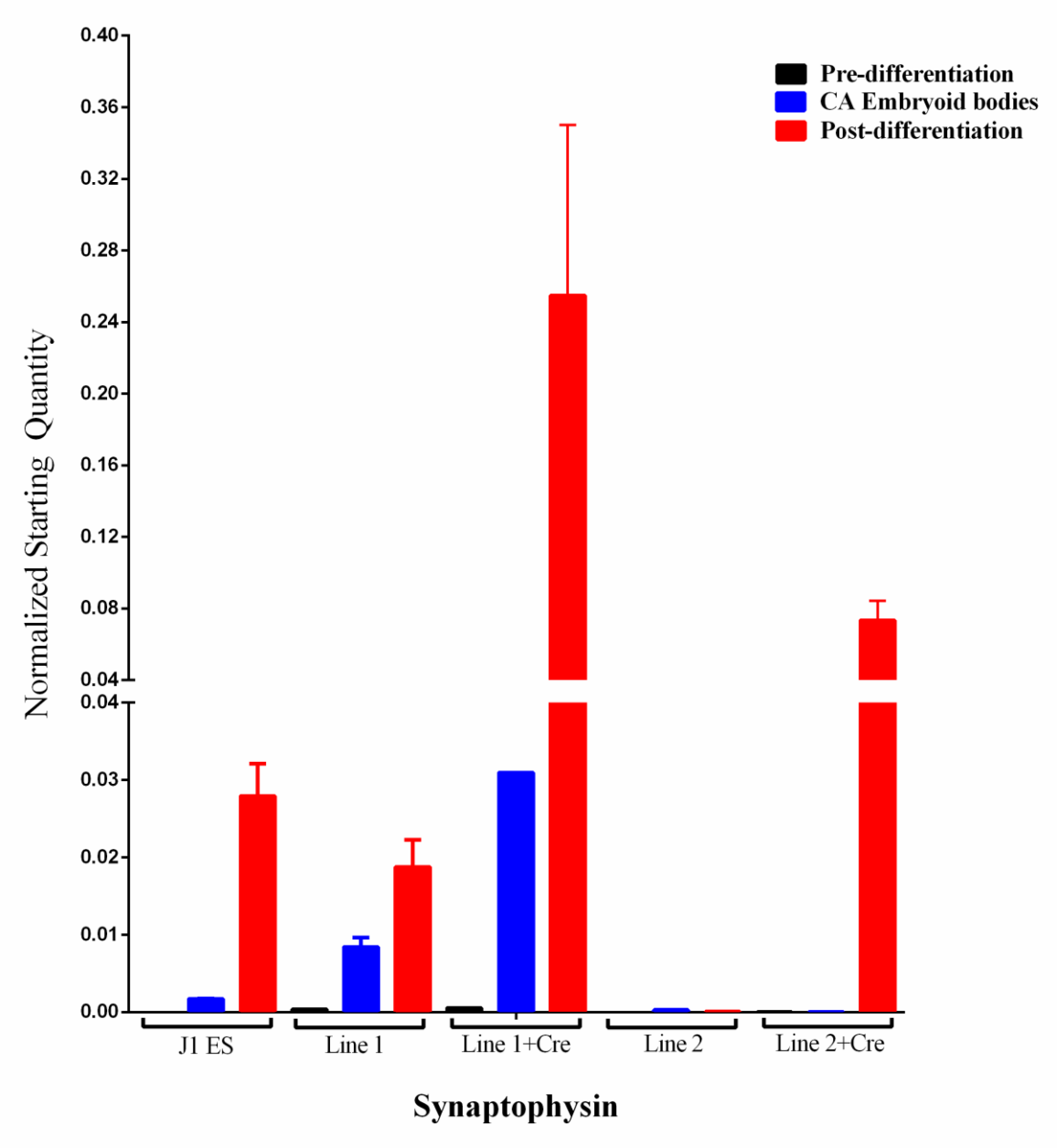


Figure 48. Neuronal cells derived from iPSC lines without active transgenes are positive for synaptophysin. Quantitative real-time (qRT-PCR) expression levels of Pax6 in iPSCs subjected to neuronal differentiation. RNA was harvested at day 0 (pre-differntiation), CA (embryoid stage) and neuronal stage (post-differentiation). Synaptophysin expression was examined in J1 ES, J1 ES CA, and J1 ES neuron compared to iPSC line 1 and 2 before and after Cre recombinase. J1 ES served as positive control. Starting quantity amount was normalized to E1A binding protein p300. Samples were prepared in duplicates. Line 1 = iPSC line 1, Line 1+Cre = iPSC line 1 after adenoviral cre transfection, Line 2 = iPSC line 2, Line 2+Cre = iPSC line 2 after adenoviral cre transfection.

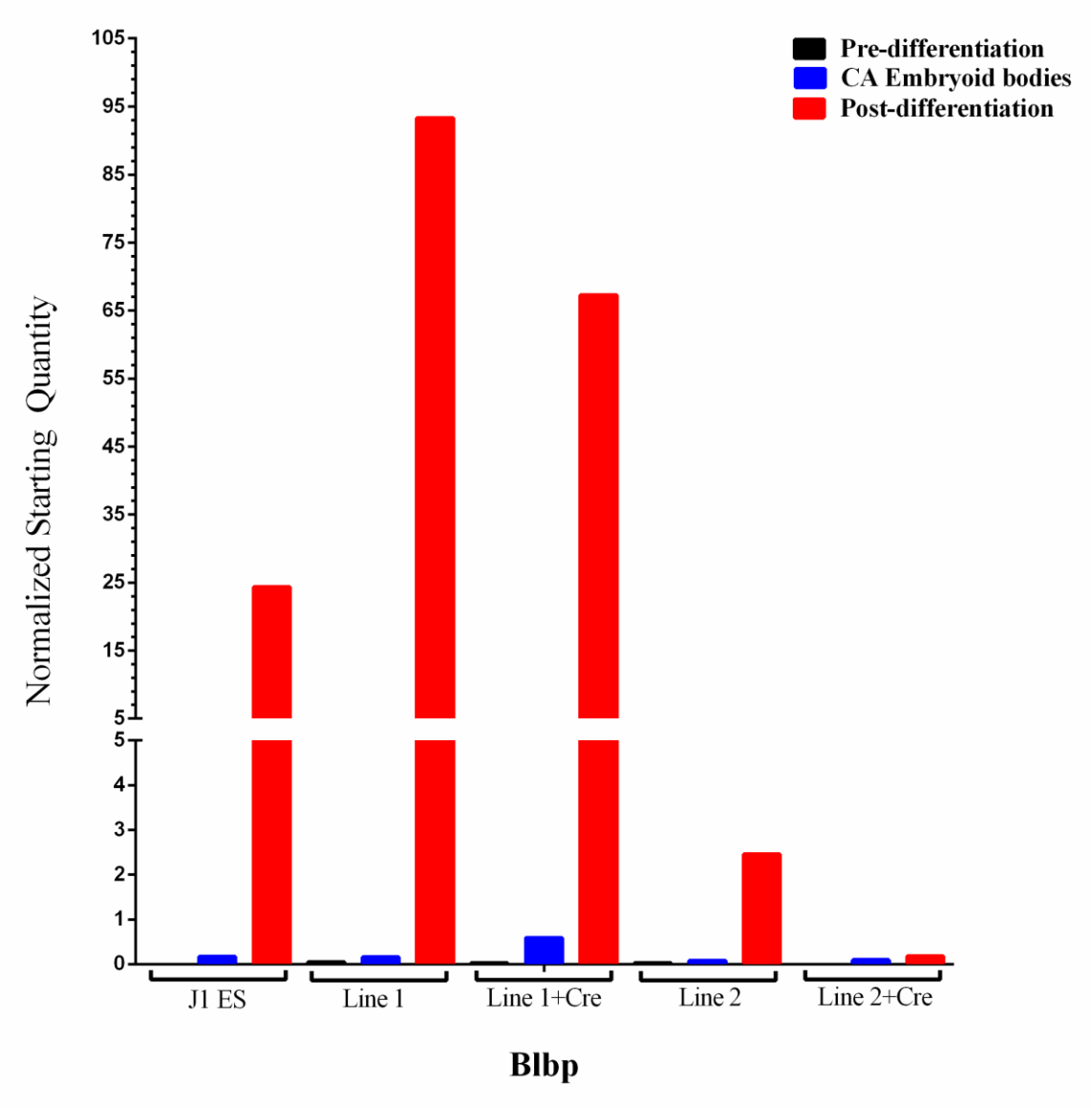


Figure 49. Neuronal cells derived from iPSC line 1 and the active iPSC line 2 are positive for Blbp**.** Neuronal cells derived from iPSC lines without active transgenes are positive for synaptophysin. Quantitative real-time (qRT-PCR) expression levels of Pax6 in iPSCs subjected to neuronal differentiation. RNA was harvested at day 0 (pre-differntiation), CA (embryoid stage) and neuronal stage (post-differentiation). Blbp expression was examined in J1 ES, J1 ES CA, and J1 ES neuron compared to iPSC line 1 and 2 before and after Cre recombinase. J1 ES served as positive control. Starting quantity amount was normalized to E1A binding protein p300. Samples were prepared in duplicates. Line 1 = iPSC line 1, Line 1+Cre = iPSC line 1 after adenoviral cre transfection, Line 2 = iPSC line 2, Line 2+Cre = iPSC line 2 after adenoviral cre transfection, Blbp = brain lipid-binding protein.

CHAPTER IV

CONVERSION OF FIBROBLASTS INTO HEMATOPOIETIC STEM CELLS

Chapter summary

Although ESCs and iPSCs can give rise to any cell type in the body, their usage is still limited in the clinic due to the potential for teratoma formation. A teratoma is a benign tumor encompassing all three germ layers: ectoderm, mesoderm, and endoderm. Additionally, the available methods for generating cells derived from PSC differentiation do not yield enough cells for cell-based therapy. Many studies have derived hematopoietic stem and progenitors cells (HSPCs) from ESCs and iPSCs. The goal of this chapter is to identify an alternative approach for generating HSPCs without using pluripotent stem cells. Although multiple cell types can be used as the platform for cellular reprogramming, here we employed fibroblasts since they share the same mesodermal origin with HSPCs [36, 99, 105, 107, 209], and thus might be more efficiently reprogrammed since both cells likely have similar epigenetic imprinting. Here we overexpressed 16 different reprogramming factors with the goal of identifing a combination of factors that would facilitate trans-differentiation of fibroblasts into HSPCs. These genetic reprogramming candidates were selected based on their importance in mouse models. The results for our initial studies were inconclusive

Introduction

Imagine the ability to regenerate a lost limb or an irreplaceable damaged nerve cell. Stem cells have the potential to replace cells that were once functional but damaged due to diseases. There are two main types of stem cells: embryonic and adult (somatic) stem cells. Embryonic stem cells (ESCs) have not committed to any cell fate and thus can become any specialized cell type. ESCs are considered pluripotent and over the past decade, scientists have generated induced pluripotent stem cells (iPSCs) that share similar phenotypes and gene expression profiles to that of ESCs. Unlike pluripotent stem cells, adult stem cells (somatic stem cells) are considered multipotent because they are committed to a specific cell fate. Somatic stem cells can only replicate or differentiate, into cells within the same compartment. For example, hematopoietic stem cells are considered multipotent because they can only produce more of themselves or cells within blood lineages. This property makes multipotent stem cells an attractive therapy target because the undifferententiated cells cannot give rise to teratomas. Even though ESCs and iPSCs have long been used to generate hematopoietic stem cells, their use in the clinic is hampered by the risk that they might form teratomas [97, 98, 110, 165, 210]. The HSPCs derived from ESC by constitutive HoxB4 overexpression were biased toward myeloid lineage [143, 144, 211-213].

Our goal is to identify ways to bypass the pluripotent state and directly convert fibroblasts into lineage-balanced HSPCs. Not only does this alternative method avoid potential tumor formation, the approach should also cut costs and happen faster as some of the reprogramming steps will be eliminated. This will be particularly useful for critically ill patients who require immediate treatment. Also, trans-differentiating fibroblasts into HSPCs would speed the development of animal models for the study of rare human blood disorders.

Recent efforts have been made to develop cellular reprogramming strategies that bypass iPSC usage. For example, multiple *in vitro* and in *vivo* studies have demonstrated that pancreatic cells can transdifferentiate into hepatocytes [214-216]. In an in vivo study, Cobaleda et al. showed that conditional knockout of Pax5 was enough to transdifferentiate B cells into T cells [217]. Myeloid cells can be converted efficiently into megakaryocytes/erythroid cells by the presence of Gata-1 [218]. Similarly, increased expression of CEBPα/β in pre-B and pre-T cells can promote them to transdifferentiate into macrophages [219, 220]. Overexpression of MyoD in fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells generated myoblasts and myotubes [221]. Fibroblasts can be converted into myoblasts in the presence of small molecule or MyoD1 [221-223]. Overexpression of PU.1 and C/EBPα/β in fibroblasts was able to generate macrophage-like cells [224]. Vierbuchen *et al.* generated functional neurons from fibroblasts by overexpressing Ascl1, and Brn2 (also known as Pou3f2). Similarly, others demonstrated that overexpression of Gata4, Mef2c, and Tbx5 was sufficient to generate functional cardiomyocytes [107]. Recently Pereir et al. showed that overexpression of Gata2, Gf1b, cFos, Etv6 (Tel1) was sufficient to induce endothelial-like precursor cells to become HSPCs like cells [105] Such studies demonstrated that it is possible to redirect one cell type to adopt another cell fate. These studies also suggested that there are master regulators for each specialized cells.

Thus, we proposed that there are specific master regulators for HSPCs can, upon favorable conditions, allow the trans-differentiation of fibroblasts into lineage-balanced HSPCs. To identify a specific gene(s) required for hematopoiesis we selected 16 candidates, including Runx1, Scl, Tel, Myb, Cdx4, Hoxb4, Sox17, Gata2, Bmi1, Lmo2, Wnt3A, Wnt5A, Lhx, Rb, Klf1 and Zfx (listed in Table 1 below). We selected these 16 candidate genes since they are

known to play a role in hematopoietic system and are important for hematopoiesis in mouse models.

Runx1-null mice exhibited abnormal hematopoiesis during development and embryos did not survive past 12.5 days in utero due to hemorrhagic and necrosis of organs [225]. Homozygous deletion of Scl/Tal1 is embryonically lethal in mice, as day 9.5 embryos lacked embryonic (nucleated) red blood cells because of deficiencies in myeloid-erythroid development [226]. Similarly c-Myb is required for T cells development [227], since homozygous-null embryos die around day 14-15, with extreme anemia. Heterozygous mutations in Myb do not lead to fatality but B or T cell development is abnormal. Mice with a homozygote deletion of Cdx4 or other mutations within this gene do not have such strong phenotypes, but have decreased erythropoiesis in yolk sac [228].

Although overexpression of Hoxb4 increases the regeneration of murine HSCs and enhances reconstitution of human severe combined immunodeficiency (SCID) cells, the HoxB4 homozygous-null mouse showed proliferative defects in HSCs [229-231]. Aberrant levels of Hoxb4 caused the hematopoietic system development to fail and led to cancer in mice [102, 232, 233]. Hoxb4 mutants showed significantly less cellularity in the spleen and bone marrow (BM), and only slight decrease in the number of erythrocytes and hemoglobin values. Otherwise, the blood lineage distribution in these mice was normal. Sox17, a Sry-related HMG box factor, is another important factor. Sox17-null mutant mice had completely lacked endoderm [234]. Fetal hematopoiesis required Sox17 for proper development, as germline knockout lead to severe defects in blood product and lack of HSCs [235]. Another important regulator of hematopoiesis is Gata2. This gene is especially important for the maintenance and proliferation of HSCs and multipotent progenitors [116, 236], and its expression is required for fetal myeloid lineage

development. Bmi1, B lymphoma Mo-MLV insertion region 1 homolog, is a polycomb, zinc finger protein that modifies chromatin structures within the cell. Homozygous Bmi1-null mice showed immune deficiencies and severe defects in hematopoiesis since Bmi1 influences the self-renewal potential of HSCs and BM microenvironments [237, 238]. Lmo2, LIM domain only 2, is another important regulator of hematopoiesis as Lmo2-null embryos die at day 10.5 in development due to a lack of yolk sac erythropoiesis [175, 239]. Lmo2 is absolutely required for adult mouse hematopoiesis [240].

Wnt3A and Wnt5A are components of the canonical Wnt signaling pathway. Both proteins have significant roles in hematopoiesis. Wnt3A knockout is embryonically lethal in mice as the embryos die around day 10.5 [241]. Analysis of early fetal livers showed a significant reduction in the numbers of hematopoietic stem and progenitor cells. These cells exhibited severe defects in long-term reconstitution upon transplantation [242]. Wnt3A has a nonredundant role in maintaining HSC functions [243]. Unlike its counterpart, Wnt5A inhibits the canonical Wnt signaling pathway to promote HSC quiescence, which should enhance short-term and long-term hematopoietic reconstitution [244].

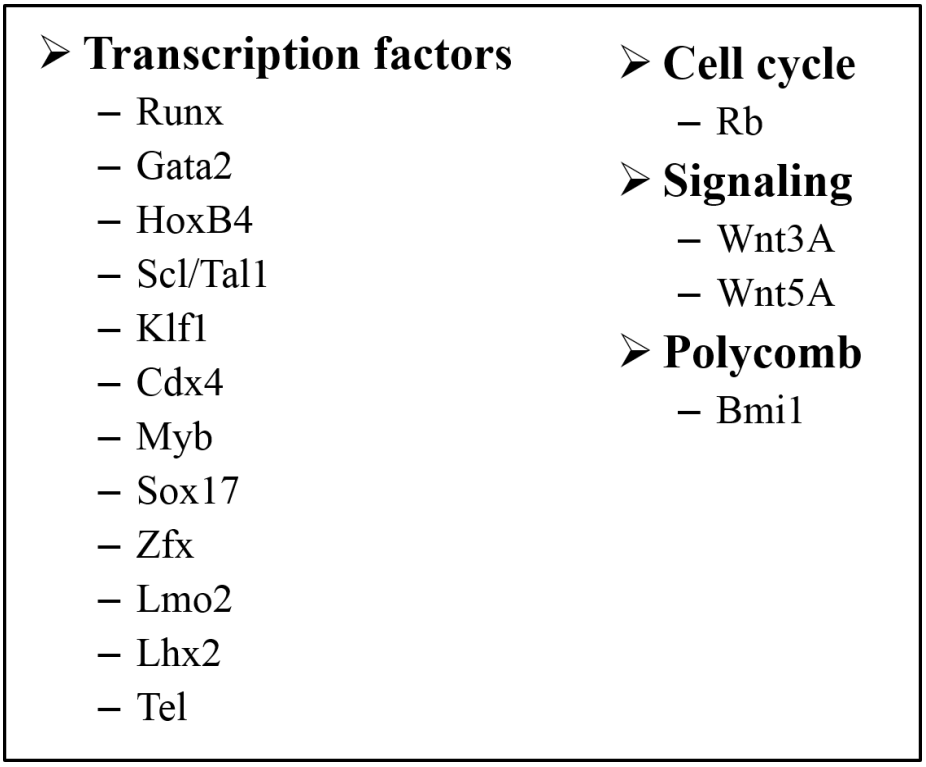
Lhx2, lim domain 2 protein, is expressed in lymphocytes, expressing at the highest level in B lymphocytes [245]. The absence of functional Lhx2 protein causes early mortality in mice fetuses as a result of defects in nervous system development and erythropoiesis [246]. Mice with both Rb1 (retinoblastoma) alleles knockout, die in utero because the liver fails to produce erythrocytes [247]. Another notable factor is the Krupel-like factor 1, also known as erythroid Klf or Klf1, which regulates definitive erythropoiesis [124, 248]. Klf1 expression is highly specific to erythroid cells and regulates the adult β-globin gene. Multiple abnormalities arise in mice embryos lacking Klf1, gene including the reduction of major transcripts in red cell

membrane proteins and heme synthesis enzymes, and defects in development of RBCs. These embryos suffer from severe anemia and fail to develop past day 16 [249-251].

Although the Ets family protein Tel1/Etv6 (translocation Ets leukemia) has redundant roles in angiogenesis and hematopoiesis with another Ets family member Fli1, deletion of Tel1 produced the strongest phenotypes among all Ets family member knockouts. Loss of Tel1 causes embryonic lethality due to lack of erythropoiesis [252]. Tel1 is critical for adult hematopoiesis as Tel1 mutation lead to defects in bone marrow erythropoiesis, myelopoiesis, and lymphopoiesis [253].

Given the important roles in hematopoiesis for our reprogramming factor (RF) candidates, we propose a combination of minimal reprogramming factors (RFs) should enable effective and efficient trans-differentiation of fibroblasts to HSPCs. We will test this idea in drop out experiments in which one RF is omitted at a time to determine whether the presence of any given RF is crucial for HSPC generation. We hypothesize that HoxB4, Gata2, Tel1, Runx, Wnt3A, and Klf1 will be crucial for reprogramming of fibroblasts into blood cells.

Table 1 List of candidate genes used for reprogramming of fibroblasts



Materials and Methods.

Generation of plasmids

The coding sequence for the 16 gene candidates (Hoxb4, Scl/Tal1, Runx1, Gata4, Sox17, Myb, Lmno2, Bmi, and Tel1) were amplified by PCR from cDNA libraries that we generated using genomic DNA isolated from 293T and MEL cells. A combination of traditional and Gateway cloning was used to generate our final lentiviral vectors. In short, the PCR product for each gene was gel purified using a Quiagen kit, Topo cloned, and subsequently cloned into a Gateway entry vector containing attsB sites. The coding sequence of specific genes was shuttled into the final destination vector containing the CMV promoter. To increase transcription efficiency, the coding sequence for mHoxB4 was amplified from of a codon-optimized

commercially available plasmid. Each plasmid candidate was checked by restriction enzyme digestion and, later, sequencing. Please refer to **Table A3** in Appendix for the list of primers used in this study.

Lentiviral production and titer

The same method performed n chapter II was used here to produce the lentiviruses. In brief, rapidly dividing 293T cells were split 1:3 for two consecutative days. On the third day, the cells were seeded at $5.5x10^6$ per 15 cm plate (2 plates per plasmid). These plasmid vectors were used at 20 μg to transfect the 293T cells in trans with 5 μg of envelop plasmid and 15 μg of gag plasmid. The media was replaced at 24 hr post transfection. Viral supernatants were collected at 48 hrs and spun down at 26,000 rpm for 90 min. The viral pellets were resuspended in 80 μg of optimum and mix well on ice for 2 hrs. Approximately, 10 μg aliquots were made for each virus and freezed down at -80^{o}C. To dermine the viral titer, viral RNA was isolated using a Qiagen kit and made into cDNA using the Promega reverse transcriptase kit. The reaction was diluted 40-fold and 5 μg of this dilution was used in qRT-PCR reaction to determine the titer using WPRE primers.

Fibroblast isolation

Timed mating for mice was carried out. Embryos were harvested from day 13.5 timed pregnant mice. Fibroblasts were isolated from these embryos as adapted from a published protocol [254]. In brief, the nervous and digestive systems were removed and the remaining portions were minced and trypsinized to isolate for single fibroblast cells. The cell suspension was plated and passaged one to two times before freezing. This was denoted as passage 2 and 3.

Transduction of fibroblasts

Fibroblasts were thawed and MOIs of 1, 3, and 10 for each of the 16 lentiviruses were used to transduce the cells. After 24 hr, the medium was replaced with fresh fibroblast medium. Between two 48 hr and 72 hr time points, depending on the proliferation state of the cells, actively proliferating transduced fibroblasts were trypsinized with 0.5% trypsin and transferred onto plates freshly coated with OP-9 stromal cells.

OP-9 cell culture

These cells were purchased from ATCC and maintained as recommended by the vendor. In short, OP-9 were grown in a-MEM medium minus riboxyde and nucleoxide, supplemented with 20% FBS, 1% pen/strep, and 1% glutamate. The cells were split 1:5 every 3 to 4 days, to no more than 80% confluence to ensure proper cell function.

Quantitative real-time PCR

RNA was isolated from each sample in 60mm plates. 500ng of RNA was reverse transcribed (Promega kit) and diluted 40-fold. Exactly 5 μl of this dilutions was used for each qRT-PCR reaction using Takara xtaq polymerase.

Results

Selected hematopoietic candidates express in the BM but not in ESCs and MEFs

Our goal was to identify the master regulator(s) of hematopoiesis that will enable efficient conversion of fibroblasts into HSPCs. To ensure a high level of transduction, we used the Trono lentiviral vector and CMV promoter to drive the expression of our transgenes as

mentioned in **Figure 50**. This Trono lentiviral vector contains a LoxP site embbeded within the 3' LTR to allow downstream excision of the transgenes [51]. Additionally, the LTRs contained modifications that inactivated any potential promoter/enhancer activity, to prevent downstream reactivation of the transgenes or activations of neighboring genes. A WPRE sequence is used to enhance mRNA stability.

Upon transduction with the reprogramming factors and incubation in cytokines and O-P9, we anticipated that a minimal genetic amount that will be necessary for the trans-differentiation of fibroblasts into HSPCs. Although the candidates we selected appeared to be necessary for hematopoiesis in mouse models, we wanted to determine the basal levels for these genes in MEF, J1 ES, and BM. Since MEF and J1 ES are not blood cells, we expected low amplification of the various RFs in these two cell types, but the RF levels would be high in BM. Since BM is where the adult hematopoiesis takes place. The qRT-PCR data in **Figure 51** suggested otherwise. Only Klf1, Lmo2, and Scl expressed at high levels in BM compared to MEF and J1 ES; other reprogramming factors (RFs) were expressed at a significantly lower level in BM. HoxB4 and Runx were expressed at similar level in MEFs and BM. Substantial evidence suggests that HoxB4 has an important role in hematopoiesis [102, 143, 211-213, 229, 230, 232]; however, its expression is not limited to the blood compartment and is also found in fibroblasts. Although these results show that our RFs are expressed in multiple cell types, their activities may be tissue-specific.

OP-9 stromal cells upon co-cultivation exhibited significant autofluorescence

We began with 14 reprogramming candidates (RFs), while Lhx2 and Tel1 were added later to our list. We transduced freshly made MEF with the 14 RFs lentiviruses for 48 hrs prior to co-cultivation on OP-9. Since OP-9 and MEF cells are both adherent, we transduced MEFs with a GFP lentivirus to distinguish between the two cell types. Because cytokines are relatively expensive and OP-9 cells are potential source, we wanted to determine whether omission of cytokines would affect the reprogramming process. We tested the following combinations: MEFs transduced with the 14 RFs in addition to cytokines. MEFs transduced with the 14 RFs with cytokines and co-cultured on OP-9, MEFs transduced with the 14 RFs co-cultured on OP-9 with no addition of cytokines, and finally MEFs transduced with only HoxB4, Runx, Klf1, and Rb.

As we observed the cells under the microscope, we found some fibroblasts that were cultured n cytokines with or without OP-9 appeared more condensed and had extensions resembling dendritic cells (blood cells). Thus the cells were stained for the surface marker CD45 which is only expressed in blood cells (**Figure 52**). As expected, unstained MEFs and MEFs transduced with GFP lentivirus only did not exhibit much CD45. There was a slight shift toward the right in the histogram for CD45 positivity in the MEF sample that was transduced with only 4 RFs with cytokines and co-cultivation on OP-9. However, this amount was not very discernable as the peak did not deviate much from the controls. CD45 positivity was dramatically higher in MEFs transduced with the 14 RFs and co-cultivated on OP-9. Cytokines had little effect on this shift, but the CD45 positivity was highest when both cytokines and OP-9 cells were present. This result suggested that reprogramming is enhanced when transduced cells were supplemented with cytokines and OP-9 co-cultivation.

Although the result seemed promising, to detect for MEF polulations that were double positive for GFP and CD45 we gated for GFP and CD45, a significant autofluorescence was observed in the APC channel (CD45). This occurred only in samples that contained OP-9 co-cultivation even though we had included compensation controls when processing the cells. This revelation interfered with the interpretations of earlier findings. When we manually overcompensated the autofluorescence signal, we found that the previously significant CD45 positivity in the reprogrammed samples were not much different from the negative controls. This autofluorescence might be coming from OP-9 as it only appeared whenever OP-9s were present and there was significant bleed-through in the APC channel. To test this hypothesis, we co-cultivated wild-type (WT) MEFs on top of OP-9s for a few days in culture prior to analyzing samples using flow cytometry. We confirmed that OP-9s were the source of the autofluorescence. Several factors could contributed to the lack of CD45 staining. First, the lentiviruses that overexpress the RFs might not be functional. Second, transducing RFs all at once might heighten cytoxicity and the probability of cells harboring all 14 transgenes might also fall. Third, an additional RF might be required in conjunction with our current RFs for cellular reprogramming to take place. We are troubleshooting this area. Meanwhile, the result does not eliminate the possibility that the 14 RFs were not master regulators of hematopoiesis as the autofluorescen could have masked an otherwise positive result.

Effective overexpression of individually transduced reprogramming factors in MEFs

To make sure that all the lentiviruses lentiviruses were independently functional, we transduced each factor individually into MEFs. RNAs were isolated from the transduced MEFs (red bars) to compare via qRT-PCR with the RNA from the mock transduced MEFs (blue bars)

(**Figure 53**). Expression of the transduced MEFs was significantly higher than the mock control (including expression of Wnt3A, Scl/Tal1, Rb, Myb, HoxB4, Lhx2, and Klf1) in Wnt5A, Runx, and Bmi expression increased moderately; however, Zfx, Sox17, Tel, Gata2, Lmo2, and Cdx4 expression remained unchanged (**Figures 53 and 54A**). The lack of gene expression after lentiviral tranductions could be either from defective lentiviruses thus preventing overexpression of their specific reprogramming factor, or alternatively from gene dosage, as some regulated genes can be silenced if significantly overexpressed. To eliminate the first possibility, we produced a fresh batch of lentiviruses for the reprogramming factors that were poorly expressed upon transduction in MEFs. Our qRT-PCR results for this fresh batch of viruses indicated expression was increased, suggesting that defective virueses were the original culprit.

Lack of expression of reprogramming factors and hematopoietic markers after cellular reprogramming

In an attempt to generate HSPCs from fibroblasts, we repeated the pilot experiment and two additional RFs and with fresh lentiviruses. To ensure optimal transduction, we tested two MOIs (10 and 3) with the 16 candidates. MOI 10 appeared to be cytotoxic to cells as detachment from the plates was observed. Subsequently, MEFs were trypsinized and co-cultured on OP-9s with added cytokines. After 7 days of co-cultivation, MEFs were isolated for qRT-PCR to detect changes in RF expression and for upregulation of hematopoietic stem cell markers CD45, c-Kit, Sca-1, and CD34. We chose the day 7 time point as the Bhatia group demonstrated that hematopoietic reprogramming happens quickly and that CD45 expression was observed very early during cellular reprogramming [99]. Thus, we anticipated that at this time point we would be able to detect CD45 expression along with other hematopoietic markers in addition to upregulation of our 16 RFs.

We used mock-transduced MEFs mock and OP-9 as our negative controls. We expected that MEFs transduced with the 16 RFs with or without OP-9 and regardless of the MOI would exhibit a higher total expression level for the RFs. The conditions tested were MEF GFP+16RF MOI3, MEF GFP+16RF MOI10, MEF GFP+10RF MOI3, MEF GFP+16RF MOI3 on OP9, MEF GFP+16RF MOI10 on OP9, and MEF GFP+10RF MOI3 (**Figures 55-59**). Interestingly the results did not turn out as expected. Only Runx1, Klf1, Myb, Lmno2, and Scl/Tal1 showed a significant increase in the 6 treated samples; these levels were drastically higher than those of Mock MEF and OP-9 controls. There were differences in RFs level amongst the treated samples, but these differences were not significantly different from the mock controls. The results suggested that not all of the RFs were efficiently transduced or there were too few cells that possessed all 1 6RFs to account for the differences in gene expressions. Additionally, some cells might not tolerate high MOIs. To overcome these possibilities, we would have to either transduce the cells with fewer reprogramming factors or co-express multiple factors on a single viral backbone. This would reduce the amount of lentiviruses used to transduce cells. Single-cell qRT-PCR analysis could be used to verify transduction consistency throughout each sample.

Finally, to determine if there was any chance that the combination of these 16 RFs could convert fibroblasts into blood-like cells, we examined the expression levels for genes that should be highly expressed in hematopoietic cells such as Sca-1/Ly6A, CD34, CD45, and Vav1 (**Figure 60)**. We anticipated significant upregulation of these genes in the treated conditions in compared to mock controls. However, since Sca-1/Ly6A is promiscuous, its level was naturally high in OP-9 cells as seen in the OP-9 mock control. Interestingly, this level decreased when MEFs and OP-9 were co-cultured together. The levels of Sca-1/Ly6A in the MEFs transduced with 16 RFs that were not co-cultivated with OP-9s were higher than those that were grown with OP-9s.

However, we cannot conclude that reprogramming worked as the levels of Sca-1/Ly6A were not significantly higher than the OP-9 mock levels. Likewise, CD34, CD45, and Vav1 exhibited similar expression patterns as Sca-1/Ly6A. These genes did not have higher expression than the MEF mock sample. The results demonstrated that the cells had not yet trans-differentiated into HSPCs as evidenced by the lack of positivity for surface antigens. Again, we did not have the ability to look at single cell expression but this would be useful as individual cell expression levels might be different than the general population. We anticipated that differences in gene expression might be much more apparent. Overall, this result suggested that HSPCs phenotype has not been observed.after 7 days using these treatment conditions.

Discussion

It is known that overexpression of Oct4, Sox2, Klf4, and cMyc can de-differentiate fibroblasts into pluripotent stem cells [36]. These cells have the potential for "personalized medicine" in which a patient's cells can be used to produce healthy cells. Despite many attractive advantages, the potential for teratoma formation as well as the low efficiency needed for cell therapy have hampered the usefulness of iPSCs. Thus the goal of this study was to find ways to overcome the mentioned barriers. We were interested in generating functional blood cells without going through the pluripotent state.

Similar to our previous study, we employed the Trono lentivirus vector. This vector was modified to minimize gene reactivation and to enhance transgene expression. We chose this strategy to ensure the optimal expression of the exogenous genes. We understand that using viruses is not favorable for clinical studies. Our design here is simply to test the principle that overexpression of hematopoietic genes can convert fibroblasts into HSPCs. After identifying a

specific set of genes that is critical for HSPCs development, alternative techniques for overexpressing these genes can be explored. Even though, lentivirus usage does not trigger formation of malignant tumors in clinical studies, upon identifying the reprogramming candidates that allow fibroblasts conversion into HSPCs we could use Zinc-finger nucleases to allow target the transgenes to specific locations within the genome without disrupting the native state of gene expression or regulation. To determine the abundance of our reprogramming candidates, we used qRT-PCR to check for the mRNA levels in MEFs, ESCs, and BM as shown in **Figure 51**. We anticipated that the expression for these genes would be limited to BM because the BM houses the blood cells. However, this trend was not consistent as the data varied depending on the gene. We found HoxB4 expression to be high in both MEFs and BM; this could be due to the same mesodermal origin between MEFs and blood cells. The Affymetrix array also showed HoxB4 expression in other cell types beside blood cells. HoxB4 might have other functions in addition to promoting proliferation and maintainance of hematopoietic cells.

We found a significant amount of autofluorescence in our first pilot study when MEFs were transduced with 14 RFs and co-cultured on OP-9 stromal cells (**Figure 52**). OP-9 is a stromal cell line isolated from mouse bone marrow, and these cells promote formation of myeloid, erythroid, and B cell lineages from ESCs upon co-cultivation [255]. We had a difficult time processing and gating the samples due to cell population heterogeneity in samples containing MEFs and OP-9s. We expected that as reprogramming took place, the fibroblasts would change shapes and become internally complex (esoniophil with granules) or change surface texture (rough versus smooth). These changes would result in higher side scatters. In samples transduced with 14 RFs in the presence of cytokines and/or OP-9 co-cultivation, we observed a unique population that was not very apparent in other samples; these cells occupied

neither the extreme low nor high on the side scatter axis. Thus, the gating (P2) for these two samples was set slightly differently to account for this unique population. This initial result suggested that our selection of candidates worked and would facilitate cellular reprogramming as evidence by the shift of the peak toward higher decade for CD45 positivity when cells were transduced with all the factors in addition to cytokines and OP-9 co-cultivation. Even though, we used compensation controls, there was a lot of autofluorescence within the APC channel set for CD45. Upon manual compensation to minimize autofluorescence, the right shift for CD45 positivity observed in MEFs with 14 RFs on OP-9s plus cytokines seemed to diminish. This autofluorescence was emitted from OP-9 stromal cells. To avoid autofluorescence in future experiments, we could avoid using APC conjugated antibodies, employ different cell sources, or use lentiviruses to label OP-9s in which we can gate them out from our flow analysis. If we were to use human fibroblasts we could use CD29 to label OP-9s and separate them from human cells which lack CD29. Unfortunately, the reprogramming efficiency of human fibroblasts is significantly lower compared to murine fibroblasts. Although labeling OP-9s prior to starting HSPC experiment will use up one fluoresecnce channel, this is neccessary to remove contaiming OP-9 cells which prevented proper analysis of the data. The method chosen to label OP-9s must allow for continous proliferation of the OP-9 cells because with each cell division the labeling will be diluted out unless we use a permanent integrated marker. Overall, this initial result was inconclusive as to whether any HSPCs was generated.

We found that some of our lentiviruses were not functional as some of the transgenes were not expressed after lentivirus transduction (**Figure 53** and **54A)**. There were significant increases in expression for HoxB4, Scl/Tal1, Rb, Myb, Lhx2 and Klf1 after transduction (red bars) compared to unstransduced MEFs (blue) but Zfx, Sox17, Tel, Cdx4, Gata2, and Lmo2

expression remained unchanged. For the RFs with low levels of expression we made fresh viruses that were effective in individually increasing transgenes expression (**Figure 54**). Thus, the overall result suggested that our RF lentiviruses were functional, and we carried on using these to transduce MEFs.

To determine whether our RFs will facilitate conversion of fibroblasts into HSPCs we transduced the 16 RF lentiviruses into MEFs at MOIs of 10 and 3. Using two different MOIs ensured that one combination would have more integration of the the transgenes with low cytotoxicity. Increased cytotoxicity was observed at a MOI of 10 as many cells lifted off from the plates and took longer to recover. At 72 hours post tranduction, most cells appeared healthy and exhibited normal proliferation although the cells transduced at a MOI of 10 were slightly less confluent. To find the most favorable culture conditions to generate HSPCs, we used several conditions with or without OP-9 co-cultivations. After 7 days with or without cytokines and OP-9, we isolated the cells to check for the expression levels of the RFs using qRT-PCR. This result will also tell us if all 16 RFs were being expressed, suggesting that all of the viruses got into the cells, and if not then the chances were high that not all of the RFs were integrated. We found that not all of the transgenes were expressed as some levels were not higher than the basal level in MEF mock-transduced and OP-9-mock transduced samples (**Figure 55-59**). The trend was inconsistent, but overall there were increases in Runx, Klf1, Myb, Lmon1, Scl/Tal1, and Tel expression while no amplifcation was seen in Zfx, Gata2, Sox17, Wnt2A, Bmi, Rb, and Wnt5A. For Runx we see an increased in its overall total expression in MEF samples that were transduced with 16 RFs of MOI 3, 16 RFs MOI of 10, 10 RFs of MOI 3, 16 RFs of MOI 3 on OP-9, 16 RFs of MOI 10 on OP-9, and 10 RFs of MOI 3 on OP-9. These samples have significantly higher Runx expression compared to the mock transduced MEFs and OP-9 samples.

Interestingly, samples that have 16 RFs at MOI 10 (green and brown bars) in general have higher Runx expression, but are highest when co-cultivated with OP-9 cells (brown bar). We re-tested Runx in the subsequent qRT-PCR run to validate that there was no technical error in qRT-PCR set up or amount of polymerase enzyme added. Klf1, Myb, Lmo2, Scl/Tal1, and Tel shared a similar overexpression trend where the highest increases in overal expression were in the MEFs transduced with 10 RFs at MOI 3 with and without OP-9 cocultivation; although, Tel1 expression was highest when MEFs were transduced with 16 RFs at MOI of 10 and on OP-9s. There were no increases in Zfx, Gata2, Sox17, Wnt3A, Bmi, Rb, and Wnt5A expression as the levels did not surpass the baseline levels of mock-transduced MEF and/or mock-transduced OP-9 controls. These qRT-PCR results suggest that our cells might not be successfully reprogrammed especially when we observed no upregulation of hematopoietic markers such as Ly6A/ Sca-1, CD34, CD45, and Vav1 (**Figure 60**). Thus, the overall result suggested that HSPCs were not generated.

The lack of upregulation in the 16 RF candidates could be due to several reasons. First, all 16 RFs may have been present but the probability of cells having all 16 RFs is low; a cell population sample for qRT-PCR would misrepresent the status of single cells. Thus, we will have to do single cell qRT-PCR to proper assess each sample. We can also examine protein levels in the transduced cells as mRNA levels do not always correlate with protein levels. Furthermore, the MOI of 10 may have enhanced transduction but this also increased cytotoxicity, and may have killed cells which successfully incorporated all 16 RFs. To resolve this issue, we can increase recovery times before futher experimentation, but we must be careful to not let the cells reach senescence as MEFs are prone to this after few passages in culture. We could also make single cell suspensions of transduced MEFs and let them grow in 96-well plates. Over

time, we can select wells that contained clumps of cells compared to the few cells from starting off.

Another potential contributor to the lack of RFs expression could be mechanisms that inhibit certain genes due to imbalance in gene dosage level. Like the iPSC studies done by Yamanaka group, certain combinations of their iPS reprogramming factor candidates enhanced colony numbers while certain other combination had detrimental effects on clonal expansion. This theory could be applied toward our case, and the optimum balance of genes are needed to enable robust trans-differentiation of fibroblasts into HSPCs would have to be discovered. Additonally, the OP-9s we used in this study might not be the optimal cell type for this system; we could try another stromal cell line, S17, that has also been shown to promote robust hematopoiesis [256].

Although our studies haven't yielded conclusive evidence in generating HSPCs from fibroblasts, we established a good starting point to proceed in the future. Of particular importance is for any similar study employing co-cultivation of OP-9s or other cells is to control for autofluorescence in order to have proper analysis of the data.

For future directions, we could repeat transduction of all 16 RFs in MEFs along with dropping each factor individually. The transduced cells can be counted and plated in 96-well plates with about 10 cells/well. If no HSPCs phenotype is observed, we could consider using keratinocytes instead of fibroblasts to increase the reprogramming efficiency. Additional blood reprogramming factor candidates could also be considered. Recent study by Riddel et al. demonstrated that matured differentiated blood cells (pro/preB cells and common myeloid progenitors) can be reprogrommed into hematopoietic stem cells (HSCs) by overexpression of Hlf, Runx, Pbx1, Lmo2, Zfp37, and Prdm5 [257]. Thus, we will consider adding Hlf, Pbx1,

Zfp37, Prdm5, Mycn, and Meis to our reprogramming list. Another option to enhance cellular reprogramming is to couple the transient Oct4 expression to open up chromatin structures which will make the cells more permissible to reprogramming as demonstrated in other studies [44, 166]. If our system is unable to efficiently gnerate HSPCs, we could revert to pluripotent stem cells such as ESCs and try to generate lineage-balanced HSPCs. Hoxb4 had been shown to be critical for hematopoietic development [102, 213, 232, 233], but its constitutive expression led to skewing towards the myeloid lineage over the lymphoid lineage [102, 143, 212]. We can control the activity of Hoxb4 temporally and spatially using the inducible tamoxifen ERT2 system as previously used in other studies [102, 211, 213, 258-260]. Our preliminary data regarding this approach is further described in the Appendix.

Our ultimate goal was to identify a specific genetic cocktail that is critical for converting fibroblasts into lineage-balanced-HSPCs. Success of this work will enhance the field's understanding cellular reprogramming and means to generate HSPCs from other cell types. In addition, this work will also reduce the time, costs, and risks of teratoma formation from using pluripotent cell lines and this will enable faster progress to be made in the clinic.

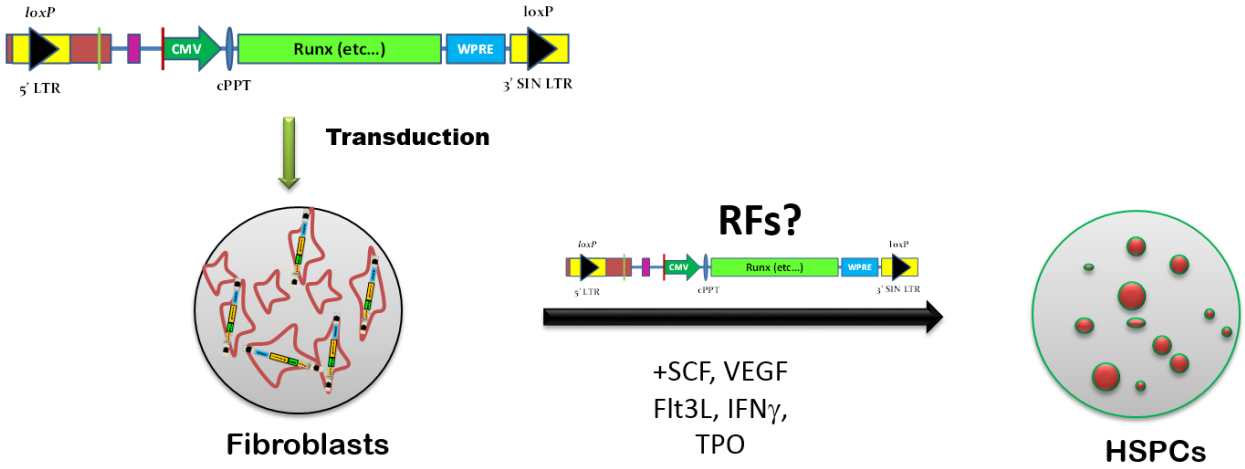


Figure 50. Schematic design for direct reprogramming of fibroblasts into hematopoietic stem and progenitor cells (HSPCs). In this strategy CDS sequence for the different transcription factors and epigenetic regulators is place inside the lentiviral backbone, and the expression is driven by the constitutive promoter CMV. Fibroblasts will be transduced with lentiviruses carrying these reprogramming candidates, and upon incubation in cytokines and co-cultivation on OP-9 cells with HSPCs phenotypes will emerge. Different lentivirus concentration and reprogramming factors (RFs) combination will be used to achieve optimum result.

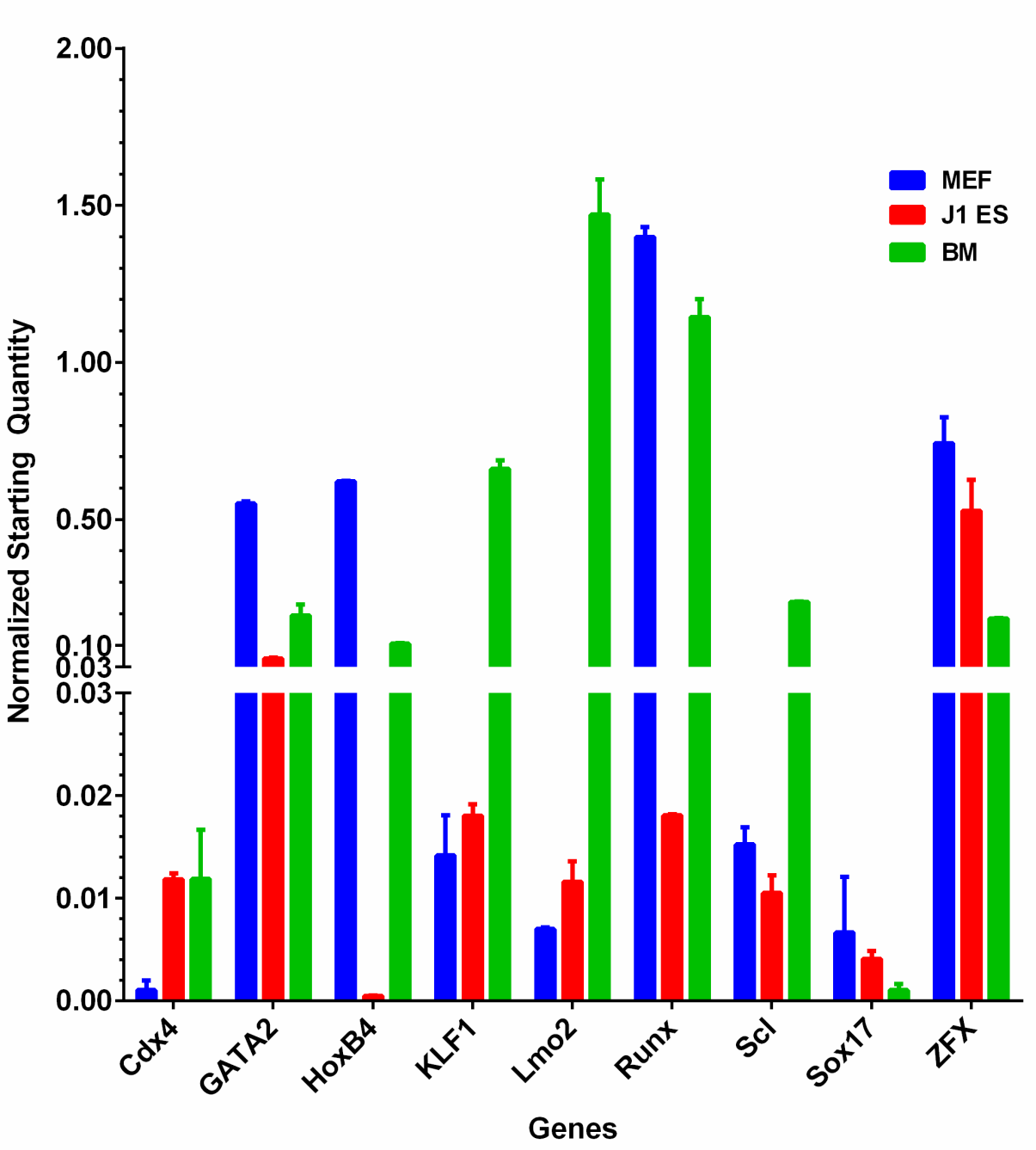


Figure 51. Endogenous expression of direct reprogramming candidates in different cell types. Quantitative real-time PCR (qRT-PCR) expression levels of hematopoietic-related genes in murine embryonic fibroblasts (MEF) – blue bars, embryonic stem (ES) - orange bars cells, and whole bone marrow (BM) – green bars, were analyzed. Levels were normalized to Hprt and Gapdh. Hprt = hypoxanthine-guanine phosphoribosyltransferase, Gapdh = glyceraldehyde-3-phosphate dehydrogenase. Samples were ran in duplicates.

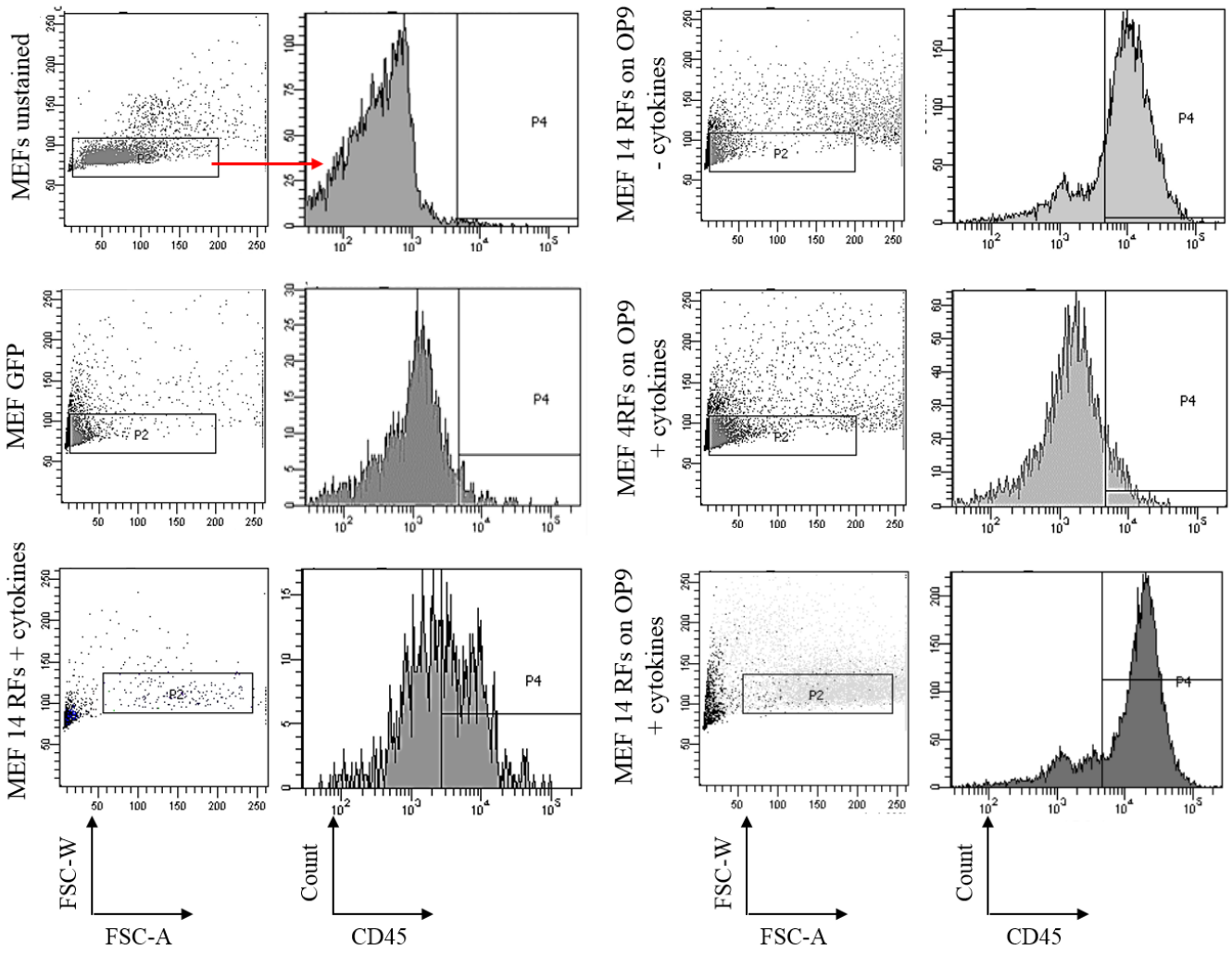


Figure 52. Cell population heterogeneity contributed to false positive CD45 staining**.** Fluorescence activated cell sorting (FACS) shown side scatters (left panel) and histogram with CD45 on X-axis (right panel) for MEFs after reprogramming and stained for CD45. Samples included: MEF mock transduced, MEF transduced with GFP virus only, MEF transduced with GFP virus along with 14 reprogramming factors (RF) in presence of cytokines, MEF transduced with GFP virus along with 14RF co-cultivation on OP-9 without cytokines, MEF transduced with GFP virus along with 4RF on OP-9 and added cytokines, and MEF transduced with GFP virus along with 14 RF in presence of cytokines and OP-9 co-cultivation. RF = reprogramming factors, GFP = green fluorescent protein, MEF = murine embryonic fibroblasts.

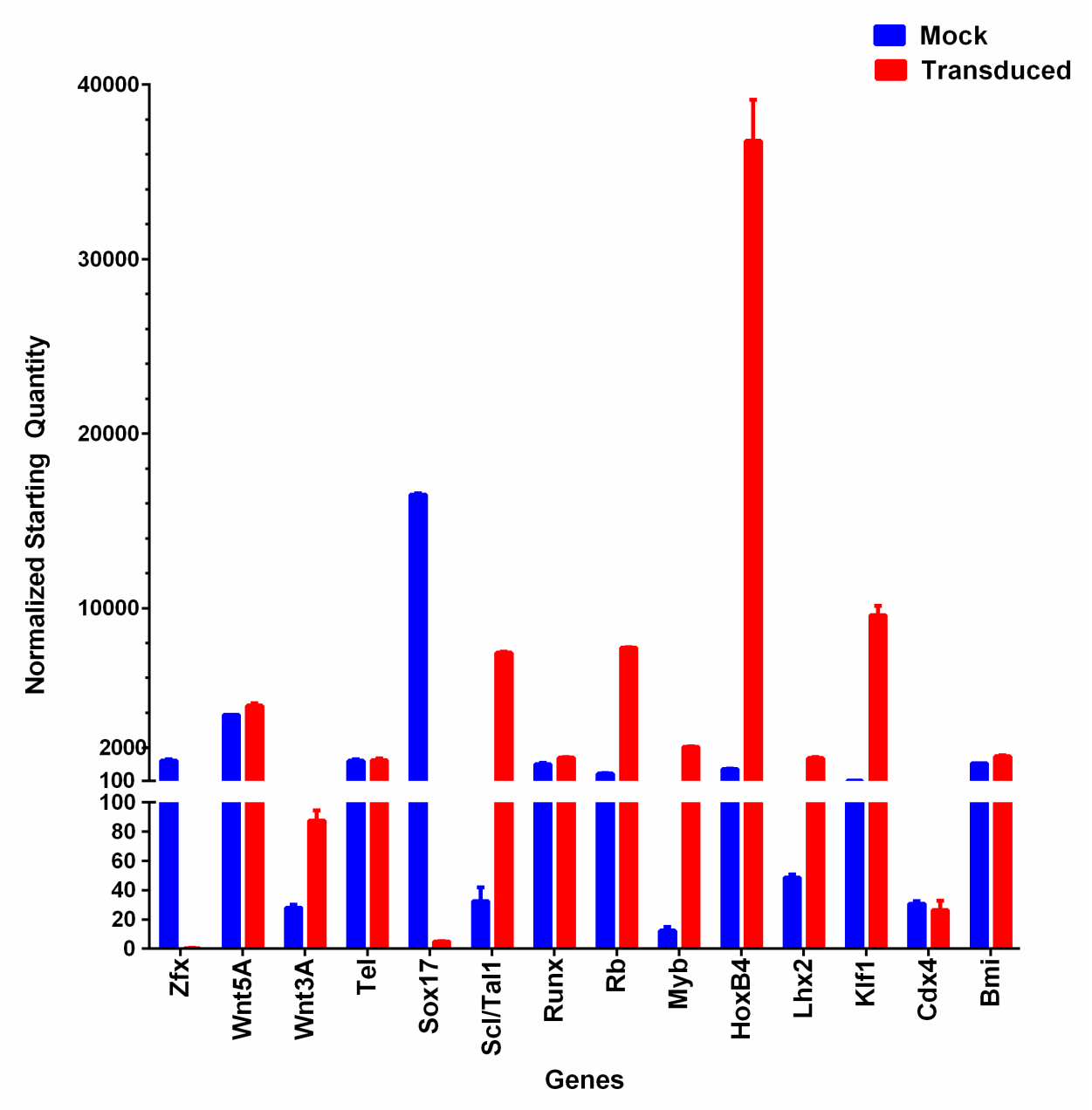


Figure 53. Expression of individual reprogramming factor after transduction in MEF in first batch of virus produced. Quantitative real-time PCR (qRT-PCR) expression levels of the reprogramming factors after 72 hour post-transduction in MEF (red bars) compared to mock transduced MEF (blue bars). Positive change in expression levels for Wnt5A, Wnt3A, Sca/Tal1, Runx, Rb1, Myb, HoxB4, Lhx2, Klf1, and Bmi observed. Negative or no change in expression for Zfx, Tel1, Sox17, and Cdx4 observed. Samples were processed in duplicates. Levels were normalized to E1A binding protein p300.

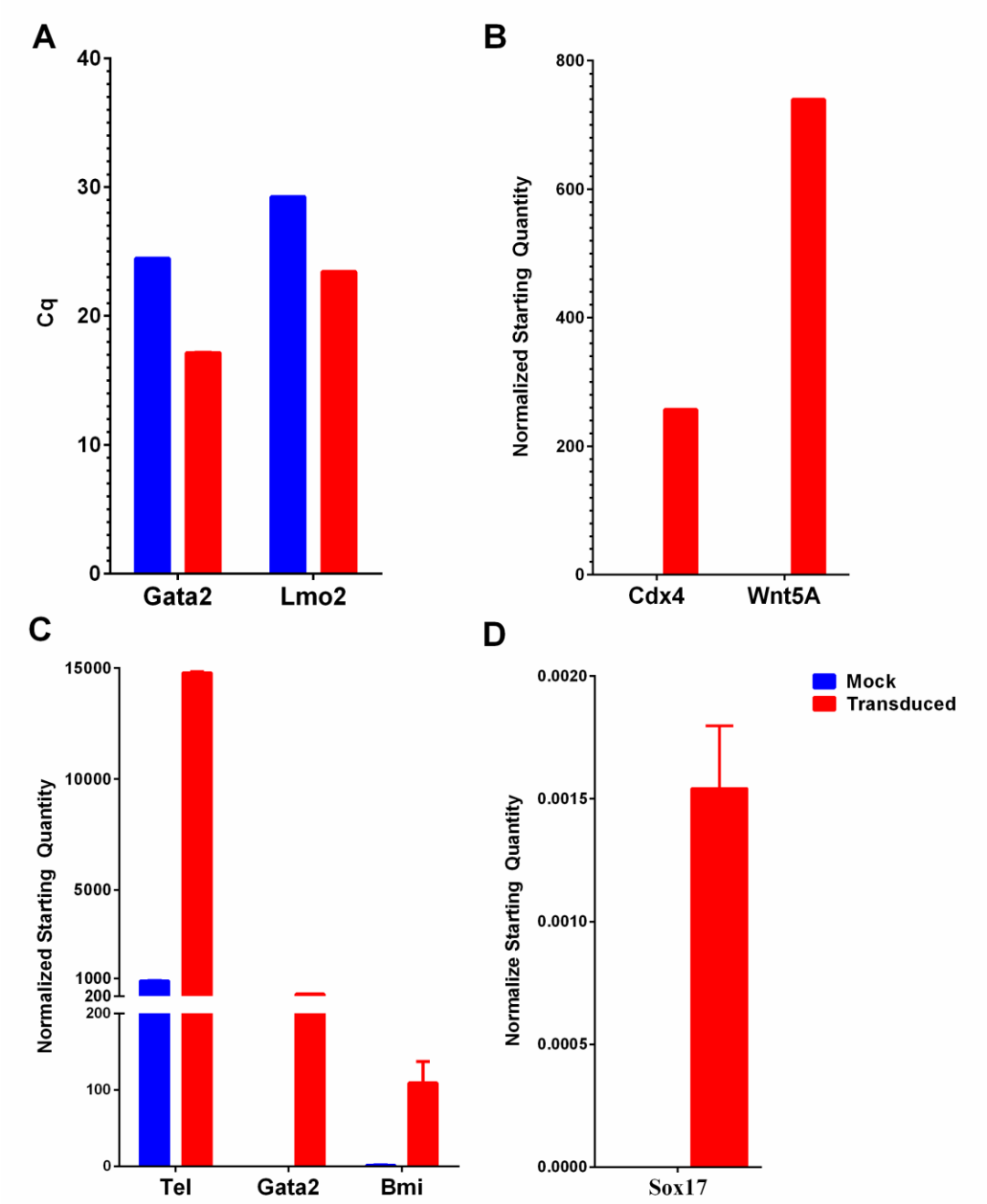


Figure 54. Expression of individual reprogramming factor after transduction in MEF and remake viruses. Quantitative real-time PCR (qRT-PCR) expression levels of the reprogramming factors after 72 hour post-transduction in MEF (red bars) compared to mock transduced MEF (blue bars). (**A**) First batch of virus Gata2 and Lmo2 expression did not enhance after transduction. (B, C, D) Remake of the viruses that previously showed no change in expression in MEFs. The results showed upregulation of Cdx4, Wnt5A, Gata2, Bmi, and Sox17 after remake of new viruses. Levels were normalized to E1A binding protein p300.

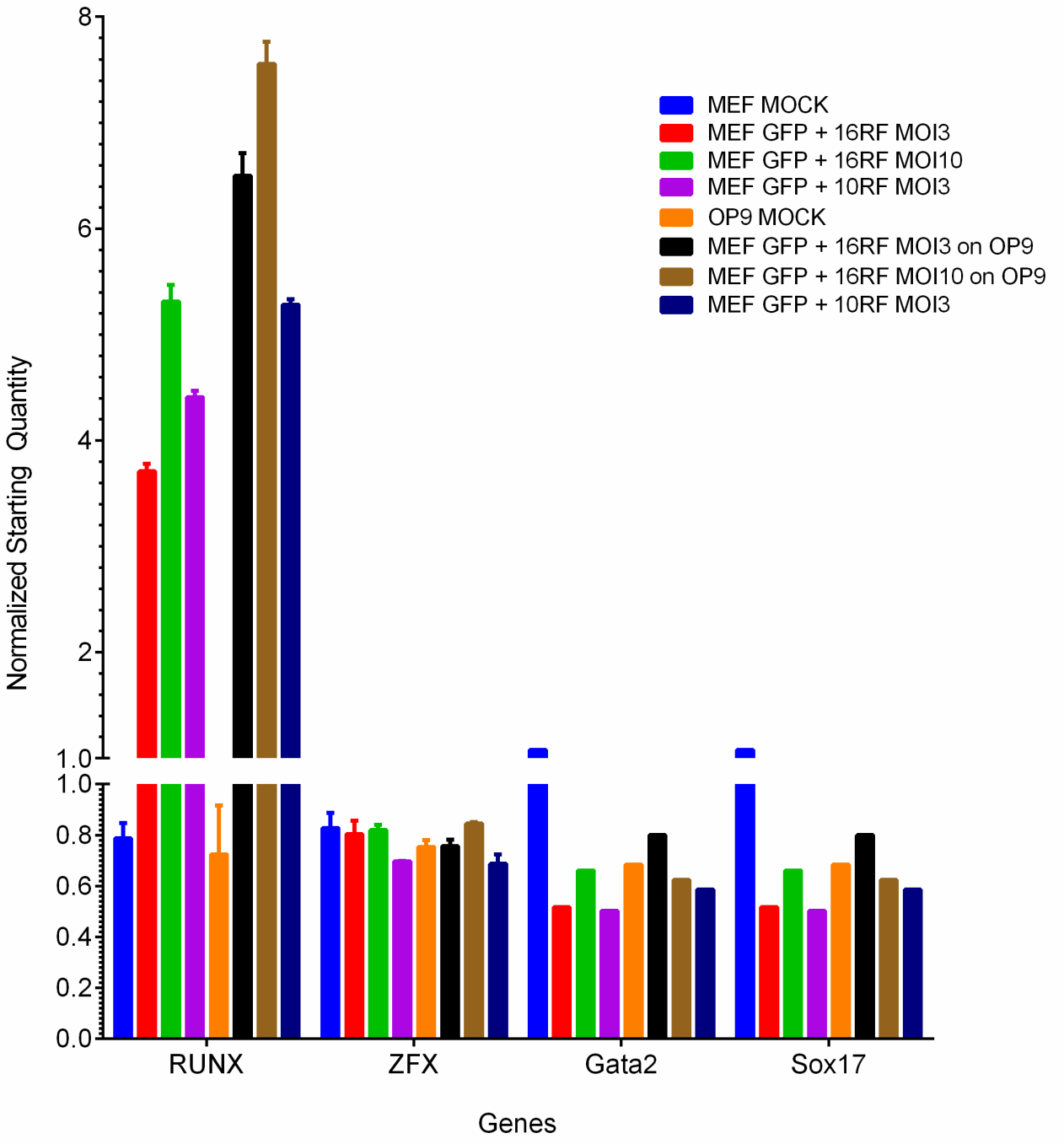

MEF MOCK
MEF GFP + 16RF MOI3
MEF GFP + 16RF MOI10
MEF GFP + 10RF MOI3
OP9 MOCK
MEF GFP + 16RF MOI3 on OP9
MEF GFP + 16RF MOI10 on OP9
MEF GFP + 10RF MOI3
Normalized Starting Quantity
8
6
4
2
1.0
1.0
0.8
0.6
0.4
0.2
0.0
RUNX
ZFX
Gata2
Sox17
Genes

Figure 55. Overall upregulation of Runx1 expression but not for Zfx, Gata2, and Sox17 in MEF/HSPCs after transduction and cultured in hematopoietic medium. Quantitative real-time PCR (qRT-PCR) expression levels of the reprogramming factors after 7 days in culture for: MEF transduced with GFP and 16 RF lentiviruses at MOI of 3 (red bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 10 (green bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 3 (purple bars), MEF transduced with GFP and 16RF lentiviruses at MOI of 3 co-cultivation with OP-9 (black bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 10 co-cultivation with OP-9 (brown bars), and MEF transduced with GFP and 10 RF lentiviruses at MOI of 3 co-cultivation with OP-9 (dark blue bars). Negative controls were MEF mock transduced (light blue bars), and OP-9 mock transduced (orange bars). Samples were run in duplicates. Levels were normalized to E1A binding protein p300. MEF = murine embryonic fibroblasts, OP-9 = bone marrow stromal cell line, MOI = multiplicity of infection, GFP = green fluorescent protein, HSPCs = hematopoietic stem and progenitor cells, RF = reprogramming factors (plural in this case to minimize space needed on graph).

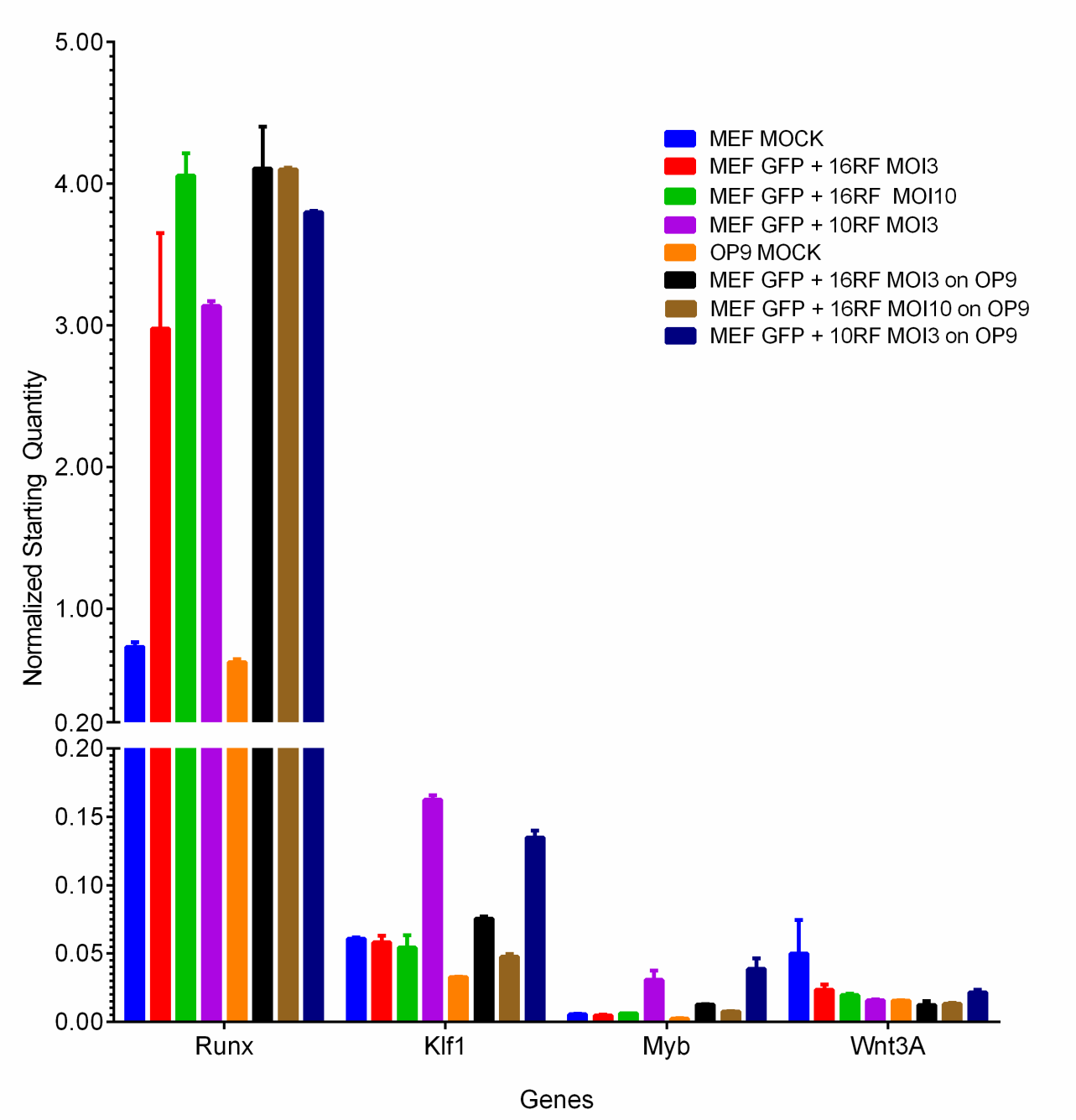

5.00
4.00
3.00
2.00
1.00
0.20
0.20
0.15
0.10
0.05
0.00
Normalized Starting Quantity
MEF MOCK
MEF GFP + 16RF MOI3
MEF GFP + 16RF MOI10
MEF GFP + 10RF MOI3
OP9 MOCK
MEF GFP + 16RF MOI3 on OP9
MEF GFP + 16RF MOI10 on OP9
MEF GFP + 10RF MOI3 on OP9
Runx
Klf1
Myb
Wnt3A
Genes

Figure 56. Overall upregulation of Runx1, Klf1, and Myb expression but not Wnt3A in MEF/HSPCs after transduction and cultured in hematopoietic medium. Quantitative real-time PCR (qRT-PCR) expression levels of the reprogramming factors after 7 days in culture for: MEF transduced with GFP and 16 RF lentiviruses at MOI of 3 (red bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 10 (green bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 3 (purple bars), MEF transduced with GFP and 16RF lentiviruses at MOI of 3 co-cultivation with OP-9 (black bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 10 co-cultivation with OP-9 (brown bars), and MEF transduced with GFP and 10 RF lentiviruses at MOI of 3 co-cultivation with OP-9 (dark blue bars). Negative controls were MEF mock transduced (light blue bars), and OP-9 mock transduced (orange bars). Samples were run in duplicates. Levels were normalized to E1A binding protein p300. MEF = murine embryonic fibroblasts, OP-9 = bone marrow stromal cell line, MOI = multiplicity of infection, GFP = green fluorescent protein, HSPCs = hematopoietic stem and progenitor cells, RF = reprogramming factors (plural in this case to minimize space needed on graph).

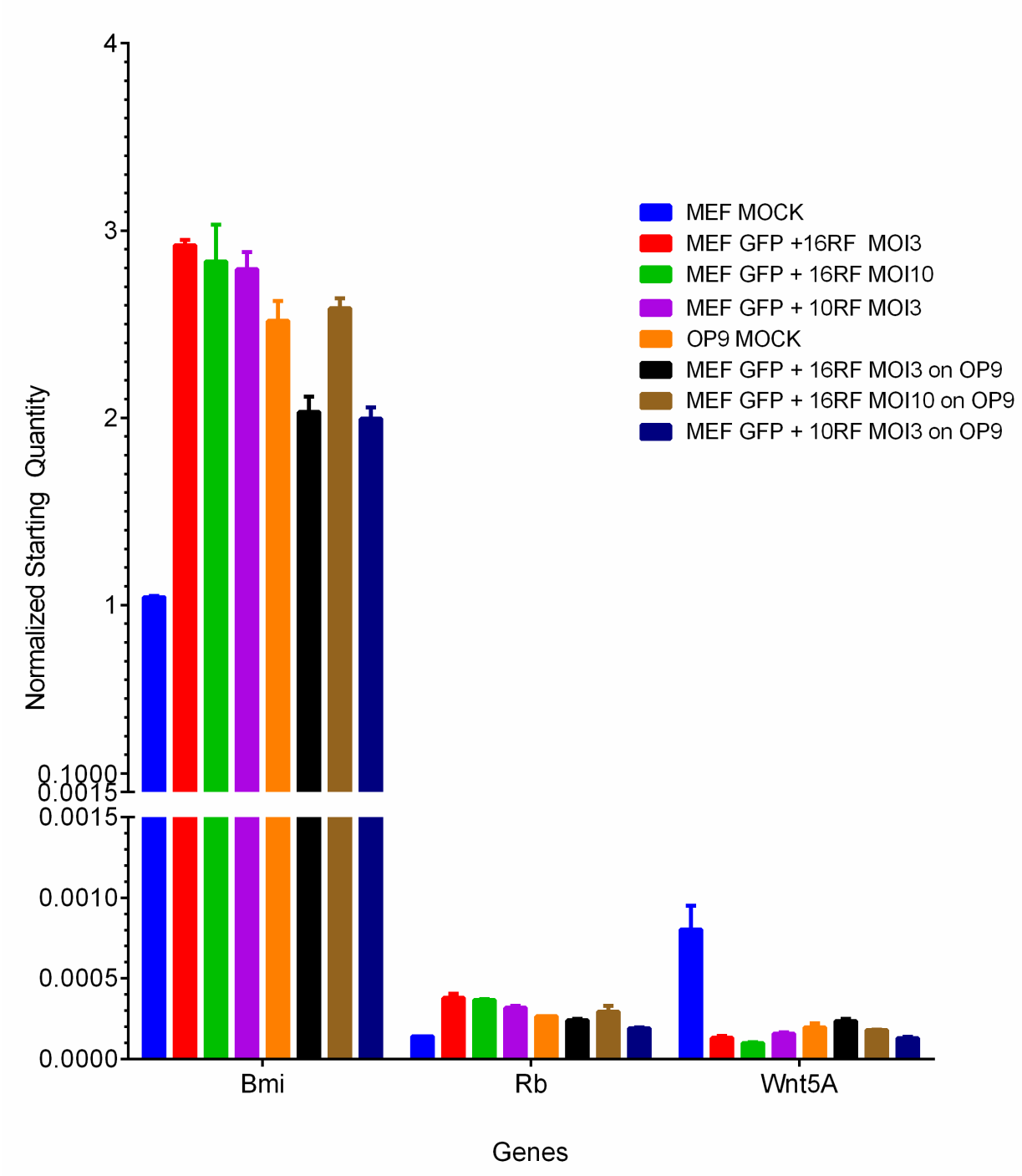

MEF MOCK
MEF GFP +16RF MOI3
MEF GFP + 16RF MOI10
MEF GFP + 10RF MOI3
OP9 MOCK
MEF GFP + 16RF MOI3 on OP9
MEF GFP + 16RF MOI10 on OP9
MEF GFP + 10RF MOI3 on OP9
Normalized Starting Quantity
4
3
2
1
0.1000
0.0015
0.0015
0.0010
0.0005
0.0000
Bmi
Rb
Wnt5A
Genes

Figure 57. No upregulation of Bmi, Rb, and Wnt5A expression was observed in MEF/HSPCs after transduction and cultured in hematopoietic medium. Quantitative real-time PCR (qRT-PCR) expression levels of the reprogramming factors after 7 days in culture for: MEF transduced with GFP and 16 RF lentiviruses at MOI of 3 (red bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 10 (green bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 3 (purple bars), MEF transduced with GFP and 16RF lentiviruses at MOI of 3 co-cultivation with OP-9 (black bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 10 co-cultivation with OP-9 (brown bars), and MEF transduced with GFP and 10 RF lentiviruses at MOI of 3 co-cultivation with OP-9 (dark blue bars). Negative controls were MEF mock transduced (light blue bars), and OP-9 mock transduced (orange bars). Samples were run in duplicates. Levels were normalized to E1A binding protein p300. MEF = murine embryonic fibroblasts, OP-9 = bone marrow stromal cell line, MOI = multiplicity of infection, GFP = green fluorescent protein, HSPCs = hematopoietic stem and progenitor cells, RF = reprogramming factors (plural in this case to minimize space needed on graph).

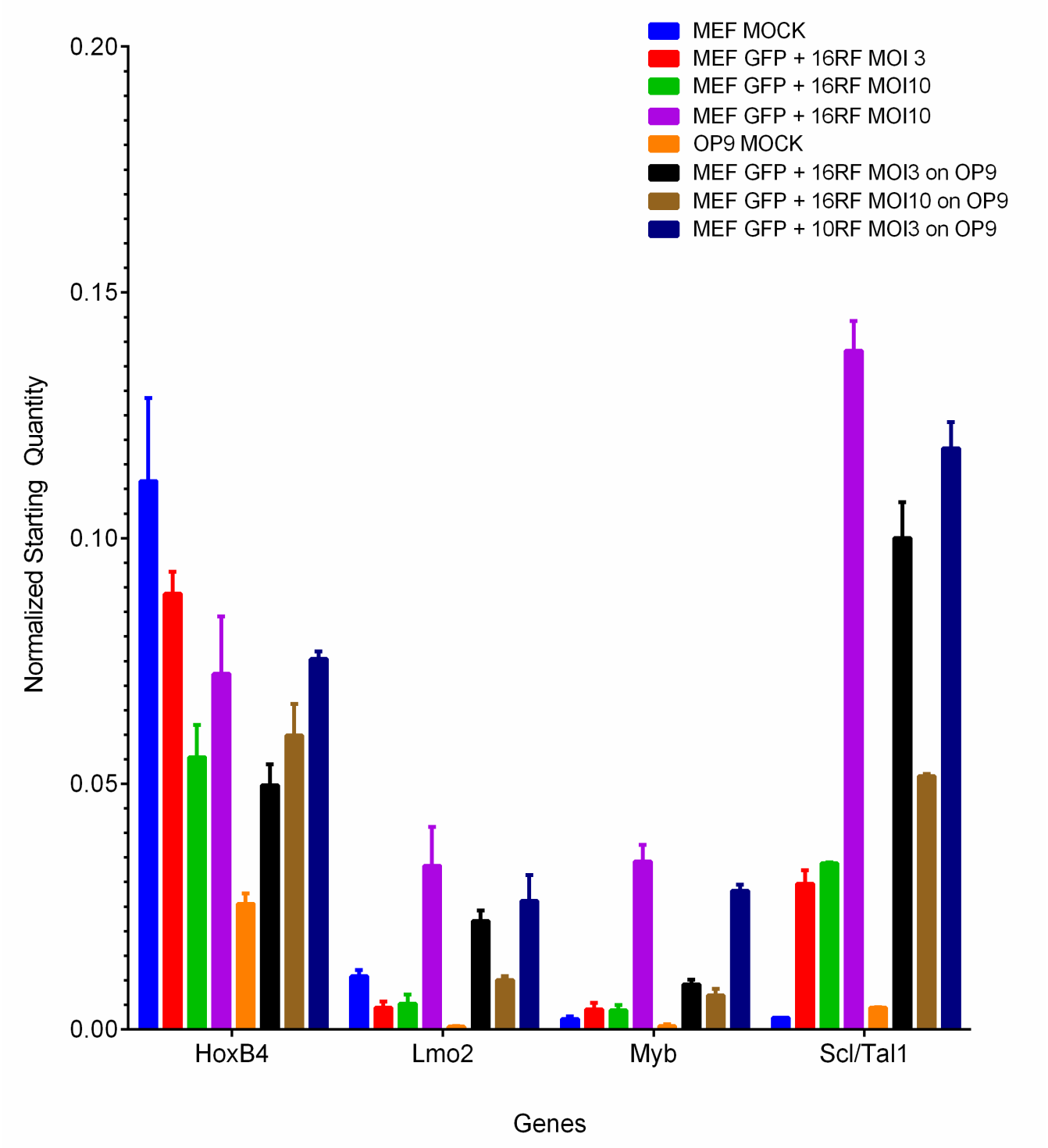

MEF MOCK
MEF GFP + 16RF MOI 3
MEF GFP + 16RF MOI10
MEF GFP + 16RF MOI10
OP9 MOCK
MEF GFP + 16RF MOI3 on OP9
MEF GFP + 16RF MOI10 on OP9
MEF GFP + 10RF MOI3 on OP9
0.20
0.15
0.10
0.05
0.00
Normalized Starting Quantity
HoxB4
Lmo2
Myb
Scl/Tal1
Genes

Figure 58. Overall upregulation of Lmo2, Myb, and Scl/Tal1 expression but not HoxB4 in MEF/HSPCs after transduction and cultured in hematopoietic medium. Quantitative real-time PCR (qRT-PCR) expression levels of the reprogramming factors after 7 days in culture for: MEF transduced with GFP and 16 RF lentiviruses at MOI of 3 (red bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 10 (green bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 3 (purple bars), MEF transduced with GFP and 16RF lentiviruses at MOI of 3 co-cultivation with OP-9 (black bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 10 co-cultivation with OP-9 (brown bars), and MEF transduced with GFP and 10 RF lentiviruses at MOI of 3 co-cultivation with OP-9 (dark blue bars). Negative controls were MEF mock transduced (light blue bars), and OP-9 mock transduced (orange bars). Samples were run in duplicates. Levels were normalized to E1A binding protein p300. MEF = murine embryonic fibroblasts, OP-9 = bone marrow stromal cell line, MOI = multiplicity of infection, GFP = green fluorescent protein, HSPCs = hematopoietic stem and progenitor cells, RF = reprogramming factors (plural in this case to minimize space needed on graph).

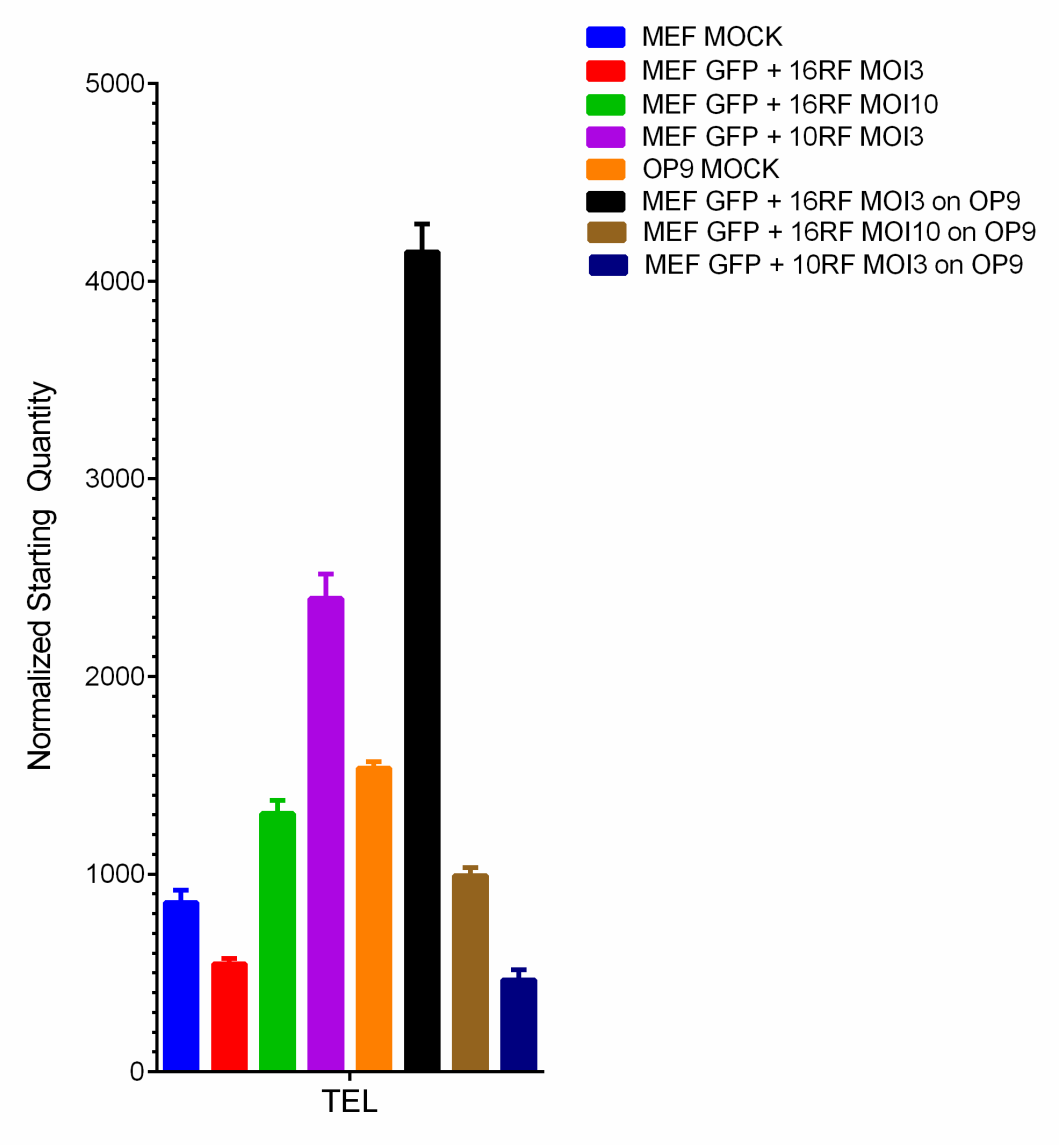


Figure 59. Increased Tel expression in MEF sample at low MOI and contain both cytokines and OP-9**.** Quantitative real-time PCR (qRT-PCR) expression levels of the reprogramming factors after 7 days in culture for: MEF transduced with GFP and 16 RF lentiviruses at MOI of 3 (red bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 10 (green bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 3 (purple bars), MEF transduced with GFP and 16RF lentiviruses at MOI of 3 co-cultivation with OP-9 (black bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 10 co-cultivation with OP-9 (brown bars), and MEF transduced with GFP and 10 RF lentiviruses at MOI of 3 co-cultivation with OP-9 (dark blue bars). Negative controls were MEF mock transduced (light blue bars), and OP-9 mock transduced (orange bars). Samples were run in duplicates. Levels were normalized to E1A binding protein p300. MEF = murine embryonic fibroblasts, OP-9 = bone marrow stromal cell line, MOI = multiplicity of infection, GFP = green fluorescent protein, HSPCs = hematopoietic stem and progenitor cells, RF = reprogramming factors (plural in this case to minimize space needed on graph).

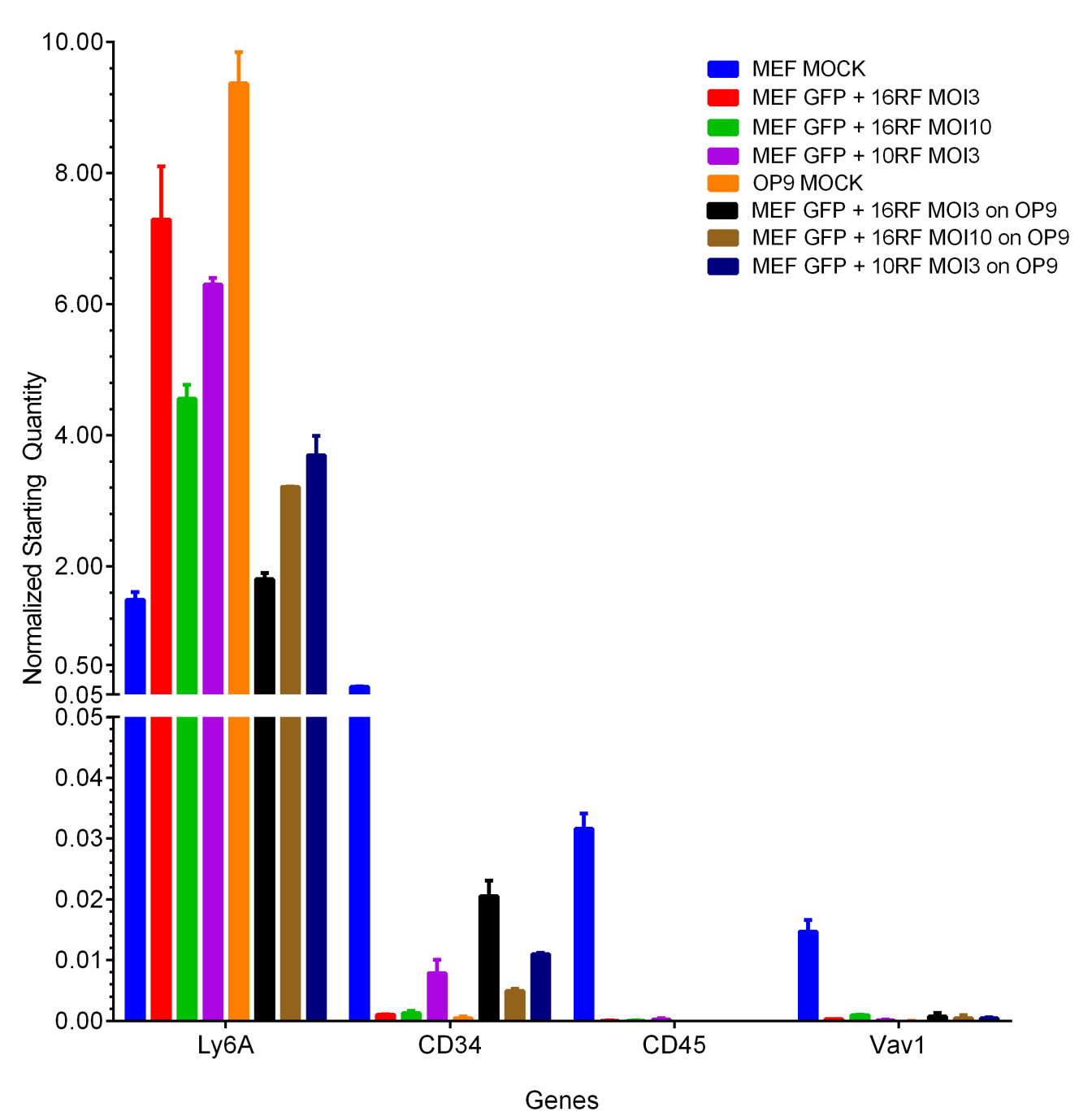

10.00
8.00
6.00
4.00
2.00
0.50
0.05
0.05
0.04
0.03
0.02
0.01
0.00
Normalized Starting Quantity
MEF MOCK
MEF GFP + 16RF MOI3
MEF GFP + 16RF MOI10
MEF GFP + 10RF MOI3
OP9 MOCK
MEF GFP + 16RF MOI3 on OP9
MEF GFP + 16RF MOI10 on OP9
MEF GFP + 10RF MOI3 on OP9
Ly6A
CD34
CD45
Vav1
Genes

Figure 60. Lack of hematopoietic markers expression after MEF/HSPCs reprogramming. Quantitative real-time PCR (qRT-PCR) expression levels of the reprogramming factors after 7 days in culture for: MEF transduced with GFP and 16 RF lentiviruses at MOI of 3 (red bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 10 (green bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 3 (purple bars), MEF transduced with GFP and 16RF lentiviruses at MOI of 3 co-cultivation with OP-9 (black bars), MEF transduced with GFP and 16 RF lentiviruses at MOI of 10 co-cultivation with OP-9 (brown bars), and MEF transduced with GFP and 10 RF lentiviruses at MOI of 3 co-cultivation with OP-9 (dark blue bars). Negative controls were MEF mock transduced (light blue bars), and OP-9 mock transduced (orange bars). Samples were run in duplicates. Levels were normalized to E1A binding protein p300. MEF = murine embryonic fibroblasts, OP-9 = bone marrow stromal cell line, MOI = multiplicity of infection, GFP = green fluorescent protein, HSPCs = hematopoietic stem and progenitor cells, RF = reprogramming factors (plural in this case to minimize space needed on graph).

CHAPTER V

GENERAL DISCUSSION

This chapter provides a summary of cellular reprogramming and the associated studies of specific biological questions this dissertation aims to answer. Since the discovery made by Shinya Yamanaka in 2006 showed that overexpression of just four genes [Oct4, Sox2, Klf4, and cMyc] was able to convert the differentiated mouse fibroblasts to become pluripotent cells, the term induced pluripotent stem cells was used [36]. The subsequent year, they discovered these same four factors can convert human fibroblasts into iPSCs abeit with longer conversion times and exhibited lower efficiency. Nevertheless, this work demonstrated a potential for "personalized medicine." Cell-based therapy treatments can be specifically tailored according to the disease the patient has, providing many great advantages but also harboring some weaknesses. For example, the capability of iPSCs to differentiate into any cell type has immense potential but could be problematic since any iPSCs that did not differentiate can give rise to teratomas upon transplantation. Second, the conversion of fibroblasts to iPSCs, and then from iPSCs to differentiated cells will take multiple time points and multiple changes in the cells' epigenetic memory. If we could find methods to minimize the number of alteration and time required, not only will patients receive the proper treatment earlier but the cost of treament will aslo be reduced as well. Thus the goals of this dissertation are to identify 1) safer methods to distinguish undifferentiated from differentiated iPSCs for the hematopoietic system, and 2) identify candiates that could convert one cell type to another cell type directly. These goals, if realized, will contribute to the advancement of specialized medicine specifically in the area of generating hematopoeitic stem and progenitor cells (HSPCs).

The first goal was to identify an efficient reporter, a tool currently unavailable. By using a reporter in HSPC generated from iPSCs one can increase the safety and efficiency of pluripotent stem cells since the reporter would enable an additional purification method to isolate for only differentiated cells. This would minimize the possibility of teratoma formation upon transplantation the iPSCs derived HSPCs into the host. A reliable hematopoietic reporter would only mark HSPCs but not other cell types; out of all of the candidates we tested, only CD45 met this criteria [190]. As described in Chapter II, CD45 is expressed exclusively in hematopoetic cells, and a small genomic region near the promoter was sufficent to drive high GFP expression [133-137, 145, 190, 261]. For future directions, one could used this CD45 reporter in generating HSPCs from iPSCs. CD45 would enhanced the purification of differentiated cells from undifferentiated iPSCs. The selected iPSCs derived HSPCs can be transplanted into animal models to demonstrate a lack of teratoma formation when CD45 reporter was used as a selectable marker. Use of this CD45 reporter would enhance the purification of differentiated cells from those undifferentiated pluripotent cells, and could reduce the costs and risks of any HSPC-intermediate method.

The second goal of this dissertation was to gain additional knowledge at the molecular and cellular level of iPSCs in order to effectively use these cells to generate hematopoietic stem cells that have direct clinical applications. Although much has been done since the discovery of iPSCs, a complete understanding of this system is still lacking [112, 165, 189, 262]. Our study in Chapter III showed that Oct4 and Sox2 were the two genes expressed at high amounts early on in the reprogramming stage, whereas Klf4 and c-Myc levels were much lower. The silenced iPSC lines generated could give rise to teratomas and differentiate into neurons. An interesting finding was that the Nanog, Essrb, and Lamin B expression was when the transgenes are being expressed

at high levels. No teratomas were observed in the NOD/SCID mice after receiving cells from the active transgenes from the iPSC line. This might be due to a lack of critical pluripotency factors such as Nanog; although, a recent publication suggested that Nanog is not required for iPSCs culture in vivo [208].

Schwartz et al. found that Nanog-deficient iPSC lines are not only transciptionally similar to normal iPSC lines, but can also support the formation of teratomas and chimeric mice. These findings conflicted with established knowledge in the field which suggested that Nanog is essential for iPSCs self-renewal and pluripotency [205-207]. The results presented in Chapter III showed that iPSC lines with active OSMK transgenes failed to form teratomas upon injection in NOD/SCID mice. However, teratomas were observed after the excision of the OSMK transgenes, restoring Nanog to normal levels. The rate of teratoma formation was much faster compared to the silent iPSC lines. However, this post-transgene excision cell line 2 could not differentiate efficiently into neurons. This could be due to several mechanistic possibilities. First, Nanog is crucial for the proper function of iPSCs, thus its aberrant expression could affects iPS differentiation capability. Second, Nanog might not contribute to the differences in differentiation efficiency among these iPSC lines, but rather the abberant level of the OSMK transgenes. High transgene expression could lead to misregulation of downstream targets of OSMK, thus interferring with the proper function of the cell. Finally studies have demonstrated that the efficiency and quality of iPSCs generated varied greatly depending on the various reprogramming techniques. Some of these iPSCs generated differ significantly from ESCs at the molecular level, as a result interferes with iPSCs developmental potential. In our study we found that iPSCs with silenced transgenes shared similar pluripotency genes expressional profile to that of ESCs. These iPSCs were much more efficient in differentiating into neuronal cells compared

to iPSCs with highly active transgenes. Even after the excision of the transgene in the iPSC line with active transgene expression, these cells still failed to differentiate effeciently and effectively into neuronal cells. Base on our result, we suggest that for studies that are generating iPSCs for clinical applications should consider using fluorescence based reprogramming factors. This system would allow researchers to select for and expanded silenced iPSC colonies, instead of the iPSC colonies with highly active OSMK transgenes (high fluorescence levels).

Finally, the third goal of this thesis was to identify a cocktail of genes that is crucial for converting fibroblasts into lineage-balanced HSPCs. Success of this work would enhance the field's understanding of cellular reprogramming. In addition, this work will also reduce time, costs, and the risk of teratoma formation from using pluripotency cell lines in therapy, which will enable faster progress made in the clinic. To carry out this study, we looked at a wide variety of genes with roles in hematopietic development in mouse models and narrowed the list down to 16 reprogramming candidates as mentioned in Chapter IV. To deliver these genes to the cells, we opted for a lentivirus system. Although this system is lentiviral-based, and thus will not be favorable for clinical use, we employed it because it is resistant to gene silencing and ensure stable expression of the transgenes. Also, the lentiviral system is an efficient methods to overexpress genes and lacks cell-cycle integration bias. Upon successful generation of HSPCs from fibroblasts, we could use different methods to express the neccassary gene cocktail without viral-based integration. After overexpression of the 16 reprogramming candidates and coculture on OP-9 stromal cells, there was a lack of amplication of markers relating to hematopoietic cells both by quantitative PCR and flow cytometry. This experiment can be repeated in the future by dividing the transduced cells into single cell suspension and plated on a 96-well plate. Over time, we anticipated that cell that gets the optimum number of transcription factors will proliferate and

trans-differentiate into HSPCs, and as a result we are expected to see a clump of cells in some well. This strategy will allow better assesment of the subpopulations arise from the single starting out fibroblast cell. Additionally, since the fibroblast reprogramming efficiency is low, we can try different cell types instead, keratinocytes for example. Although these cells are of a different germ layer (ectoderm instead of mesoderm) keratinocytes are much more efficient in cellular reprogramming. One additional option to increase the success rate is to couple the transient Oct4 expression, as used by other systems to induce the cells to become more permissive to reprogramming [44, 166]. Additional blood reprogramming factor candidates could also be considered. Recent study by Riddel et al. demonstrated that matured differentiated blood cells (pro/preB cells and common myeloid progenitors) can be reprogrommed into hematopoietic stem cells (HSCs) by overexpression of Hlf, Runx, Pbx1, Lmo2, Zfp37, and Prdm5 [257]. Thus, we will consider adding Hlf, Pbx1, Zfp37, Prdm5, Mycn, and Meis to our reprogramming list.

A promising method to generating a lineage-balanced HSPCs is to control the levels of Hoxb4 in pluripotent stem cells. Hoxb4 has been shown to be critical for hematopoietic development [102, 213, 232, 233]. However, its constitutive expression appears to skewed toward the myeloid lineage [102, 143, 212]. To regulate the expression of Hoxb4 temporally and spatially, we employed the inducible tamoxifen ERT2 system that has been used widely [102, 211, 213, 258-260]. The preliminary results shown in the Appendix establish that co-expressing fluorescence proteins along with the HoxB4 contruct does not affect the translocation of Hoxb4 to the nucleus (There was a band observed in the nuclear fraction after the addition of Tamoxifen). This construct was then used to transduce ESCs to facilitate differentiation among the cells toward a hematopoietic fate. The future goal is to use this contruct to overexpress

HoxB4 in iPS and ESC lines to determine exactly when and how much exposure to Hoxb4 is optimal to generate HSPCs that can give rise to both myeloid and lymphoid cells in equal proportion.

In conclusion, the ultimate goal for this dissertation is to gain additional knowledge about cellular reprogramming and to effectively derive HSPCs from iPSCs or fibroblasts in the hopes that this will aid in the advancement of personalized medicine. Although we realized that cell-based therapy with direct clinical applications is still far way, working toward developing an effective therapy to treat patients with different blood illness will minimize costs, making gene/cell therapy more accessible to Americans, and decrease suffering for those stricken with life threatening diseases that cannot be treated by other methods.

APPENDIX

ELUCIDATING MECHANISMS FOR GENERATING LINEAGE-BALANCED HSPCS BASE ON HOXB4 EXPRESSION

Temporal and spatial regulation of HoxB4 activity

Previous studies suggested that ESC differentiation skewed toward the myeloid lineage [102]. Although this is desirable to reconstitute patients with myeloid related diseases, ideally we would want to reconstitute both myeloid and lymphoid lineages, especially for patients who have undergone chemotherapy. Studies by Jackson et al. and Sorrentino et al. showed controlling temporal expression of Hoxb4 had an effect on lymphoid formation [211, 212]. Short pulses of HoxB4 enhanced lymphoid formation in HSPCs enriched from BM whereas constitutive expression of HoxB4 greatly enhanced number of myeloid colony on methocellulose plate [211, 212]. Thus, our goal is to identify a new method to generate HSPCs balanced in hematopoietic lineages through the use of the combination of HoxB4 expression and small molecules.

We fused the ERT2 domain hs/mmHoxB4 which will allow HoxB4 expression to be regulated temporally and spatially. Overlapping was used to assemble the final construct as shown (**Figure A1**) (Please refer to **Table A4** in Appendix for the list of primers used in this study). The completed plasmid was overexpressed in trans with gag and envelope plasmids in 293T to produce lentiviral particles. We used the concentrated virus to transduce J1 ESCs at MOIs of 10 and 30. A MOI of 30 was most efficient; at this MOI, most of the cells were positive for the mOrange fluorescent protein. Next we isolated mRNA from these cells to run qRT-PCR. This approach will help us determine whether ERT2 is functional; as upon addition of 4-hydroxytamoxifen (4OH-TAM), ERT2 will allow HoxB4 to translocate into the nucleus to

upregulate its downstream target. As shown in **Figure 61**, our negative controls include mock transduced J1 ESCs and J1 ESCs transduced with constitutive HoxB4 with and without 4OH-TAM addition. In the mock J1 ESCs, there were very little to no HoxB4 detected whereas in J1 ES samples that were transduced with regulatable HoxB4 (ERT2) or constitutive HoxB4 we saw an increase in HoxB4. However, there was no increase in Hemgn expression, a downstream HoxB4 target, in any transduced cells. We also looked at the expression levels of additional targets for HoxB4 including Scl, Sox17, Gata2, and Runx to improve our understanding mechanisms of HoxB4 activity. However, there was also no apparent change in the expression for these genes upon overexpression of HoxB4. These findings could be due to multiple reasons. First, the cell system we used in this study might not be appropriate. Previous studies examined the expressional changes in HoxB4 target genes used using marrow lineage depleted cells or ESCs undergoing hematopoietic differentiation [211-213]. Second, HoxB4 alone might not be enough to trigger the downstream signaling cascade; additional stimuli or unmet conditions could be required for HoxB4 function.

Next we wanted to determine whether HoxB4 can translocate into the nucleus in the presence of TAM. We exposed the transduced cells with 4OH-TAM for 3 hrs prior to collecting cell lysates for Western blot analysis. **A2** showed the result for two cell fractions cytosolic and nuclear extract (NE) for J1 ES mock, J1 ES HoxB4-ERT2 without 4OH-TAM, J1 ES HoxB4-ERT2 plus 4OH-TAM at 3hr, and J1 ES TAM at 24hr. Lamin B was used as a loading control. Prior to 4OH-TAM addition to the medium, we expected the majority of HoxB4 protein to reside in the cytosolic fraction. Upon addition of 4OH-TAM, the majority of the HoxB4 protein should translocate to the nucleus. As expected, no HoxB4 was observed in the mock-transduced J1 ESCs. Upon transduction, bands were observed around the molecular weight for HoxB4 in the

cytosolic fraction. No band was observed for HoxB4 in the NE portion before 4OH-TAM addition. After 24 hrs of 4OH-TAM treatment, HoxB4 can be seen in the NE. Although this band is tinted, this sample was underloaded compared to the other samples suggesting that the fusion of ERT2 segment did not interfered with the ability of HoxB4 translocation into the nucleus. Thus the design of this construct preserved the function of HoxB4, and the construct would be useful to determine the exact dosage of HoxB4 that would lead to the generatio of lineage-balanced HSPCs.

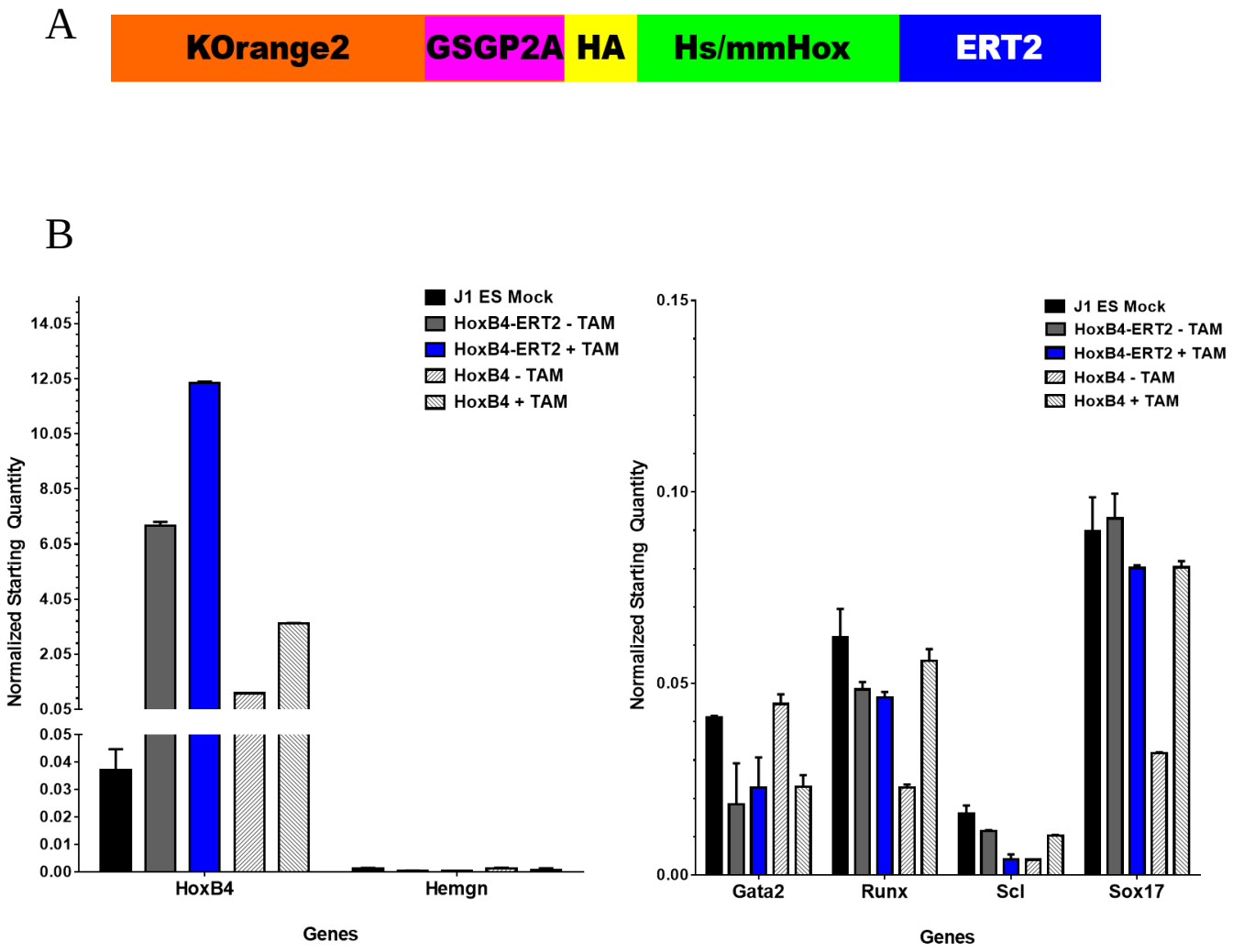


Figure A1. No upregulation of Hemgn in ESCs upon HoxB4 overexpression. (**A**) HoxB4 construct used to transduce ESCs. (**B**) Quantitative real-time polymerase (qRT-PCR) expression levels of hematopoietic genes: HoxB4, Hemgn, Gata2, Runx, Scl, and Sox17. J1 embryonic stem cell (J1 ES) non-transduced (mock), J1 ES transduced with HoxB4-ERT2 construct without addition of 4-hydroxytamoxifen (TAM) to the culture medium (HoxB4-ERT2 – TAM), J1 ES transduced with HoxB4-ERT construct with addition of 4-hydroxytamoxifen (TAM) to the culture medium (HoxB4-ERT2 + TAM), J1 ES transduced with HoxB4 construct without ERT2 fusion in absence of 4-hydroxytamoxifen (TAM) added to the culture medium (HoxB4 – TAM), and J1 ES transduced with HoxB4 without ERT2 fusion construct with addition of 4-hydroxytamoxifen (TAM) to the culture medium (HoxB4 + TAM) were analyzed. Each sample was loaded in duplicate (error bars). Levels were normalized to E1A binding protein p300 (p300).

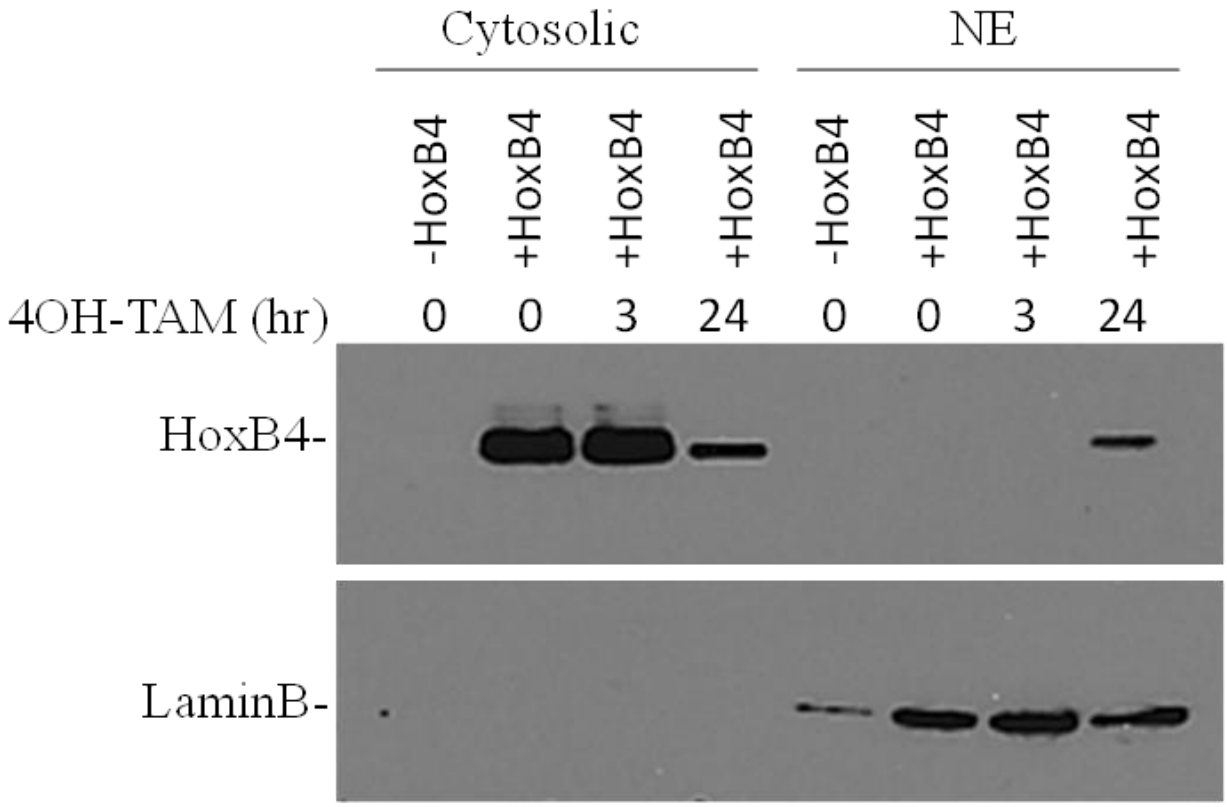


Figure A2. HoxB4 translocated into nucleus upon presence of 4OH-TAM. Western blot of lysate isolated from J1 ESCs transduced with HoxB4-ERT2 lentivirus. Showed protein faction isolated from cytosolic fraction (left) and nuclear extract (right) with 4OH-TAM administered at O, 6, 24, and 48 hr. NE = nuclear extract; hr = hour, TAM = 4-Hydroxytamoxifen.

Primers used in this thesis

Table A1. List of primers used in Chapter II Reporter study (5' → 3')

PCR primers to amplify the reporter constructs (AttB1/AttB2 = red, purple = genomic region, hs = homo sapiens, F = forward, R = reverse)	
CD45 prom -144-AttB1-F1	ggggacaagtttgtacaaaaaagcaggcttacctttagaggaaaattgagacga
CD45 prom -112-AttB1-F1	ggggacaagtttgtacaaaaaagcaggcttaggaagtaagaagccattgcac
CD45 prom -99-AttB1-F1	ggggacaagtttgtacaaaaaagcaggcttacattgcactgactttgaacgac
CD45 prom +4-AttB2-R1	ggggaccactttgtacaagaaagctgggtatgtcagattcggttttactaacttc
CD45 prom +99-AttB2-R1	ggggaccactttgtacaagaaagctgggtactttgcactctgtcttatcagatg
CD45 prom +168-AttB2-R1	ggggaccactttgtacaagaaagctgggtaataatggttatctctcttgaagtttg
CD34 prom -207-AttB1-F1	ggggacaagtttgtacaaaaaagcaggcttaggaattgtgttgtttcctgctt
CD34 prom +128-AttB2-R1	ggggaccactttgtacaagaaagctgggtagagatgcgctcggtgcag
CD34 prom -95-AttB1-F1	ggggacaagtttgtacaaaaaagcaggcttaaatttgaccctgccgagag
CD34 prom -29-AttB1-F1	ggggacaagtttgtacaaaaaagcaggcttagggctggcctcaccaagac
CD34 prom +73-AttB2-R1	ggggaccactttgtacaagaaagctgggtagcaggggtgggggatattg
hsCD45 prom -75-AttB1-F1	ggggacaagtttgtacaaaaaagcaggcttattgaattaaggaagtgcaatgc
hsCD45 prom -67-AttB1-F1	ggggacaagtttgtacaaaaaagcaggcttaaggaagtgcaatgccattc
Quantitative real-time PCR primers	
CD34 qRT-PCR-F	ggtagctctctgcctgatgagt
CD34 qRT-PCR-R	ttggtaggaactgatggggata
CD45 qRT-PCR-F	ggcagatgatattccaaagaaa
CD45 qRT-PCR-R	atctccacttccatgtctccat
Myb qRT-PCR-F	gagatgtgtgaccatgactacga
Myb qRT-PCR-R	gttctgttccaccagcttcttc
Vav1 qRT-PCR-F	gcagtgaactctttgaggcttt
Vav1 qRT-PCR-R	gattcctttgttctgggcaat
Hprt1 qRT-PCR-F	aactttgctttccctggttaa
Hprt1 qRT-PCR-R	tcaagggcatatccaacaaca
Gapdh qRT-PCR-F	tgtgtccgtcgtggatctga
Gapdh qRT-PCR-R	cctgcttcaccaccttcttga
Lentiviral vector titering primers	
LV WPRE-F	accacctgtcagctcctttc
LV WPRE-R	caacaccacggaattgtcagt
LV VSV-G-F	ccattattgcccgtcaagctc
LV VSV-G-R	ggagtgaaggatcggatggact
LV gag-F	atgggtgcgagagcgtcagt
LV gag-R	caggccaggattaactgcga

Table A2. List of primers used in Chapter III iPSCs study (shown 5' → 3')

PCR primers to amplify the mouse OSMK (Oct4, Sox2, c-Myc, and Klf4) constructs (m = mouse, F = forward, R = reverse)_	
EGFP P2A fusion F	caccgtcgaccaccatggtgagcaagggcgagga
EGFP P2A fusion R	ggatcccttgtacagctcgtccatgcc
mKOrange2 P2A fusion F	caccgtcgaccaccatggtgagcgtgatcaagccc
mKOrange2 P2A fusion R	ggatccggagtgggccacggcgtc
mTagBFP P2A fusion F	caccgtcgaccaccatgagcgagctgattaag
mTagBFP P2A fusion R	ggatccattaagcttgtgccc
GSG sequence	ggatccggcgccaccaacttcagcctgctgaagcaggccggcgacgtggaggagaaccccg gcccc
mOct4 P2A fusion F	gctctggatccggcgccaccaacttcagcctgctgaagcaggccggcgacgtggaggagaac cccggccccatggctggacacctggctt
mOct4 P2A fusion R	gctctgcactagttcagtttgaatgcatgggagagcc
mSox2 P2A fusion F	gctctggatccggcgccaccaacttcagcctgctgaagcaggccggcgacgtggaggagaac cccggccccatgtataacatgatggaga
mSox2 P2A fusion R	gctctgcactagttcacatgtgcgacaggggca
mc-Myc P2A fusion F	gctctggatccggcgccaccaacttcagcctgctgaagcaggccggcgacgtggaggagaac cccggcccccctggatttcctttgggcg
mc-Myc P2A fusion R	gctctgcactagtttatgcaccagagtttcgaag
mKLF4 P2A fusion F	gctctggatccggcgccaccaacttcagcctgctgaagcaggccggcgacgtggaggagaac cccggccccatggctgtcagcgacgctc
mKLF4 P2A fusion R	gctctgcactagtttaaaagtgcctcttcatgtgtaagg
Quantitative real-time PCR primers	
Sox2 qRT-PCR-F	aagaacagcccggaccgcgtcaaga
Sox2 qRT-PCR-R	ttgctgatctccgagttgtgcatc
Oct4 qRT-PCR-F	agcacgagtggaaagcaact
Oct4 qRT-PCR-R	ttcatgtcctgggactcctc
Klf4 qRT-PCR-F	gcagtcacaagtcccctctc
Klf4 qRT-PCR-R	agtgtgggtggctgttcttt
cMyc qRT-PCR-F	tagtgctgcatgaggagacac
cMyc qRT-PCR-R	ggtttgcctcttctccacag
Nanog qRT-PCR-F	tgcttacaagggtctgctact
Nanog qRT-PCR-R	aagaggcaggtcttcagagg
Esrrb qRT-PCR-F	atgaaatgcctcaaagtggg
Esrrb qRT-PCR-R	aggttcaggtaggggctgtt
Sox2 qRT-PCR-F	atggcccagcactaccag
Oct4 qRT-PCR-R	tctgttcccgtcactgctct
Klf4 qRT-PCR-F	tcagtgccagaagtgtgaca
WPRE qRT-PCR-F	cacaaactcgaacagcttcg
WPRE qRT-PCR-R	gtaatccagaggttgattatcgg
p300 qRT-PCR-F	gggttcagacgccaaatctt
p300 qRT-PCR-R	tcttcatgccactgttttcg

Table A3. Primers used in Chapter IV Direct reprograming study (5' → 3')

PCR primers to amplify the reporter constructs (AttB1/AttB2 = red, purple = genomic region, hs = homo sapiens, F = forward, R = reverse)	
Scl/Tal CDS Forward	ctgtctagtcgacaatatgccccaggatgacg
Scl/Tal CDS Reverse	ctgtctaactagtagccccagacgcatcacc
Lmo2 CDS Forward	ctgtctagtcgacatgattcgctctctctctttt
Lmo2 CDS Reverse	ctgtctaactagtctagatgatcccattgatct
Zfx CDS Forward	ctgtctagtcgacatggatgaagatggacttga
Zfx CDS Reverse	ctgtctaactagtaagagtactgttagggcagg
Sox17 CDS Forward	ctgtctagtcgactctggagagccatgagcagc
Sox17 CDS Reverese	ctgtctaactagttcaaatgtcggggtagttgc
HoxB4 CDS Forward	cacctcccagaaattaatggctatgag
HoxB4 CDS Reverse	ccttcggtgctcgctgtt
HoxB4 CDS AttB1 Forward	ggggacaagtttgtacaaaaaagcaggcttacaccatggccatgagcagctttc
HoxB4 CDS AttB2 Reverse	ggggaccactttgtacaagaaagctgggtatcacagggcgggagggccgccatt
Gata2 CDS Forward	ctgtctagtcgaccctggttcccaagacacagta
Gata2 CDS Reverse	ctgtctaactagtaccacccttgatgtccatgt
Cdx4 CDS Forward	caccatgtatggaagctgcctttt
Cdx4 CDS Reverse	tgctcctgtttcattcagaa
Aml1/Runx1 CDS Forward	ctgtctactcgagatgcgtatccccgtagatgc
Aml1/Runx1 CDS Reverse	ctgtctaactagtatggcgctcagctcagtagg
Bmi1 CDS Forward	ctgtctagtcgacatgcatcgaacaaccagaat
Bmi1 CDS Reverse	ctgtctaactagtctaaccagatgaagttgctg
mRB CDS Forward	ctgtctagtcgacttatttttgtaacgggagtcg
mRB CDS Reverse	ctgtctaactagtaggtcctcacttttcctcct
mKLF1/EKLF CDS Forward	ctgtctagtcgaccatagcccatgaggcagaag
mKLF1/EKLF CDS Reverse	ctgtctaactagttgtgcaggatcactcagagg
mWnt3A CDS-F2	ctgtctagtcgacatggctcctctcggatacct
mWnt3A CDS-R2	ctgtctaactagtgaaccctgctcccgtgttag
mWnt5A CDS-F1	caccatgaagaagcccattg
mWnt5A CDS-R1	ctatttgcacacgaactgat
Quantitative real-time PCR primers	
Runx1-QRT-F1	ggcagaactgagaaatgctacc
Runx1-QRT-R1	ttgtaaagacggtgatggtcag
SCL-QRT-F1	agcgctgctctatagccttagc
SCL-QRT-R1	ctcttcacccggttgttgtt
Sox17-QRT-F1	ctttatggtgtgggccaaag
Sox17-QRT-R1	gcttctctgccaaggtcaac
ZFX-QRT-F1	caacaacaagcaccaaactcat
ZFX -QRT-R1	tgggtcatcgggaacatagt
Cdx4-QRT-F1	acatcaccatcaggaggaagtc

Cdx4-QRT-R1	gtcactttgcacggaacctc
GATA2-QRT-F1	ccacaagatgaatggacagaac
GATA2-QRT-R1	cataaggtggtggttgtcgtct
Hoxb4-QRT-F1	caaagagcccgtcgtctacc
Hoxb4-QRT-R1	gaaactccttctccaactccag
Lmo2-QRT-F1	tactacaagctgggacggaaat
Lmo2-QRT-R1	actttgtctttcacccgcatc
Wnt5a qRT-F1	cacgctaagggttcctatgaga
Wnt5a qRT-R1	gccagacactccatgacactta
Wnt3a qRT-F1	ccatctttggccctgttctg
Wnt3a qRT-R1	tgtcactgcgaaagctactcc
Rb qRT-F1	ggagtccaaattccaacagaaa
Rb qRT-R1	tcctccttgtttgagacatcct
Lhx2 qRT-F1	actttgccattaaccacaatcc
Lhx2 qRT-R1	gtaaaaggttgcgcctgaact
WPRE Tg qRT-R1	gtaatccagaggttgattatcgg
Lmo2 LV Tg-F1	acgagtggaccaagatcaat
Gata2 LV Tg-F1	cacccttcctccagtctctct
Cdx4 LV Tg-F1	gatacagcaggtcatagtttctg
Bmi1 LV Tg-F1	atgggtcatcagcaacttca
Wnt5a LV Tg-F1	gtggatcagttcgtgtgcaa
Zfx LV Tg-F1	cccggacagttacacagagg
Sox17 LV Tg-F1	agctcagcggtctactattgc
Runx1 LV Tg-F1	caactcgcccaccaacatg
Rb LV Tg-F1	acacgaatgcaaaagcagaga
Klf1 LV Tg-F1	cctctgcccacgtgctttt
Wnt3a LV Tg-F1	gtctatgacgtgcacacctg
Tel LV Tg-F1	gctaggggtggctgtaaagg
Myb tx1/2 LV Tg-F1	aacgcgttctcagctcgaa
Scl-Tal1 LV Tg-F1	cccggacagttacacagagg
Scl-Tal1 LV Tg-R1	tgctggaattcgcccttctg
Hoxb4 LV Tg-F1	aagctgcccaacaccaagat
Hoxb4 LV Tg-R1	cgtttcgatcgctcacaggg

Table A4. Primers used in Regulating HoxB4 activity study

PCR primers used to generate HoxB4-ERT2 construct (OL = overlapping, F = forward, R = reverse)	
OL1F	ggggacaagtttgtacaaaaaagcaggctgatggtgagcgtgatcaagcccga
OL2F	ccgtggcccactccggatccggatccggcgccaccaactt
OL2R	ccgtggcccactccggatccggatccggagccacgaactt
OL3F	cagaaagctgctcatggccatcaaagcgtagtctgggacgt
OL4F	agctctcatgtctccagcagagagcgcgcgggggcctccatt
OL4R	gctctcatgtctccagcagacagggcgggagggccgccat
OL5F	aatggaggcccccgcgcgctctctgctggagacatgagagct
OL6R	ggggaccactttgtacaagaaagctgggttcagactgtggcagggaaaccctc
OL7F	cgatcttcatggtgagcgtgatcaagcc
OL8R	ttcgcatgtcagactgtggcagggaaac
OL9-P2A F	aagttggtggcgccggatccggatccggagtgggccacggcgtc
OL10-HoxB4 F	acgtcccagactacgctttggctatgagttcttttttgatc
OL11	tggaggagaaccccggcccctacccatacgacgtcccaga
OL12	tggaggagaaccccggcccctacccatacgacgtcccaga
OL14P2A-Orange	ggggccggggttctcctccacgtcgccggcctgcttcagcaggctgaagttggtggcgccggatcc ggatccggagtgggccacggc
Quantitative real-time PCR primers	
Hemgn qRT-F	ggacagggaggagcctaaac
Hemgn qRT-R	gagcttccggatcttgatct

REFERENCES

1. Ema, H., et al., *Isolation of murine hematopoietic stem cells and progenitor cells.* Curr Protoc Immunol, 2005. **Chapter 22**: p. Unit 22B 1.

2. Ema, H., et al., *Quantification of self-renewal capacity in single hematopoietic stem cells from normal and Lnk-deficient mice.* Dev Cell, 2005. **8**(6): p. 907-14.

3. Orkin, S.H. and L.I. Zon, *Hematopoiesis: an evolving paradigm for stem cell biology.* Cell, 2008. **132**(4): p. 631-44.

4. Kollman, C., et al., *Donor characteristics as risk factors in recipients after transplantation of bone marrow from unrelated donors: the effect of donor age.* Blood, 2001. **98**(7): p. 2043-51.

5. Mehta, J., et al., *Does younger donor age affect the outcome of reduced-intensity allogeneic hematopoietic stem cell transplantation for hematologic malignancies beneficially?* Bone Marrow Transplant, 2006. **38**(2): p. 95-100.

6. Grosse, S.D., et al., *A public health framework for rare blood disorders.* Am J Prev Med, 2011. **41**(6 Suppl 4): p. S319-23.

7. de Paulis, A., et al., *Basophils infiltrate human gastric mucosa at sites of Helicobacter pylori infection, and exhibit chemotaxis in response to H. pylori-derived peptide Hp(2-20).* J Immunol, 2004. **172**(12): p. 7734-43.

8. Grattan, C.E., et al., *Blood basophil numbers in chronic ordinary urticaria and healthy controls: diurnal variation, influence of loratadine and prednisolone and relationship to disease activity.* Clin Exp Allergy, 2003. **33**(3): p. 337-41.

9. Lucas, A.D. and D.R. Greaves, *Atherosclerosis: role of chemokines and macrophages.* Expert Rev Mol Med, 2001. **3**(25): p. 1-18.

10. Bingle, L., N.J. Brown, and C.E. Lewis, *The role of tumour-associated macrophages in tumour progression: implications for new anticancer therapies.* J Pathol, 2002. **196**(3): p. 254-65.

11. Weisberg, S.P., et al., *Obesity is associated with macrophage accumulation in adipose tissue.* J Clin Invest, 2003. **112**(12): p. 1796-808.

12. Karaiskos, C., et al., *Defective macrophage function in crohn's disease: role of alternatively activated macrophages in inflammation.* Gut, 2011. **60**(Suppl 1): p. A143-A144.

13. Smith, J.E., *Erythrocyte membrane: structure, function, and pathophysiology.* Vet Pathol, 1987. **24**(6): p. 471-6.

14. Bookchin, R.M. and V.L. Lew, *Red cell membrane abnormalities in sickle cell anemia.* Prog Hematol, 1983. **13**: p. 1-23.

15. Sanders, D.B., et al., *Sickle cell disease and complex congenital cardiac surgery: a case report and review of the pathophysiology and perioperative management.* Perfusion, 2014. **29**(2): p. 153-8.

16. Vivier, E., et al., *Functions of natural killer cells.* Nat Immunol, 2008. **9**(5): p. 503-10.

17. Brilot, F., T. Strowig, and C. Munz, *NK cells interactions with dendritic cells shape innate and adaptive immunity.* Front Biosci, 2008. **13**: p. 6443-54.

18. Terunuma, H., et al., *Potential role of NK cells in the induction of immune responses: implications for NK cell-based immunotherapy for cancers and viral infections.* Int Rev Immunol, 2008. **27**(3): p. 93-110.

19. Briggs, R., *An analysis of the inactivation of the frog sperm nucleus by toluidine blue.* J Gen Physiol, 1952. **35**(5): p. 761-80.

20. Briggs, R. and T.J. King, *Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs.* Proc Natl Acad Sci U S A, 1952. **38**(5): p. 455-63.

21. King, T.J. and R. Briggs, *Changes in the Nuclei of Differentiating Gastrula Cells, as Demonstrated by Nuclear Transplantation.* Proc Natl Acad Sci U S A, 1955. **41**(5): p. 321-5.

22. Gurdon, J.B., *The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles.* J Embryol Exp Morphol, 1962. **10**: p. 622-40.

23. Gurdon, J.B., *The transplantation of nuclei between two species of Xenopus.* Dev Biol, 1962. **5**: p. 68-83.

24. Gurdon, J.B., *Adult frogs derived from the nuclei of single somatic cells.* Dev Biol, 1962. **4**: p. 256-73.

25. Gurdon, J.B., *Multiple genetically identical frogs.* J Hered, 1962. **53**: p. 5-9.

26. Gurdon, J.B., *Nuclear transplantation and the analysis of gene activity in early amphibian development.* Adv Exp Med Biol, 1975. **62**: p. 35-44.

27. Gurdon, J.B., R.A. Laskey, and O.R. Reeves, *The developmental capacity of nuclei transplanted from keratinized skin cells of adult frogs.* J Embryol Exp Morphol, 1975. **34**(1): p. 93-112.

28. Wilmut, I., et al., *Viable offspring derived from fetal and adult mammalian cells.* Nature, 1997. **385**(6619): p. 810-3.

29. Tamashiro, K.L., et al., *Cloned mice have an obese phenotype not transmitted to their offspring.* Nat Med, 2002. **8**(3): p. 262-7.

30. Wakayama, T. and R. Yanagimachi, *Cloning of male mice from adult tail-tip cells.* Nat Genet, 1999. **22**(2): p. 127-8.

31. Stevens, L.C. and C.C. Little, *Spontaneous Testicular Teratomas in an Inbred Strain of Mice.* Proc Natl Acad Sci U S A, 1954. **40**(11): p. 1080-7.

32. Kleinsmith, L.J. and G.B. Pierce, Jr., *Multipotentiality of Single Embryonal Carcinoma Cells.* Cancer Res, 1964. **24**: p. 1544-51.

33. Evans, M.J. and M.H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos.* Nature, 1981. **292**(5819): p. 154-6.

34. Martin, G.R., *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.* Proc Natl Acad Sci U S A, 1981. **78**(12): p. 7634-8.

35. Tam, P.P. and J. Rossant, *Mouse embryonic chimeras: tools for studying mammalian development.* Development, 2003. **130**(25): p. 6155-63.

36. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.

37. Boland, M.J., et al., *Adult mice generated from induced pluripotent stem cells.* Nature, 2009. **461**(7260): p. 91-4.

38. Kang, L., et al., *iPS cells can support full-term development of tetraploid blastocyst-complemented embryos.* Cell Stem Cell, 2009. **5**(2): p. 135-8.

39. Zhao, X.Y., et al., *iPS cells produce viable mice through tetraploid complementation.* Nature, 2009. **461**(7260): p. 86-90.

40. Stadtfeld, M., et al., *A reprogrammable mouse strain from gene-targeted embryonic stem cells.* Nat Methods, 2010. **7**(1): p. 53-5.

41. Hester, M.E., et al., *Two factor reprogramming of human neural stem cells into pluripotency.* PLoS One, 2009. **4**(9): p. e7044.

42. Kim, J.B., et al., *Direct reprogramming of human neural stem cells by OCT4.* Nature, 2009. **461**(7264): p. 649-3.

43. Wang, J., et al., *Generation of Induced Pluripotent Stem Cells with High Efficiency from Human Umbilical Cord Blood Mononuclear Cells.* Genomics, Proteomics & Bioinformatics, 2013. **11**(5): p. 304-311.

44. Grinnell, K.L., et al., *De-differentiation of mouse interfollicular keratinocytes by the embryonic transcription factor Oct-4.* J Invest Dermatol, 2007. **127**(2): p. 372-80.

45. Aasen, T., et al., *Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes.* Nat Biotechnol, 2008. **26**(11): p. 1276-84.

46. Sun, N., et al., *Feeder-free derivation of induced pluripotent stem cells from adult human adipose stem cells.* Proc Natl Acad Sci U S A, 2009. **106**(37): p. 15720-5.

47. Fusaki, N., et al., *Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome.* Proc Jpn Acad Ser B Phys Biol Sci, 2009. **85**(8): p. 348-62.

48. Okita, K., T. Ichisaka, and S. Yamanaka, *Generation of germline-competent induced pluripotent stem cells.* Nature, 2007. **448**(7151): p. 313-7.

49. Chang, C.W., et al., *Polycistronic lentiviral vector for "hit and run" reprogramming of adult skin fibroblasts to induced pluripotent stem cells.* Stem Cells, 2009. **27**(5): p. 1042-9.

50. Zufferey, R., et al., *Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo.* Nat Biotechnol, 1997. **15**(9): p. 871-5.

51. Zufferey, R., et al., *Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery.* J Virol, 1998. **72**(12): p. 9873-80.

52. Zufferey, R., et al., *Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors.* J Virol, 1999. **73**(4): p. 2886-92.

53. Woltjen, K., et al., *piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells.* Nature, 2009. **458**(7239): p. 766-70.

54. Desponts, C. and S. Ding, *Using small molecules to improve generation of induced pluripotent stem cells from somatic cells.* Methods Mol Biol, 2010. **636**: p. 207-18.

55. Patel, M. and S. Yang, *Advances in reprogramming somatic cells to induced pluripotent stem cells.* Stem Cell Rev, 2010. **6**(3): p. 367-80.

56. Okita, K., et al., *Generation of mouse induced pluripotent stem cells without viral vectors.* Science, 2008. **322**(5903): p. 949-53.

57. Stadtfeld, M., et al., *Induced pluripotent stem cells generated without viral integration.* Science, 2008. **322**(5903): p. 945-9.

58. Zhou, W. and C.R. Freed, *Adenoviral gene delivery can reprogram human fibroblasts to induced pluripotent stem cells.* Stem Cells, 2009. **27**(11): p. 2667-74.

59. Vorburger, S.A. and K.K. Hunt, *Adenoviral gene therapy.* Oncologist, 2002. **7**(1): p. 46-59.

60. Li, H.O., et al., *A cytoplasmic RNA vector derived from nontransmissible Sendai virus with efficient gene transfer and expression.* J Virol, 2000. **74**(14): p. 6564-9.

61. Seki, T., et al., *Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells.* Cell Stem Cell, 2010. **7**(1): p. 11-4.

62. Ban, H., et al., *Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors.* Proc Natl Acad Sci U S A, 2011. **108**(34): p. 14234-9.

63. Warren, L., et al., *Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA.* Cell Stem Cell, 2010. **7**(5): p. 618-30.

64. Kim, D., et al., *Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins.* Cell Stem Cell, 2009. **4**(6): p. 472-6.

65. Mallanna, S.K. and A. Rizzino, *Emerging roles of microRNAs in the control of embryonic stem cells and the generation of induced pluripotent stem cells.* Dev Biol, 2010. **344**(1): p. 16-25.

66. Chivukula, R.R. and J.T. Mendell, *Abate and switch: miR-145 in stem cell differentiation.* Cell, 2009. **137**(4): p. 606-8.

67. Xu, N., et al., *MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells.* Cell, 2009. **137**(4): p. 647-58.

68. Rosa, A. and A.H. Brivanlou, *A regulatory circuitry comprised of miR-302 and the transcription factors OCT4 and NR2F2 regulates human embryonic stem cell differentiation.* EMBO J, 2011. **30**(2): p. 237-48.

69. Orkin, S.H. and K. Hochedlinger, *Chromatin connections to pluripotency and cellular reprogramming.* Cell, 2011. **145**(6): p. 835-50.

70. Kim, V.N., *Cell cycle micromanagement in embryonic stem cells.* Nat Genet, 2008. **40**(12): p. 1391-2.

71. Kanellopoulou, C., et al., *Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing.* Genes Dev, 2005. **19**(4): p. 489-501.

72. Murchison, E.P., et al., *Characterization of Dicer-deficient murine embryonic stem cells.* Proc Natl Acad Sci U S A, 2005. **102**(34): p. 12135-40.

73. Wang, Y., et al., *DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal.* Nat Genet, 2007. **39**(3): p. 380-5.

74. Barroso-del Jesus, A., G. Lucena-Aguilar, and P. Menendez, *The miR-302-367 cluster as a potential stemness regulator in ESCs.* Cell Cycle, 2009. **8**(3): p. 394-8.

75. Lin, S.L., et al., *MicroRNA miR-302 inhibits the tumorigenecity of human pluripotent stem cells by coordinate suppression of the CDK2 and CDK4/6 cell cycle pathways.* Cancer Res, 2010. **70**(22): p. 9473-82.

76. Wilson, K.D., et al., *MicroRNA profiling of human-induced pluripotent stem cells.* Stem Cells Dev, 2009. **18**(5): p. 749-58.

77. Anokye-Danso, F., et al., *Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency.* Cell Stem Cell, 2011. **8**(4): p. 376-88.

78. Subramanyam, D., et al., *Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells.* Nat Biotechnol, 2011. **29**(5): p. 443-8.

79. Kawamura, T., et al., *Linking the p53 tumour suppressor pathway to somatic cell reprogramming.* Nature, 2009. **460**(7259): p. 1140-4.

80. Utikal, J., et al., *Immortalization eliminates a roadblock during cellular reprogramming into iPS cells.* Nature, 2009. **460**(7259): p. 1145-8.

81. Marion, R.M., et al., *A p53-mediated DNA damage response limits reprogramming to ensure iPSC genomic integrity.* Nature, 2009. **460**(7259): p. 1149-53.

82. Sachdeva, M., et al., *p53 represses c-Myc through induction of the tumor suppressor miR-145.* Proc Natl Acad Sci U S A, 2009. **106**(9): p. 3207-12.

83. Suzuki, H.I., et al., *Modulation of microRNA processing by p53.* Nature, 2009. **460**(7254): p. 529-33.

84. Suh, S.O., et al., *MicroRNA-145 is regulated by DNA methylation and p53 gene mutation in prostate cancer.* Carcinogenesis, 2011. **32**(5): p. 772-8.

85. Judson, R.L., et al., *Embryonic stem cell-specific microRNAs promote induced pluripotency.* Nat Biotechnol, 2009. **27**(5): p. 459-61.

86. Duenas-Gonzalez, A., et al., *Valproic acid as epigenetic cancer drug: preclinical, clinical and transcriptional effects on solid tumors.* Cancer Treat Rev, 2008. **34**(3): p. 206-22.

87. Huangfu, D., et al., *Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds.* Nat Biotechnol, 2008. **26**(7): p. 795-7.

88. Huangfu, D., et al., *Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2.* Nat Biotechnol, 2008. **26**(11): p. 1269-75.

89. Shi, Y., et al., *Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds.* Cell Stem Cell, 2008. **3**(5): p. 568-74.

90. Shi, Y., et al., *A combined chemical and genetic approach for the generation of induced pluripotent stem cells.* Cell Stem Cell, 2008. **2**(6): p. 525-8.

91. Feng, B., et al., *Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells.* Cell Stem Cell, 2009. **4**(4): p. 301-12.

92. Maherali, N. and K. Hochedlinger, *Guidelines and techniques for the generation of induced pluripotent stem cells.* Cell Stem Cell, 2008. **3**(6): p. 595-605.

93. Li, W. and S. Ding, *Small molecules that modulate embryonic stem cell fate and somatic cell reprogramming.* Trends Pharmacol Sci, 2010. **31**(1): p. 36-45.

94. Yamanaka, S., *Strategies and new developments in the generation of patient-specific pluripotent stem cells.* Cell Stem Cell, 2007. **1**(1): p. 39-49.

95. Yamanaka, S. and K. Takahashi, *[Induction of pluripotent stem cells from mouse fibroblast cultures].* Tanpakushitsu Kakusan Koso, 2006. **51**(15): p. 2346-51.

96. Hanna, J., et al., *Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(20): p. 9222-7.

97. Blum, B. and N. Benvenisty, *The tumorigenicity of human embryonic stem cells.* Adv Cancer Res, 2008. **100**: p. 133-58.

98. Gutierrez-Aranda, I., et al., *Human induced pluripotent stem cells develop teratoma more efficiently and faster than human embryonic stem cells regardless the site of injection.* Stem Cells, 2010. **28**(9): p. 1568-70.

99. Szabo, E., et al., *Direct conversion of human fibroblasts to multilineage blood progenitors.* Nature, 2010. **468**(7323): p. 521-6.

100. Shizuru, J.A., R.S. Negrin, and I.L. Weissman, *Hematopoietic stem and progenitor cells: clinical and preclinical regeneration of the hematolymphoid system.* Annual review of medicine, 2005. **56**: p. 509-38.

101. Kaufman, D.S., et al., *Hematopoietic colony-forming cells derived from human embryonic stem cells.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10716-21.

102. Kyba, M., R.C. Perlingeiro, and G.Q. Daley, *HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors.* Cell, 2002. **109**(1): p. 29-37.

103. Hanna, J., et al., *Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin.* Science, 2007. **318**(5858): p. 1920-3.

104. Ran, D., et al., *RUNX1a enhances hematopoietic lineage commitment from human embryonic stem cells and inducible pluripotent stem cells.* Blood, 2013.

105. Pereira, C.F., et al., *Induction of a hemogenic program in mouse fibroblasts.* Cell Stem Cell, 2013. **13**(2): p. 205-18.

106. Vierbuchen, T., et al., *Direct conversion of fibroblasts to functional neurons by defined factors.* Nature, 2010. **463**(7284): p. 1035-41.

107. Ieda, M., et al., *Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors.* Cell, 2010. **142**(3): p. 375-86.

108. Robinton, D.A. and G.Q. Daley, *The promise of induced pluripotent stem cells in research and therapy.* Nature, 2012. **481**(7381): p. 295-305.

109. Smith, Z.D., et al., *Dynamic single-cell imaging of direct reprogramming reveals an early specifying event.* Nat Biotechnol, 2010. **28**(5): p. 521-6.

110. Chan, E.M., et al., *Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells.* Nat Biotechnol, 2009. **27**(11): p. 1033-7.

111. Samavarchi-Tehrani, P., et al., *Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming.* Cell Stem Cell, 2010. **7**(1): p. 64-77.

112. Stadtfeld, M., et al., *Defining molecular cornerstones during fibroblast to iPSC reprogramming in mouse.* Cell Stem Cell, 2008. **2**(3): p. 230-40.

113. Hou, P., et al., *Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds.* Science, 2013. **341**(6146): p. 651-4.

114. Levasseur, D.N., et al., *Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells.* Genes Dev, 2008. **22**(5): p. 575-80.

115. Rossi, L., et al., *Less is more: unveiling the functional core of hematopoietic stem cells through knockout mice.* Cell Stem Cell, 2012. **11**(3): p. 302-17.

116. Suzuki, N., et al., *Combinatorial Gata2 and Sca1 expression defines hematopoietic stem cells in the bone marrow niche.* Proc Natl Acad Sci U S A, 2006. **103**(7): p. 2202-7.

117. Hills, D., et al., *Hoxb4-YFP reporter mouse model: a novel tool for tracking HSC development and studying the role of Hoxb4 in hematopoiesis.* Blood, 2011. **117**(13): p. 3521-8.

118. Kataoka, K., et al., *Evi1 is essential for hematopoietic stem cell self-renewal, and its expression marks hematopoietic cells with long-term multilineage repopulating activity.* J Exp Med, 2011. **208**(12): p. 2403-16.

119. Zhou, Y., et al., *Rescue of the embryonic lethal hematopoietic defect reveals a critical role for GATA-2 in urogenital development.* EMBO J, 1998. **17**(22): p. 6689-700.

120. Su, A.I., et al., *A gene atlas of the mouse and human protein-encoding transcriptomes.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(16): p. 6062-7.

121. Ma, X., et al., *The Ly-6A (Sca-1) GFP transgene is expressed in all adult mouse hematopoietic stem cells.* Stem Cells, 2002. **20**(6): p. 514-21.

122. Levasseur, D.N., et al., *Correction of a mouse model of sickle cell disease: lentiviral/antisickling beta-globin gene transduction of unmobilized, purified hematopoietic stem cells.* Blood, 2003. **102**(13): p. 4312-9.

123. Stella, C.C., et al., *CD34-positive cells: biology and clinical relevance.* Haematologica, 1995. **80**(4): p. 367-87.

124. Blanchette, M., et al., *Aligning multiple genomic sequences with the threaded blockset aligner.* Genome Res, 2004. **14**(4): p. 708-15.

125. Sutherland, D.R. and A. Keating, *The CD34 antigen: structure, biology, and potential clinical applications.* J Hematother, 1992. **1**(2): p. 115-29.

126. Andrews, R.G., J.W. Singer, and I.D. Bernstein, *Precursors of colony-forming cells in humans can be distinguished from colony-forming cells by expression of the CD33 and CD34 antigens and light scatter properties.* J Exp Med, 1989. **169**(5): p. 1721-31.

127. Brunstein, C.G. and D.J. Weisdorf, *Future of cord blood for oncology uses.* Bone Marrow Transplant, 2009. **44**(10): p. 699-707.

128. Bregni, M., R.W. Childs, and N.T. Ueno, *4th Edition of the International Autologous and Allogeneic Cell Therapy for Solid Tumors (ATST) Meeting.* J Cancer, 2011. **2**: p. 307-8.

129. Delaney, C., M.Z. Ratajczak, and M.J. Laughlin, *Strategies to enhance umbilical cord blood stem cell engraftment in adult patients.* Expert Rev Hematol, 2010. **3**(3): p. 273-83.

130. Berenson, R.J., et al., *Antigen CD34+ marrow cells engraft lethally irradiated baboons.* J Clin Invest, 1988. **81**(3): p. 951-5.

131. Berenson, R.J., et al., *Engraftment after infusion of CD34+ marrow cells in patients with breast cancer or neuroblastoma.* Blood, 1991. **77**(8): p. 1717-22.

132. Krause, D.S., et al., *Characterization of murine CD34, a marker for hematopoietic progenitor and stem cells.* Blood, 1994. **84**(3): p. 691-701.

133. Hermiston, M.L., Z. Xu, and A. Weiss, *CD45: a critical regulator of signaling thresholds in immune cells.* Annual review of immunology, 2003. **21**: p. 107-37.

134. Timon, M. and P.C. Beverley, *Structural and functional analysis of the human CD45 gene (PTPRC) upstream region: evidence for a functional promoter within the first intron of the gene.* Immunology, 2001. **102**(2): p. 180-9.

135. Anson, D.S. and T. Occhiodoro, *Transcriptional activity of the CD45 gene promoter in retroviral vector constructs.* Biochim Biophys Acta, 1994. **1219**(1): p. 81-8.

136. Virts, E.L. and W.C. Raschke, *The role of intron sequences in high level expression from CD45 cDNA constructs.* J Biol Chem, 2001. **276**(23): p. 19913-20.

137. Yang, J., et al., *Transgenic tools for analysis of the haematopoietic system: knock-in CD45 reporter and deletor mice.* J Immunol Methods, 2008. **337**(2): p. 81-7.

138. Wang, J., et al., *A protein interaction network for pluripotency of embryonic stem cells.* Nature, 2006. **444**(7117): p. 364-8.

139. Lattin, J.E., et al., *Expression analysis of G Protein-Coupled Receptors in mouse macrophages.* Immunome Res, 2008. **4**: p. 5.

140. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.

141. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.

142. John, S., et al., *Chromatin accessibility pre-determines glucocorticoid receptor binding patterns.* Nat Genet, 2011. **43**(3): p. 264-8.

143. Chan, K.M., et al., *Hematopoiesis and immunity of HOXB4-transduced embryonic stem cell-derived hematopoietic progenitor cells.* Blood, 2008. **111**(6): p. 2953-61.

144. McKinney-Freeman, S.L., O. Naveiras, and G.Q. Daley, *Isolation of hematopoietic stem cells from mouse embryonic stem cells.* Curr Protoc Stem Cell Biol, 2008. **Chapter 1**: p. Unit 1F 3.

145. Thomas, M.L., *The leukocyte common antigen family.* Annual review of immunology, 1989. **7**: p. 339-69.

146. Hao, Q.L., et al., *A functional comparison of CD34 + CD38- cells in cord blood and bone marrow.* Blood, 1995. **86**(10): p. 3745-53.

147. Chen, W., et al., *Lentiviral vector transduction of hematopoietic stem cells that mediate long-term reconstitution of lethally irradiated mice.* Stem Cells, 2000. **18**(5): p. 352-9.

148. Wilson, N.K., et al., *Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators.* Cell Stem Cell, 2010. **7**(4): p. 532-44.

149. Whitlock, C.A. and O.N. Witte, *Long-term culture of B lymphocytes and their precursors from murine bone marrow.* Proc Natl Acad Sci U S A, 1982. **79**(11): p. 3608-12.

150. Sudo, T., et al., *Interleukin 7 production and function in stromal cell-dependent B cell development.* J Exp Med, 1989. **170**(1): p. 333-8.

151. Huang, P., et al., *Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors.* Nature, 2011. **475**(7356): p. 386-9.

152. Yomogida, K., et al., *Developmental stage- and spermatogenic cycle-specific expression of transcription factor GATA-1 in mouse Sertoli cells.* Development, 1994. **120**(7): p. 1759-66.

153. Ran, F.A., et al., *Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity.* Cell, 2013. **154**(6): p. 1380-9.

154. Mostoslavsky, G., et al., *Efficiency of transduction of highly purified murine hematopoietic stem cells by lentiviral and oncoretroviral vectors under conditions of minimal in vitro manipulation.* Mol Ther, 2005. **11**(6): p. 932-40.

155. Wang, X., et al., *Dynamic tracking of human hematopoietic stem cell engraftment using in vivo bioluminescence imaging.* Blood, 2003. **102**(10): p. 3478-82.

156. Tian, X., et al., *Bioluminescent imaging demonstrates that transplanted human embryonic stem cell-derived CD34(+) cells preferentially develop into endothelial cells.* Stem Cells, 2009. **27**(11): p. 2675-85.

157. Yu, S., et al., *The TCF-1 and LEF-1 transcription factors have cooperative and opposing roles in T cell development and malignancy.* Immunity, 2012. **37**(5): p. 813-26.

158. Sadelain, M., E.P. Papapetrou, and F.D. Bushman, *Safe harbours for the integration of new DNA in the human genome.* Nat Rev Cancer, 2012. **12**(1): p. 51-8.

159. Hong, J.S., et al., *Excellent in vivo bystander activity of fludarabine phosphate against human glioma xenografts that express the escherichia coli purine nucleoside phosphorylase gene.* Cancer Res, 2004. **64**(18): p. 6610-5.

160. Gafni, O., et al., *Derivation of novel human ground state naive pluripotent stem cells.* Nature, 2013. **504**(7479): p. 282-6.

161. Pinto do, O.P., A. Kolterud, and L. Carlsson, *Expression of the LIM-homeobox gene LH2 generates immortalized steel factor-dependent multipotent hematopoietic precursors.* EMBO J, 1998. **17**(19): p. 5744-56.

162. Thomson, J.A., et al., *Embryonic stem cell lines derived from human blastocysts.* Science, 1998. **282**(5391): p. 1145-7.

163. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.

164. Unternaehrer, J.J. and G.Q. Daley, *Induced pluripotent stem cells for modelling human diseases.* Philos Trans R Soc Lond B Biol Sci, 2011. **366**(1575): p. 2274-85.

165. Miura, K., et al., *Variation in the safety of induced pluripotent stem cell lines.* Nature biotechnology, 2009. **27**(8): p. 743-5.

166. Racila, D., et al., *Transient expression of OCT4 is sufficient to allow human keratinocytes to change their differentiation pathway.* Gene Ther, 2011. **18**(3): p. 294-303.

167. Kim, J.B., et al., *Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors.* Nature, 2008. **454**(7204): p. 646-50.

168. Li, W., et al., *Generation of rat and human induced pluripotent stem cells by combining genetic reprogramming and chemical inhibitors.* Cell Stem Cell, 2009. **4**(1): p. 16-9.

169. Utikal, J., et al., *Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells.* J Cell Sci, 2009. **122**(Pt 19): p. 3502-10.

170. Staerk, J., et al., *Reprogramming of human peripheral blood cells to induced pluripotent stem cells.* Cell Stem Cell, 2010. **7**(1): p. 20-4.

171. Bennett, E.M., et al., *Designer gene therapy using an Escherichia coli purine nucleoside phosphorylase/prodrug system.* Chem Biol, 2003. **10**(12): p. 1173-81.

172. Papapetrou, E.P., et al., *Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation.* Proc Natl Acad Sci U S A, 2009. **106**(31): p. 12759-64.

173. Hay, D.C., et al., *Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells.* Stem Cells, 2004. **22**(2): p. 225-35.

174. McCormack, M.P., et al., *The LMO2 T-cell oncogene is activated via chromosomal translocations or retroviral insertion during gene therapy but has no mandatory role in normal T-cell development.* Mol Cell Biol, 2003. **23**(24): p. 9003-13.

175. Hacein-Bey-Abina, S., et al., *LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1.* Science, 2003. **302**(5644): p. 415-9.

176. Akashi, K., et al., *A clonogenic common myeloid progenitor that gives rise to all myeloid lineages.* Nature, 2000. **404**(6774): p. 193-7.

177. Coffin, J.M., S.H. Hughes, and H.E. Varmus, *The Interactions of Retroviruses and their Hosts*, in *Retroviruses*, J.M. Coffin, S.H. Hughes, and H.E. Varmus, Editors. 1997: Cold Spring Harbor (NY).

178. Mitchell, R.S., et al., *Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences.* PLoS Biol, 2004. **2**(8): p. E234.

179. Montini, E., et al., *Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration.* Nat Biotechnol, 2006. **24**(6): p. 687-96.

180. Check, E., *Gene therapy: shining hopes dented - but not dashed.* Nature, 2002. **420**(6917): p. 735.

181. Check, E., *Safety panel backs principle of gene-therapy trials.* Nature, 2002. **420**(6916): p. 595.

182. Hacein-Bey-Abina, S., G. de Saint Basile, and M. Cavazzana-Calvo, *Gene therapy of X-linked severe combined immunodeficiency.* Methods Mol Biol, 2003. **215**: p. 247-59.

183. Hacein-Bey-Abina, S., et al., *A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency.* N Engl J Med, 2003. **348**(3): p. 255-6.

184. Di Nunzio, F., et al., *HIV-derived vectors for therapy and vaccination against HIV.* Vaccine, 2012. **30**(15): p. 2499-509.

185. Maude, S.L., et al., *Chimeric antigen receptor T cells for sustained remissions in leukemia.* N Engl J Med, 2014. **371**(16): p. 1507-17.

186. Biffi, A., et al., *Lentiviral vector common integration sites in preclinical models and a clinical trial reflect a benign integration bias and not oncogenic selection.* Blood, 2011. **117**(20): p. 5332-9.

187. Desbordes, S.C., et al., *High-throughput screening assay for the identification of compounds regulating self-renewal and differentiation in human embryonic stem cells.* Cell Stem Cell, 2008. **2**(6): p. 602-12.

188. Yu, J., et al., *Human induced pluripotent stem cells free of vector and transgene sequences.* Science, 2009. **324**(5928): p. 797-801.

189. Stadtfeld, M. and K. Hochedlinger, *Induced pluripotency: history, mechanisms, and applications.* Genes Dev, 2010. **24**(20): p. 2239-63.

190. Duong, K.L., et al., *Identification of hematopoietic-specific regulatory elements from the CD45 gene and use for lentiviral tracking of transplanted cells.* Exp Hematol, 2014.

191. Bibel, M., et al., *Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells.* Nat Protoc, 2007. **2**(5): p. 1034-43.

192. Funston, G.M., et al., *Expression of heterologous genes in oncolytic adenoviruses using picornaviral 2A sequences that trigger ribosome skipping.* J Gen Virol, 2008. **89**(Pt 2): p. 389-96.

193. Park, D., et al., *Nestin is required for the proper self-renewal of neural stem cells.* Stem Cells, 2010. **28**(12): p. 2162-71.

194. Wiedenmann, B., et al., *Synaptophysin: a marker protein for neuroendocrine cells and neoplasms.* Proc Natl Acad Sci U S A, 1986. **83**(10): p. 3500-4.

195. Feng, L., M.E. Hatten, and N. Heintz, *Brain lipid-binding protein (BLBP): a novel signaling system in the developing mammalian CNS.* Neuron, 1994. **12**(4): p. 895-908.

196. Li, R., et al., *A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts.* Cell Stem Cell, 2010. **7**(1): p. 51-63.

197. Masui, S., et al., *Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells.* Nat Cell Biol, 2007. **9**(6): p. 625-35.

198. Dechat, T., K. Gesson, and R. Foisner, *Lamina-independent lamins in the nuclear interior serve important functions.* Cold Spring Harb Symp Quant Biol, 2010. **75**: p. 533-43.

199. Dechat, T., et al., *Nuclear lamins.* Cold Spring Harb Perspect Biol, 2010. **2**(11): p. a000547.

200. Vergnes, L., et al., *Lamin B1 is required for mouse development and nuclear integrity.* Proc Natl Acad Sci U S A, 2004. **101**(28): p. 10428-33.

201. Coffinier, C., L.G. Fong, and S.G. Young, *LINCing lamin B2 to neuronal migration: growing evidence for cell-specific roles of B-type lamins.* Nucleus, 2010. **1**(5): p. 407-11.

202. Yang, S.H., et al., *An absence of both lamin B1 and lamin B2 in keratinocytes has no effect on cell proliferation or the development of skin and hair.* Hum Mol Genet, 2011. **20**(18): p. 3537-44.

203. Coffinier, C., et al., *Abnormal development of the cerebral cortex and cerebellum in the setting of lamin B2 deficiency.* Proc Natl Acad Sci U S A, 2010. **107**(11): p. 5076-81.

204. Coffinier, C., et al., *Deficiencies in lamin B1 and lamin B2 cause neurodevelopmental defects and distinct nuclear shape abnormalities in neurons.* Mol Biol Cell, 2011. **22**(23): p. 4683-93.

205. Silva, J., et al., *Nanog is the gateway to the pluripotent ground state.* Cell, 2009. **138**(4): p. 722-37.

206. Miyanari, Y. and M.E. Torres-Padilla, *Control of ground-state pluripotency by allelic regulation of Nanog.* Nature, 2012. **483**(7390): p. 470-3.

207. Theunissen, T.W., et al., *Reprogramming capacity of Nanog is functionally conserved in vertebrates and resides in a unique homeodomain.* Development, 2011. **138**(22): p. 4853-65.

208. Schwarz, Benjamin A., et al., *Nanog Is Dispensable for the Generation of Induced Pluripotent Stem Cells.* Current Biology. **24**(3): p. 347-350.

209. Polo, J.M., et al., *Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells.* Nat Biotechnol, 2010. **28**(8): p. 848-55.

210. Abad, M., et al., *Reprogramming in vivo produces teratomas and iPS cells with totipotency features.* Nature, 2013. **502**(7471): p. 340-345.

211. Jiang, J., et al., *Hemgn is a direct transcriptional target of HOXB4 and induces expansion of murine myeloid progenitor cells.* Blood, 2010. **116**(5): p. 711-9.

212. Jackson, M., et al., *HOXB4 can enhance the differentiation of embryonic stem cells by modulating the hematopoietic niche.* Stem Cells, 2012. **30**(2): p. 150-60.

213. Forrester, L.M. and M. Jackson, *Mechanism of action of HOXB4 on the hematopoietic differentiation of embryonic stem cells.* Stem Cells, 2012. **30**(3): p. 379-85.

214. Tosh, D., C.N. Shen, and J.M. Slack, *Differentiated properties of hepatocytes induced from pancreatic cells.* Hepatology, 2002. **36**(3): p. 534-43.

215. Scarpelli, D.G. and M.S. Rao, *Differentiation of regenerating pancreatic cells into hepatocyte-like cells.* Proc Natl Acad Sci U S A, 1981. **78**(4): p. 2577-81.

216. Eberhard, D., et al., *In vitro reprogramming of pancreatic cells to hepatocytes.* Methods Mol Biol, 2010. **636**: p. 285-92.

217. Cobaleda, C., W. Jochum, and M. Busslinger, *Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors.* Nature, 2007. **449**(7161): p. 473-7.

218. Graf, T., *Differentiation plasticity of hematopoietic cells.* Blood, 2002. **99**(9): p. 3089-101.

219. Laiosa, C.V., et al., *Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors.* Immunity, 2006. **25**(5): p. 731-44.

220. Xie, H., et al., *Stepwise reprogramming of B cells into macrophages.* Cell, 2004. **117**(5): p. 663-76.

221. Choi, J., et al., *MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes.* Proc Natl Acad Sci U S A, 1990. **87**(20): p. 7988-92.

222. Davis, R.L., H. Weintraub, and A.B. Lassar, *Expression of a single transfected cDNA converts fibroblasts to myoblasts.* Cell, 1987. **51**(6): p. 987-1000.

223. Tapscott, S.J., et al., *MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts.* Science, 1988. **242**(4877): p. 405-11.

224. Feng, R., et al., *PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells.* Proc Natl Acad Sci U S A, 2008. **105**(16): p. 6057-62.

225. Wang, Q., et al., *Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis.* Proc Natl Acad Sci U S A, 1996. **93**(8): p. 3444-9.

226. Shivdasani, R.A., E.L. Mayer, and S.H. Orkin, *Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL.* Nature, 1995. **373**(6513): p. 432-4.

227. Bender, T.P., et al., *Critical functions for c-Myb at three checkpoints during thymocyte development.* Nat Immunol, 2004. **5**(7): p. 721-9.

228. Wang, Y., et al., *Cdx gene deficiency compromises embryonic hematopoiesis in the mouse.* Proc Natl Acad Sci U S A, 2008. **105**(22): p. 7756-61.

229. Magnusson, M., et al., *Hoxa9/hoxb3/hoxb4 compound null mice display severe hematopoietic defects.* Exp Hematol, 2007. **35**(9): p. 1421-8.

230. Bjornsson, J.M., et al., *Reduced proliferative capacity of hematopoietic stem cells deficient in Hoxb3 and Hoxb4.* Mol Cell Biol, 2003. **23**(11): p. 3872-83.

231. Brun, A.C., et al., *Hoxb4-deficient mice undergo normal hematopoietic development but exhibit a mild proliferation defect in hematopoietic stem cells.* Blood, 2004. **103**(11): p. 4126-33.

232. Lawrence, H.J., et al., *The role of HOX homeobox genes in normal and leukemic hematopoiesis.* Stem Cells, 1996. **14**(3): p. 281-91.

233. Magli, M.C., C. Largman, and H.J. Lawrence, *Effects of HOX homeobox genes in blood cell differentiation.* J Cell Physiol, 1997. **173**(2): p. 168-77.

234. Kanai-Azuma, M., et al., *Depletion of definitive gut endoderm in Sox17-null mutant mice.* Development, 2002. **129**(10): p. 2367-79.

235. Kim, I., T.L. Saunders, and S.J. Morrison, *Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells.* Cell, 2007. **130**(3): p. 470-83.

236. Vicente, C., et al., *The role of the GATA2 transcription factor in normal and malignant hematopoiesis.* Crit Rev Oncol Hematol, 2012. **82**(1): p. 1-17.

237. Oguro, H., et al., *Differential impact of Ink4a and Arf on hematopoietic stem cells and their bone marrow microenvironment in Bmi1-deficient mice.* J Exp Med, 2006. **203**(10): p. 2247-53.

238. Guo, Y., et al., *Polycomb group gene Bmi1 plays a role in the growth of thymic epithelial cells.* Eur J Immunol, 2011. **41**(4): p. 1098-107.

239. Terano, T., et al., *Transcriptional control of fetal liver hematopoiesis: dominant negative effect of the overexpression of the LIM domain mutants of LMO2.* Exp Hematol, 2005. **33**(6): p. 641-51.

240. Yamada, Y., et al., *The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis.* Proc Natl Acad Sci U S A, 1998. **95**(7): p. 3890-5.

241. Takada, S., et al., *Wnt-3a regulates somite and tailbud formation in the mouse embryo.* Genes Dev, 1994. **8**(2): p. 174-89.

242. Luis, T.C. and F.J. Staal, *WNT proteins: environmental factors regulating HSC fate in the niche.* Ann N Y Acad Sci, 2009. **1176**: p. 70-6.

243. Luis, T.C., et al., *Wnt3a nonredundantly controls hematopoietic stem cell function and its deficiency results in complete absence of canonical Wnt signaling.* Blood, 2010. **116**(3): p. 496-7.

244. Nemeth, M.J., et al., *Wnt5a inhibits canonical Wnt signaling in hematopoietic stem cells and enhances repopulation.* Proc Natl Acad Sci U S A, 2007. **104**(39): p. 15436-41.

245. Xu, Y., et al., *LH-2: a LIM/homeodomain gene expressed in developing lymphocytes and neural cells.* Proc Natl Acad Sci U S A, 1993. **90**(1): p. 227-31.

246. Porter, F.D., et al., *Lhx2, a LIM homeobox gene, is required for eye, forebrain, and definitive erythrocyte development.* Development, 1997. **124**(15): p. 2935-44.

247. Jacks, T., et al., *Effects of an Rb mutation in the mouse.* Nature, 1992. **359**(6393): p. 295-300.

248. Donze, D., T.M. Townes, and J.J. Bieker, *Role of erythroid Kruppel-like factor in human gamma- to beta-globin gene switching.* J Biol Chem, 1995. **270**(4): p. 1955-9.

249. Nuez, B., et al., *Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene.* Nature, 1995. **375**(6529): p. 316-8.

250. Perkins, A.C., A.H. Sharpe, and S.H. Orkin, *Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF.* Nature, 1995. **375**(6529): p. 318-22.

251. Pang, C.J., et al., *Kruppel-like factor 1 (KLF1), KLF2, and Myc control a regulatory network essential for embryonic erythropoiesis.* Mol Cell Biol, 2012. **32**(13): p. 2628-44.

252. Wang, L.C., et al., *Yolk sac angiogenic defect and intra-embryonic apoptosis in mice lacking the Ets-related factor TEL.* EMBO J, 1997. **16**(14): p. 4374-83.

253. Wang, L.C., et al., *The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow.* Genes Dev, 1998. **12**(15): p. 2392-402.

254. Okita, K., et al., *Generation of mouse-induced pluripotent stem cells with plasmid vectors.* Nat Protoc, 2010. **5**(3): p. 418-28.

255. Nakano, T., H. Kodama, and T. Honjo, *Generation of lymphohematopoietic cells from embryonic stem cells in culture.* Science, 1994. **265**(5175): p. 1098-101.

256. Cumano, A., et al., *The influence of S17 stromal cells and interleukin 7 on B cell development.* Eur J Immunol, 1990. **20**(10): p. 2183-9.

257. Riddell, J., et al., *Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors.* Cell, 2014. **157**(3): p. 549-64.

258. Furth, P.A., et al., *Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter.* Proc Natl Acad Sci U S A, 1994. **91**(20): p. 9302-6.

259. JAISSER, F., *Inducible Gene Expression and Gene Modification in Transgenic Mice.* Journal of the American Society of Nephrology, 2000. **11**(suppl 2): p. S95-S100.

260. Jaisser, F., *Inducible gene expression and gene modification in transgenic mice.* J Am Soc Nephrol, 2000. **11 Suppl 16**: p. S95-S100.

261. Trevisan, M. and N.N. Iscove, *Phenotypic analysis of murine long-term hemopoietic reconstituting cells quantitated competitively in vivo and comparison with more advanced colony-forming progeny.* J Exp Med, 1995. **181**(1): p. 93-103.

262. Hyun, I., et al., *New advances in iPSC research do not obviate the need for human embryonic stem cells.* Cell Stem Cell, 2007. **1**(4): p. 367-8.