In Vitro Models to Study the Properties of the TCRα Locus Control Region

by

Armin Lahiji

A dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2013

©2013

ARMIN LAHIJI

All rights reserved

This manuscript has been read and accepted for the Graduate Faculty in Biology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

9/10/13	Dr. Benjamin D. Ortiz, Hunter College
Date	Chair of Examining Committee

9/17/13	Dr. Laurel A. Eckhardt
Date	Executive Officer

Dr. Laurel Eckhardt, Hunter College

Dr. Diego Loayza, Hunter College

Dr. Derek B. Sant'Angelo, Robert Wood Johnson Medical School

Dr. Jayanta Chaudhuri, Memorial Sloan-Kettering Cancer Center

Supervising Committee

The City University of New York

ABSTRACT

In Vitro Models to Study the Properties of the TCRα Locus Control Region

by

Armin Lahiji

Advisor: Benjamin D. Ortiz, Ph.D.

Locus Control Regions (LCR) are *cis*-acting gene regulatory elements with the unique, integration site-independent ability to transfer the characteristics of their locus-of-origin's gene expression pattern to a linked transgene. LCR activities have been discovered in numerous T cell-lineage gene loci. These elements can be adapted to the design of gene therapy vectors that direct robust therapeutic gene expression to the T cell progeny of engineered stem cells. Currently, transgenic mice provide the only experimental approach that wholly supports all the critical aspects of LCR activity. Herein, we report two cell culture models to study the properties of the T cell receptor (TCR)-α LCR. The first is an *in vitro* embryonic stem cell differentiation model that has been optimized to manifest all key features of mouse TCRα LCR function. High level, copy number-related TCRα LCR-linked reporter gene expression levels are cell type-restricted in this system, and are upregulated during the expected stage transition of T cell development. The ability of this LCR to overcome position effects may be due to barrier insulator-like activity within. To further

explore this possibility, we established a second model that seems to support this notion. The characterization and identification of the sequences involved in this possible barrier insulator will provide an additional vertebrate model for the study of insulators. This study additionally validates a novel, tractable and more rapid approach for the study of LCR activity in T cells, and its translation to therapeutic genetic engineering.

ACKNOWLEDGEMENTS

Dr. Ben Ortiz, from day one, has been an inspiration to me. Thank you for giving me the opportunity to be part of your lab, and for guiding me on my path to becoming a better scientist and a stronger writer. You are an amazing mentor, teacher and a friend. You have always pushed me to make myself better in all aspects of my life. I aspire to be as good of a scientist and teacher as you are one day.

Thank you to the Ortiz laboratory, past and present, for truly making every day an enjoyable experience. To Martina Kučerová-Levisohn for being my sounding board through everything. I turned to you, literally, every time I was stuck and you were always there. The way you juggle family, lab, life and everything in between is awe inspiring and motivating.

Thank you to those who have helped me on my project(s). Stefan Knirr, Zayd Daruwala, Randy Arroyave, Blanca Andino, Kyle Wong, Judy Rosena, Forest Ray, Joel Morales-Rosado, Mike Rist, Rosa Mejia and Aneta Rogoz. To the new graduate students who are taking parts of my project to better places, Gayathri Devi Raghupathy and Jordana Lovett: I leave knowing my work is in your capable hands. Thank you to Dr. Janette Gomos-Klein for being an ear and giving me your perspective.

I am grateful to my committee members Dr. Laurel Eckhardt, Dr. Diego Loayza, Dr. Derek Sant'Angelo, and Dr. Jayanta Chaudhuri, for your support through out the years. Thank you for your advice and guidance.

I would like to express my gratitude to Dr. Juan Carlos Zúñiga-Pflücker and Roxanne Holmes for their endless expertise, support, and advice during our attempts to establish the embryonic stem cell differentiation model. I also would like to thank Dr. Michel Sadelain for his support and for providing us with lentiviral vectors.

I would also like to thank Dr. Adrienne Alaie for giving me an opportunity to teach throughout these years. It has been your guidance and effective teaching style that has shaped my abilities as a teacher.

Thank you to Joon Kim for your help and advice regarding flow cytometry. I would also like to acknowledge all the faculty and staff at Hunter College with whom I have interacted throughout the years.

Thank you to my family: my Dad, Mansoor, my Mom, Nahid, my sister, Arta, and my brother, Arad. Thank you for being supportive from far away and for always believing in me. Thank you for the visits, the care packages, the love, and everything in between. Thank you for being an ear and a rock for support through ups and downs. Thank you to Samantha for your love and support. You have always pushed me and encouraged me to strive for success. I love you all.

TABLE OF CONTENTS

TABLE OF FIGURES

Chapter 1: Introduction

1.1 Gene Regulation

The genome of an organism consists of tens of thousands of genes coding for different proteins. However, not all of these genes are transcribed concurrently in a given cell. Furthermore, there will be distinctive gene transcription patterns in the different cells in an organism. The expression of these genes can be regulated at various stages, ranging from the onset of their transcription to the translation of their mRNA products. At the transcriptional level, the most basic of regulatory units for genes are promoters, enhancers, and transcription factors. These regulatory elements directly influence the activity of RNA polymerase on DNA. Other regulatory components such as histone and DNA modifying enzymes[1], barrier insulators[2,3], and locus control regions (LCRs)[4-7] can regulate the expression of a gene by remodeling the surrounding chromatin. These latter elements support higher order regulatory mechanisms that can coordinate the formation of heterochromatic (closed chromatin) or euchromatic (open chromatin) regions to modulate the access of regulatory factors to a gene. For example, the formation of either of these two states of chromatin can be regulated by modifying enzymes that will acetylate or methylate histone proteins, around which the DNA is wrapped[8]. How these acetylation or methylation "marks" can influence the surrounding chromatin is still not completely understood. But, at least two modes of influence are supported by the literature. One is that the marks directly influence chromatin structure by changing the electrostatic charge or by modifying the interactions between different nucleosomes[8]. The second is that these marks, in turn,

recruit other chromatin remodeling proteins[8]. The regulation of chromatin structure at a given locus is one of the many pivotal aspects of regulating gene expression.

The interplay of these regulatory elements contribute to the appropriate expression of a gene. Understanding how these regulatory elements function and interact can have significant implications across both basic and therapeutic sciences.

1.2 Locus Control Regions

LCRs are gene regulatory elements known to confer specific properties upon the expression of a linked transgene. These properties include integration site-independence and copy number-dependent mRNA production levels with predictable spatiotemporal characteristics resulting in their *in vivo* expression in specific tissues. LCRs are composed of a collection of DNase I hypersensitive sites (HS) that are thought to collectively regulate the chromatin structure of a linked gene locus. Each HS can have a different function contributing to the overall regulatory properties of the LCR.

The mechanisms by which LCRs function have yet to be completely revealed. But, they are hypothesized to recruit histone modifying enzymes that alter the structure of surrounding chromatin allowing for the correct temperospatial expression of a linked transgene[4-7,9,10]. By possibly inhibiting the encroachment of surrounding heterochromatin into the domain of a linked gene, the LCR supports predictable expression of a linked transgene irrespective of its position of integration in a genome. This property leads to the mRNA expression levels to be linearly correlated with the number of transgene integrants, which is commonly referred to as copy number dependence[11].

There are multiple identified LCRs that are active in a variety of tissues. Some LCRs include human β-globin[11], human Growth Hormone[12], human HLA-B7[13], human CD2[14] and the mouse TCRα LCRs[15].

The first identified LCR was found in the locus containing the human β globin gene[11]. Much of our knowledge regarding LCRs stems from this particular LCR, as it is the most characterized one. The importance of this LCR was made evident in patients with β-thalassemia who have impaired synthesis of β-globin protein. These patients have a deletion, termed Hispanic deletion, in a region upstream of the β-globin gene that was later determined to be the locus control region[16]. The globin genes encode part of the oxygen-carrying molecule, hemoglobin, carried by red blood cells (RBCs). There are five globin genes (ε, Gγ, Aγ, β, and δ) within this locus that are activated at distinct developmental stages[17]. In humans, during development, the initial site of hematopoiesis is the yolk sac. Between day 16-19 of human embryonic development, the site of hematopoiesis switches to the fetal liver, and finally after birth, hematopoiesis occurs in the bone marrow[17]. A different globin gene is expressed in each of these sites. The ε globin gene is expressed in the yolk sac, while the two γ genes are expressed in the fetal liver and subsequently the β or δ globin genes are expressed in the RBCs arising from the bone marrow[17]. This successive developmental switching of globin genes is regulated by the β-globin LCR located upstream of these five genes[4].

The importance of the β-globin LCR to the proper expression of globin genes was further exemplified by *in vivo* studies. Transgenic mice bearing only the human β-globin gene with its promoter failed to express physiological levels of β-globin[18,19]. Furthermore this low

level expression was not consistent between different mouse lines[20]. This low, inconsistent expression in transgenic mice was only rescued when the gene fragment also included the upstream LCR[11].

While a number of LCRs have been identified, only a subset of these have been shown to confer tissue specific, copy number dependent, and integration site independent expression on heterogeneous or unrelated transgenes[21,22]. This subset includes the human β globin[23], human CD2[14] and the mouse TCRα LCR[15].

1.3 TCRα/*Dad1* Locus

We study the mouse T cell receptor (TCR) α LCR, which is located in an especially complex gene locus flanked by differentially expressed genes. Upstream of the LCR are the αβ T cell specific Cα exons of the TCR α gene. Downstream, the ubiquitously expressed *Dad1* gene is present[24] **(Figure 1)**.

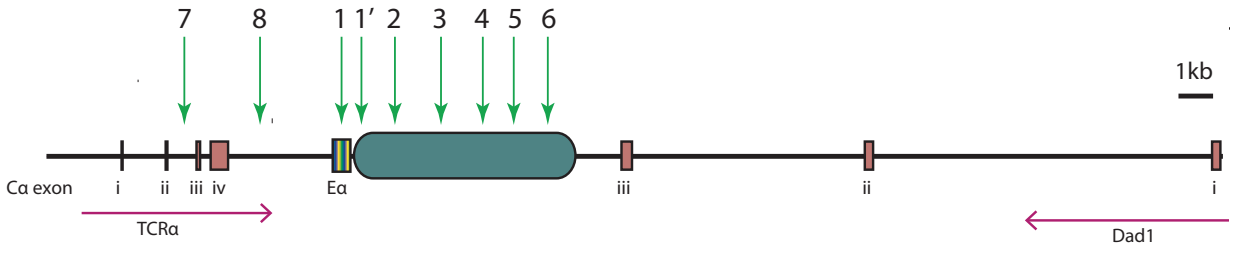


FIGURE 1 The TCRα LCR genomic region.

Diagram of the TCRα/Dad1 locus. Vertical arrows depict DNase HS 1–8 of the TCRα LCR. The multicolored box marks the Eα classical transcriptional enhancer. All other boxes indicate exons of their respective genes. Horizontal arrows indicate the transcription orientation of the genes. Diagram is drawn to scale.

The TCR is a heterodimer composed of either αβ or γδ chain subunits. This receptor binds to major histocompatibility complex (MHC) molecules that display specific peptide antigens. The binding of an antigen to the TCR will lead to the clonal selection and proliferation of that T cell. The functionality of this receptor is crucial for the cellular adaptive immune response. Circulating T cells predominately express αβ TCRs. For αβ T cells to function properly, it is essential for the TCRα receptor to be appropriately expressed.

Dad1 has been shown to be involved in N-linked glycosylation and can act as an anti-apoptotic factor. Furthermore, the deletion of *Dad1* results in embryonic lethality[24]. The *TCRα* and *Dad 1* genes in this locus are differentially expressed. *TCRα* is expressed only in T lymphocytes, whereas, the *Dad1* gene is expressed in all cells, lymphoid and non-lymphoid[24]. Dad1 expression can be seen as early as day seven of embryonic development, whereas, the earliest TCRα mRNA detected is day sixteen of embryonic development[24]. It remains unknown if the LCR has any effect on Dad1 expression. However, Dad1 and TCRα mRNA expression are upregulated concurrently when T cells are fully matured[25].

The location of the TCRα LCR, in between these two differentially expressed genes, might indicate its role in coordinating the two distinct expression patterns of these genes. Therefore, understanding of the intricacies involved in the regulation of this locus is of great interest.

1.4 Mouse TCRα LCR

The mouse TCRα LCR is located on chromosome 14 and consists of nine hypersensitive sites spanning 13.5kb of DNA[26,27]**(Figure 1)**. To date we have identified five regions of functionality, contained within four of these HS, that are required for complete LCR activity. The 5' end of the LCR is thought to confer its spatiotemporal regulatory properties, whereas the 3' end is proposed to have barrier insulator like activity and is required for protection against position effects (**Figure 2)**.

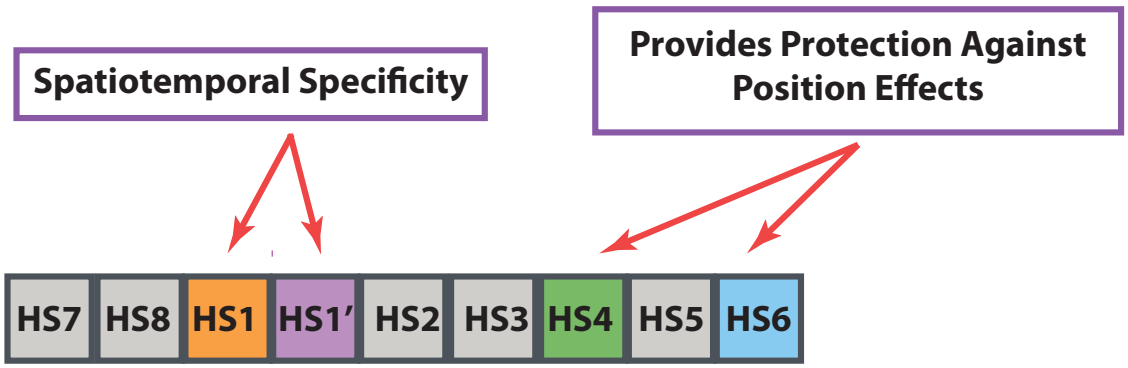


FIGURE 2 Overview of the TCRα LCR.

Colored boxes indicate the identified functional HS and are required for LCR activity. HS1 is the classical Eα enhancer. HS1' contains CTCF sites and along with HS1 confer tissue specificity. HS6 includes two functions regions, TF1-2-3 and HS6-316.

HS1 contains the classical enhancer (Eα) and along with HS1', confer tissue specificity upon a linked transgene[27,28]. Within HS1' there are two binding sites for the CTCF (CCTC-binding factor) transcription factor[29,30]. CTCF has been previously implicated in the enhancer blocking activity of insulators and the activation or repression of some genes[31]. However, these two CTCF sites seem not to be strictly required for LCR activity[30].

The third indispensible HS is HS4. It has been shown to be required for full LCR activity and its deletion leads to significantly reduced levels of transgene expression[30]. HS4 DNA is methylated in all tissues examined with the exception of lymphoid organs[32]. This mostly tissue selective de-methylation in lymphoid tissue is contingent on the presence of HS1'[27]. The absence of methylation in this region is significant because DNA methylation can lead to the recruitment and binding of chromatin remodeling proteins which, in turn, can lead to silencing of nearby genes by formation of heterochromatin[8].

The patterns and levels of a transgene expression can be dependent on the location of integration within the host genome. A transgene integrated in a euchromatic region will most likely be expressed. In contrast, if the same transgene were to become integrated into a heterochromatic region it would likely be silenced. This can lead to the phenomenon commonly referred to as position effect variegation (PEV) [33].

HS6 possesses position effect suppression capacity on its own, both *in vivo* and *in vitro*[34,35]. There are two identified functional regions within HS6. One is a region consisting of three factor-binding sites termed thymic footprint 1, 2 and 3 (TF1-2-3)[34]. TF2 binds to AML-1/Runx1 and TF3 binds Elf-1 factors[35]. Families of both proteins have been shown to be involved in chromatin remodeling and the deletion of TF1-2-3 leads to lower expression of

the reporter transgenes per copy[35]. Elf-1 has been shown to be active in all stages of T cell development[36]. AML-1 has been implicated in CD4 gene silencing during T cell lineage specification[37]. Furthermore, AML-1 interacts with proteins involved in histone modifications[38]. The other identified functional region of HS6 is a 316bp sequence; its deletion leads to significant reduction of the ability of the TCRα LCR to suppress position effects [35].

To date, no known function has been attributed to the HS2, 3 and 5 regions. Located 5' to HS 1 are HS7 and HS8 and there have been some reports of silencer elements located within these regions[26]. However, their presence is not required for full LCR activity[15]. Altogether, these data suggest that components in HS1, HS1', and HS4 and HS6 combined may reproduce full LCR activity.

1.5 Studies of Locus Control Regions Using Cell Culture

Until recently, the predominant model used to study the activity of LCRs at ectopic genomic sites has been transgenic mice. This model has provided valuable knowledge to advance the field of LCRs. However, studying the complexity of the LCR activity, in particular the structure-function relationships that contribute to LCR activity, in transgenic mice is resource intensive and involves protracted experimental timetables. This has led to attempts to establish a cell culture model to serve as a more rapid model to study LCRs.

Previous attempts to use cultured cell lines to study LCRs have only been partially successful. For instance, β globin LCR activity is incomplete in directly transfected erythroid cell lines[39]. An important clue to explaining this puzzling result was reported the

following year. In this work, mouse embryonic fibroblasts, transfected with a reporter gene linked to the β globin LCR, were subsequently fused with an erythroid cell line. The fusion resulted in restoration of the copy number dependent reporter transgene expression indicative of full LCR activity[40]. These studies suggested the need for the β globin LCR DNA to be present, initially, in the genome of an uncommitted cell type, prior to the establishment of its complete activity upon later exposure to a differentiated nuclear environment[39,40]. Inspired by this work, we sought to design an *in vitro* cell culture model of complete LCR activity that would meet these apparent requirements.

Chapter 2: Embryonic Stem Cell *in Vitro Differentiation* model

2.1 Embryonic Stem Cells

Embryonic stem cells are pluripotent cells, which, *in vivo,* can give rise to all the cells that make up an embryo[41]. ESCs can be kept in an undifferentiated state in culture with the addition of leukemia inhibitory factor (LIF). When LIF is removed, ES cells will differentiate into three-dimensional structures called embryoid bodies and will give rise to many different cell types spontaneously[42]. This differentiation, however, can be directed to quantitatively yield a particular cell type, *in vitro,* upon the provision of the proper microenvironment and certain soluble factors. For example, ESC can be directed to differentiate into cells of the blood (hematopoiesis) in a bone marrow-like microenvironment[43].

2.2 Hematopoiesis

In a developing embryo, hematopoiesis occurs in the fetal liver. During mammalian embryogenesis, the fetal liver is seeded with hematopoietic stem cells (HSCs) by the yolk sac, placenta or AGM (aorta, gonads, mesonephros region)[44,45]. In adult mammals, the location of hematopoiesis shifts to the bone marrow. Here, HSCs will differentiate into progenitors that will eventually give rise to the organism's common myeloid or common lymphoid progenitors (CMP or CLP, respectively). CMPs will continue differentiating and give rise to red blood cells (RBCs), macrophages, myeloid dendritic cells (mDCs) and other

non-lymphocytes[46]. CLPs, on the other hand, will give rise to natural killer (NK), lymphoid dendritic (lDCs), B and T cells[46].

Over the past decades, the understanding of the factors that orchestrate hematopoietic differentiation has vastly increased[47,48]. For instance, erythropoietin drives differentiation of HSCs to RBCs[49], while macrophage colony-stimulating factor (M-CSF) drives their differentiation to macrophages[50]. Other factors involved in hematopoiesis have also been identified. Steel factor (SCF), expressed by stromal cells, has been identified as a ligand for the c-kit tyrosine kinase receptor[51]. Along with other cytokines, SCF drives the proliferation of the various blood cells[52]. SCF can work synergistically with interleukin 7 (IL-7) to promote granulocyte formation[53]. FMS like tyrosine kinase 3 ligand (Flt3-L) is initially involved in the differentiation of ESCs to HSCs[54]. Subsequently, during hematopoiesis, the presence of Flt3-L, SCF and IL-7 drive differentiation of HSCs towards B cell proliferation[54,55].

Although T cell precursors arise in the bone marrow, they complete their development in the thymus. The Notch pathway has been implicated in several binary cell fate decisions, including the differentiation of T over B cells[56]. Under normal conditions, T-lineage directing Notch ligands are absent in the bone marrow and present in the thymus. However, constitutive Notch expression in the bone marrow favors T cell rather than B cell development[57]. Conversely, tissue-specific knockout of Notch in the thymus leads to ectopic development of B cells[56]. It is still unknown how, exactly, Notch drives the differentiation of T cells over B cells. However, it is known that Notch is required for the proper expression lineage activating transcription factors[58].

Upon the binding of its ligand, the Notch receptor is cleaved at two different positions, liberating the intracellular domain of Notch (ICN), from the extracellular portion[58,59]. The ICN is then translocated in the nucleus where it acts with other transcription factors either activating or repressing several genes. In CLPs, Notch activates T cell specific genes such as GATA-3, Bcl11b, and HEB while negatively influencing genes required for B cell differentiation such as E2A, PAX5, and EBF[58].

Thymic epithelial stromal cells highly express Dll-4, while the thymic vasculature mainly expresses Dll-1[58]. Both Dll-1 and Dll-4 are ligands for the Notch pathway[59].

2.3 *In Vitro* Hematopoiesis and T Cell Development

In order to test whether the *in vivo* environments in which hematopoiesis occurs can be recapitulated in an *in vitro* model, ESCs were co-cultured with bone marrow stromal cell lines. In these studies, ESCs gave rise predominantly to macrophages[60]. However, co-cultures with a stromal cell line deficient for MCSF (OP9 cells) yielded various cell types within the hematopoietic lineage including neutrophils, macrophages, erythroid cells, mast cells, megakaryocytes, and B cells[61]. Analysis of OP9 cells showed high expression of SCF, but moderate levels of IL-7 and low levels of Flt3-L. Thus the addition of these two cytokines led to more robust differentiation of B cells but not T cells[62].

Prior to the development of the OP9 differentiation system, *in vitro* T cell development required the presence of a 3D thymic environment[63]. The addition of flk1^{+} blood mesoderm cells derived from ESCs/OP9 co-cultures onto a 3D thymic environment does, however, support inefficient T cell formation[63]. OP9 cells do not express Dll-1 or 4, for this reason the

co-culture of ES cells with OP9 cells supplemented with Flt3-L and IL-7 never yielded T cells[64]. ESCs co-cultured with OP9 cells transduced to express Dll1 or Dll4 can be driven to form T cells *in vitro*[63-65].

As T cells develop in mice, they progress through several stages. The first four of stages are termed double negative (DN) 1-4, characterized by the lack of cell surface markers CD4 and CD8. Two other cell surface proteins, CD44 and CD25, distinguish the four stages from each other (**Figure 3**). After the DN stages, the differentiating cells will express both CD4 and CD8, indicative of progression to the double positive (DP) stage (**Figure 3**). Finally, these cells will either differentiate to helper T cells, $CD4^+$, or cytotoxic T cells, $CD8^+$ (**Figure 3**). This final stage is called single positive (SP). Differentiation of ESCs on OP9-Dll1 cells will give rise to $CD8^+$ SP T cells. All stages of development can be tracked via flow cytometry. $CD4^+$ SP cells do not differentiate in this culture system. One reason is that OP9 cells lack expression of MHC II which is required for proper $CD4^+$ T cell development[37,66].

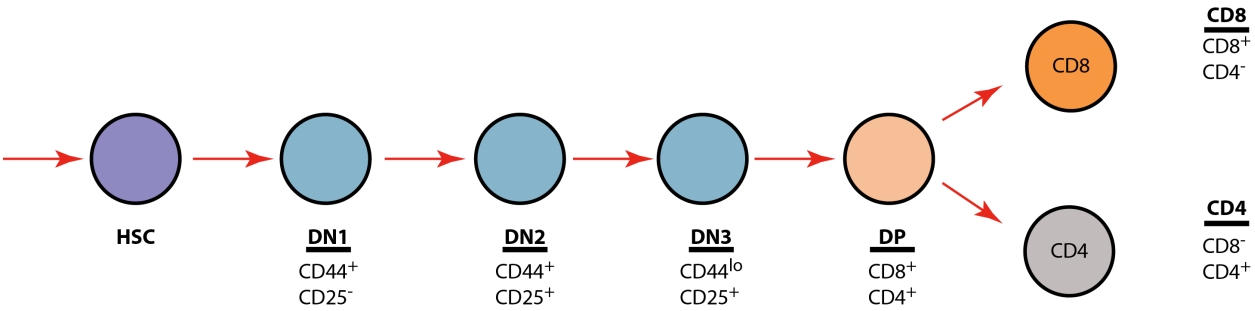


FIGURE 3 Mouse T cell development.

The development of T cells from HSCs. HSCs in the bone marrow can give rise to all blood cell types including lymphocytes. HSCs will give rise to the common lymphoid progenitor, which migrates to the thymus. In the thymus, T cell development can be tracked via flow cytometry by cell surface proteins. The hallmark of the first four stages of T cell development, called double negative (DN) stages, is the absence of both the CD4 and CD8 cell surface co-receptors. These four stages can be distinguished from each by two cell surface markers, CD44 and CD25. DN1 is $CD44^{+}$ and $CD25^{-}$, DN2 is $CD44^{+}$ $CD25^{+}$, DN3 is $CD44^{lo}$ and $CD25^{+}$, and DN4 (not shown) $CD44^{-}$ and $CD25^{-}$. The next stage is the double positive (DP) stage; these T cell progenitors express both CD4 and CD8 proteins. These cells can either differentiate to T helper cells ($CD4^{+}$, $CD8^{-}$) or cytotoxic T cells ($CD4^{-}$, $CD8^{+}$).

2.4 Applying the *in vitro* ESC Differentiation Model to Study the TCRα LCR

ESCs can be kept in an undifferentiated state given the proper cultured conditions. Once ESCs are cultured with OP9 cells (on day 0), they will start differentiating. The bone marrow environment, provided by the OP9 cells, will support the differentiation of ESC to blood mesoderm and these cells will display a tyrosine kinase receptor on their cell surface called FMS like-1, ($flk\text{-}1^{+}$)[67,68]. On day 5 of the co-culture, the differentiation of mesoderm-derived precursors can then be driven to HSCs with the addition of Flt3-L. The HSCs will differentiate to erythrocytes, monocytes, and B cells after the addition of IL-7 on day 8. Further inclusion of a Notch ligand Dll-1/4 in the OP9 cells signals differentiation of HSCs into T-lineage cells by day 18[69]. The differentiation of ESCs into the various lineage of the hematopoietic system in this model can be tracked by flow cytometry (**Figures 4 & 5**).

Although ES cells can be grown with OP9Dll-1 cells from the onset, a higher amount of mesodermal cells are formed when initially grown on OP9 cells and subsequently transferred onto OP9Dll-1/4 cells[69]. We aimed to study if this ESC *in vitro* differentiation system can recapitulate the properties of the TCRα LCR seen in transgenic mice.

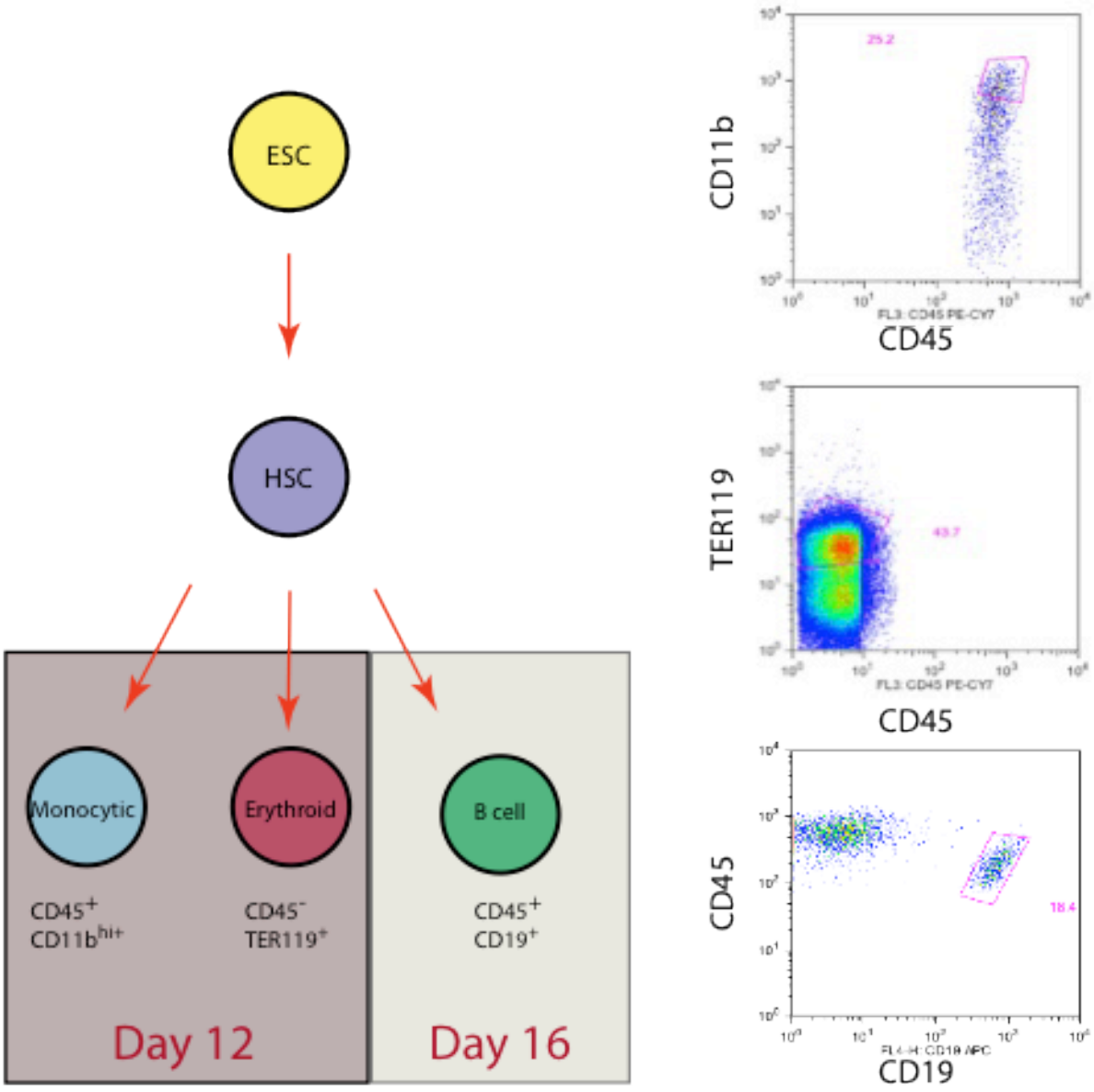


FIGURE 4 Analysis of cell differentiation in ESC differentiation system using OP9 cells by flow cytometry.

ESCs co-cultured with OP9 cells will give rise to monocytic, erythrocytic and B-cells over the span of sixteen days. These lineages can be tracked via flow cytometry. On day 12 of co-culture, one can detect monocytic and erythroid cells. Monocytes are distinguished by their expression of CD45 and CD11b, while erythrocytes are $CD45^{-}$ and $TER119^{+}$. On day 16 of this co-culture, differentiated cells should predominately be B cells, identified by the expression of CD45 and CD19 cell surface markers. Gates for flow cytometry shown on the right.

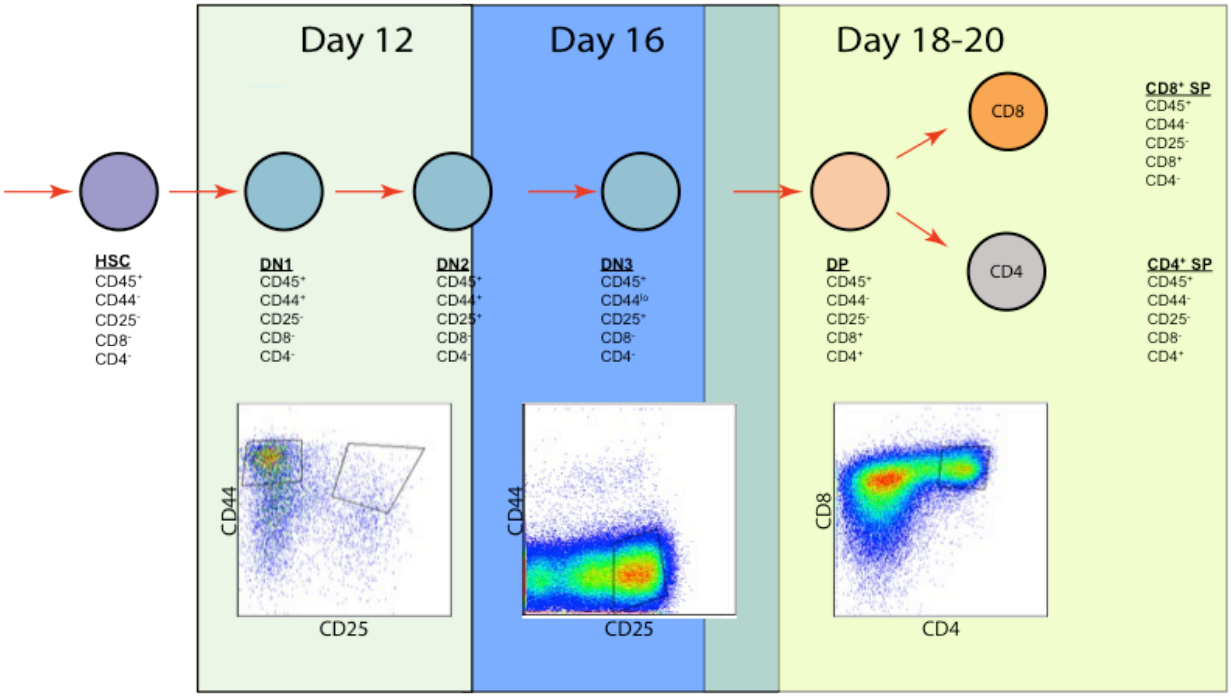


FIGURE 5 Analysis of mouse T cell development in OP9 differentiation system by flow cytometry.

The development of T cells in the OP9 differentiation system can be easily tracked by immunostaining. On day 12 of a typical co-culture on top of OP9-DL1 cells, DN1 and DN2 cells can be detected. DN1 cells are $CD45^+$, $CD44^+$, $CD25^-$, $CD4^-$, and $CD8^-$. DN2 cells are $CD45^+$, $CD44^+$, $CD25^+$, $CD4^-$, and $CD8^-$. Day 16 of a co-culture with OP9-DL1 cells mainly contains DN3 and DN4 cells with some DP cells. These can be detected with the following parameters: DN3: $CD45^+$, $CD44^{lo}$, $CD25^+$, $CD4^-$, and $CD8^-$; DN4: $CD45^+$, $CD44^-$, $CD25^-$, $CD4^-$, and $CD8^-$; DP: $CD45^+$, $CD44^-$, $CD25^-$, $CD4^+$, and $CD8^+$; and SP CD8: $CD45^+$, $CD44^-$, $CD25^-$, $CD4^-$, and $CD8^+$. On day 20, the culture mainly contains DP and SP $CD8^+$ cells with some DN3 and DN4 cells. Gates for flow cytometry shown on the bottom.

Chapter 3: Optimization of the *in vitro* ESC model to study the TCRα LCR

3.1 Selection of Reporter Gene

In selecting a reporter gene for assaying LCR activity in T cells differentiated *in vitro* from ESCs, we initially chose a cDNA encoding for a yellow fluorescent protein (YFP), a variant of the green fluorescent protein found in *Aequorea victoria*[70]. This reporter gene is widely used with transfected cell lines. The advantages of YFP include that it is a relatively short cDNA sequence encoding a fluorescent protein detectable without the use of antibodies. Unfortunately, these ESC clones did not produce detectable YFP levels once differentiated to T cells. The lack of expression of the reporter gene could be due to the known silencing of cDNA after differentiation or, to the presence of non-mammalian DNA[33,71].

Studies of the TCRα LCR in transgenic mice have employed both cognate (TCRα)[26] and heterologous mammalian reporter genes (human β-globin[15], HLAB7[60], and human CD2[72]). All of these reporter gene fragments are susceptible to position effects in transgenic mice in the absence of their own cognate LCR. However, when these reporter genes are linked to the TCRα LCR, their expression pattern closely resembles that of the endogenous TCRα gene in time and space, and displays copy number dependent mRNA levels irrespective of the site of integration. An ideal reporter gene for testing LCR activity in the progeny of *in vitro* differentiated ESCs would be one that produces a cell surface protein detectable by

flow cytometry. Of the aforementioned reporter genes, the hCD2 reporter protein is consistently detected on the cell surface. Therefore, we selected this gene for use as a reporter of the TCRα LCR activity in *in vitro* differentiated ESCs.

We use a hCD2 gene fragment which encodes for a non-signaling hCD2 protein (hCD2ΔT) as our reporter gene[73] (**Figure 6)**. Transgenic mice bearing an unlinked hCD2 reporter gene did not express the reporter gene at detectable levels[74]. However, when this construct was linked to the TCRα LCR (hCD2:1-8) (**Figure 6)**, the T cells of transgenic mice expressed high levels of this reporter gene. This expression was copy number dependent and site of integration independent. Furthermore, hCD2 protein expression was upregulated at the DN3 to DP transition, congruent with the developmental timing of endogenous TCRα expression[72].

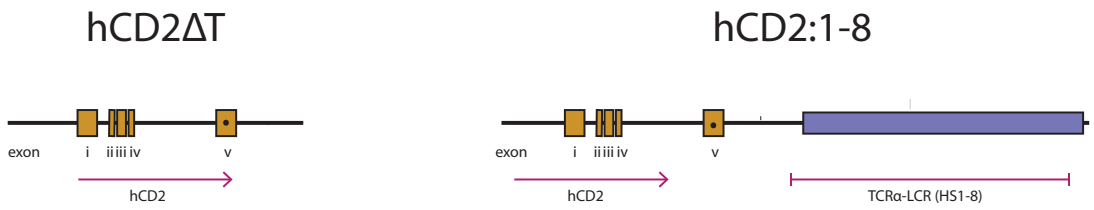


FIGURE 6 hCD2 and hCD2:1-8 Reporter transgenes.

Depiction of the hCD2ΔT and hCD2:1-8 transgenes. A premature stop codon (•) was introduced in exon V prior to the codons of the cytoplasmic tail[73]. In hCD2:1-8, the TCRα LCR cassette[15] (purple box) containing an exon-free HS1-8 fragment is linked to hCD2ΔT gene fragment.

3.2 Transfection of ESCs

An effective way of introducing a transgene into cells is electroporation. A common variation on this technique is nucleofection, which requires lower cell numbers and less DNA and results in higher transfection efficiency in ESCs. However, upon characterization of nucleofected transgene integration sites by Southern blot, severe truncations of the transgenes were prevalent in most stable-transfected ESC clones. Truncations within the reporter transgene can lead to skewed gene expression results, due to the potential for deletion of coding and/or regulatory components in some or all integrated transgene copies.

The hCD2:1-8 transgene is approximately 21kb in length. The observation that this large construct would be susceptible to shearing upon nucleofection has been corroborated by a recent report[75]. We thus switched to exponential decay pulsing for electroporation (BioRAD) as the method for transfection. Although this approach did require more cells and more DNA, and resulted in a smaller number of stable transfected clones, the frequency of truncated transgene integrants was much lower in the resultant clones.

The inclusion of bacterial DNA in reporter transgenes can lead to the silencing of the reporter gene[33]. Therefore, we liberated the vector backbone from our constructs. We then co-transfected our reporter gene fragment with a neomycin resistance gene fragment into ESCs.

3.3 Selection of Stable-transfected ESC Clones

Twenty-four hours after ESCs were co-transfected, G418 antibiotic was added at a concentration of 0.25mg/ml for two days and then reduced to 0.18mg/ml. This concentration was maintained for the duration of the selection period. If the live, remaining neomycin resistant colonies were not adequately expanding after seven days of selection, the concentration of G418 was reduced to 0.15mg/ml, to encourage colony proliferation. After ten days of selection, individual colonies were picked using a dissecting a microscope and propagated in 0.18mg/ml of G418.

3.4 Analysis of Integration Sites

We took multiple steps to characterize the transgene integration sites of the isolated clones. The first step was to confirm the presence of the transcriptional start site of the construct by PCR analysis. We designed primers to detect the non-endogenous hCD2 promoter region. The clones that contained this region of the transgene were then analyzed by Southern blot with a probe detecting the HS6 region, located at the 3'end of the LCR. Cells that incorporate multiple copies of a transgene in their genome tend to integrate these in tandem arrays in a "head to tail" arrangement. Therefore, the Southern blot strategy was designed to detect these head to tail junctions and any truncations thereof. The presence of intact head to tail tandem arrays was further confirmed via a PCR scheme designed with a forward primer located in HS6 and a reverse primer located in the 5' sequence flanking the hCD2 promoter.

One of the properties of the TCRα LCR is its ability to confer an mRNA expression level that bears a linear relationship to the number of integrated transgene copies. To test if the ESC differentiation model can support the copy number dependence property of the TCRα LCR one must carefully estimate the number of integrated transgene copies.

We derived the copy number in each clone by multiple Southern blots. We used a probe detecting the HS6 region, which is common to both the endogenous and transgene loci. To distinguish between the endogenous and transgene regions, a digestion strategy was designed that would generate a distinct HS6 containing fragment size for each. The probe hybridization signal intensities of transgene and endogenous bands were then quantified using a PhosphorImager. We calculate relative copy number estimates for each clone from an average of at least three independent Southern blots, each of which includes DNA samples from all clones to be directly compared in further analyses.

3.5 Differentiation of ESC Clones to Cell Types of the Hematopoietic System

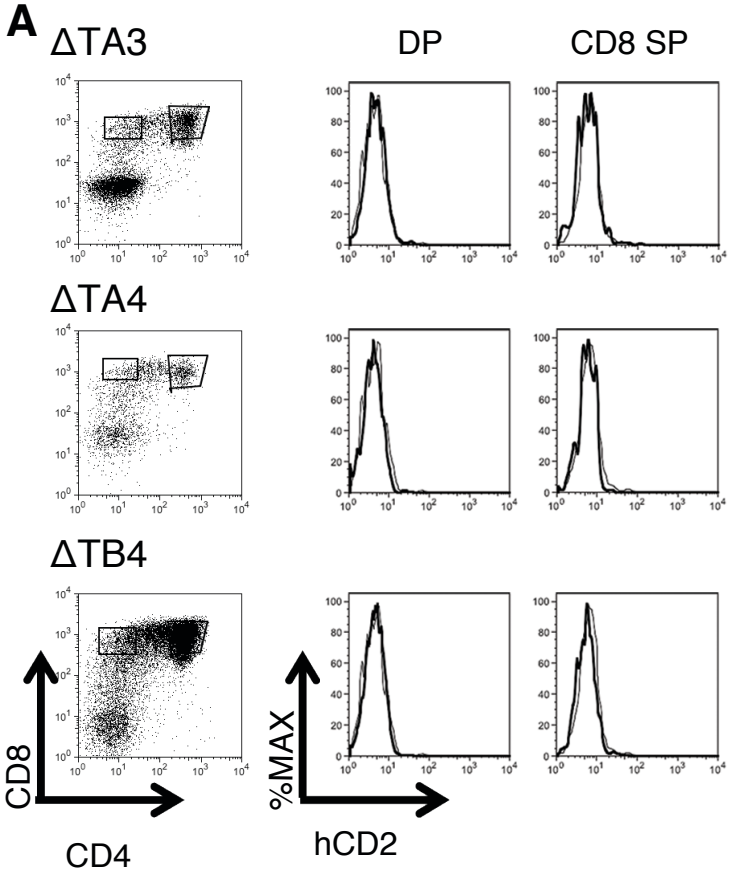
The original ESC differentiation protocol calls for a co-culture of $5x10^4$ ESCs with an 80% confluent layer of OP9 bone marrow stromal cells. After five days of co-culture, the ESCs should display qualitative differentiation to blood forming mesoderm-like structures. On this day, $5x10^5$ cells from the co-culture are re-plated onto a fresh layer of stromal cells in one 10cm plate for each future time point to be analyzed by flow cytometry (see Materials and Methods). Eight days after the start of the co-culture, the mesoderm should give rise to HSCs. These HSCs are now gently harvested and placed onto either an OP9 cell layer to generate monocytic, erythroid, and B cells or onto an OP9-DL1 cell layer to generate T cells.

To optimize this model for studies of the TCRα LCR, we made small, but important adjustments to this published protocol. We had observed that the progression of differentiation of transfected ESC clones tend to be at a slower rate compared to non-transfected ESCs. Therefore, to increase the yield of differentiating cells, we introduced an extra day of differentiation at two key time-points. The fifth and eight day are crucial junctures in this co-culture. If the formation of blood mesoderm on day five is less than 80%, we allow for an extra day of co-culture (day 5+1) at this point prior to transfer. We have observed that this delay improves the robustness of subsequent co-culture. On day eight, a significant amount of HSCs should be visible. If the apparent numbers of HSCs is low, then an extra day of co-culture (day 8+1) is introduced to increase these numbers prior to HSC harvest and transfer. These adjustments allowed for robust differentiation of ESCs transfected with our reporter constructs. Despite higher proliferation, we confirmed that non-transfected ESC, display the same developmental kinetics as the transfected ESC under these adjusted co-culture conditions.

Chapter 4: An *in vitro* Model to Study the Activity of TCRα LCR

4.1 Applying *in vitro* Embryonic Stem Cell Differentiation to the Study of the TCRα LCR

Virtually the entire course of T cell development in the thymus can be modeled in the ESC OP9-DL1 co-culture system, with each developmental stage readily distinguishable by multi-parameter flow cytometry. Multiple LCRs have been discovered in gene loci expressed at varying stages of T cell development and function. Thus, we believed this system offered the opportunity to model the activity of LCRs that function in T-lineage cells after their *in vitro* differentiation from reporter gene transfected ESCs. To investigate if the *in vitro* ESC differentiation system will support the same properties of the TCRα LCR seen in the transgenic mouse models, we generated ESCs clones transfected with either hCD2ΔT or hCD2:1-8. The ESC clones that were transfected with the hCD2ΔT transgene did not express hCD2 proteins in DP T cells **(Figure 7A)**. In contrast, all of the clones that were transfected with the hCD2:1-8 transgene produced hCD2 positive DP T-cells **(Figure 7B)**. Taken together, these two observations showed that the expression of the hCD2:1-8 reporter gene construct is not dependent on the transgene site of integration.


A
ΔTA3
DP
CD8 SP
ΔTA4
ΔTB4
CD8
CD4
%MAX
hCD2

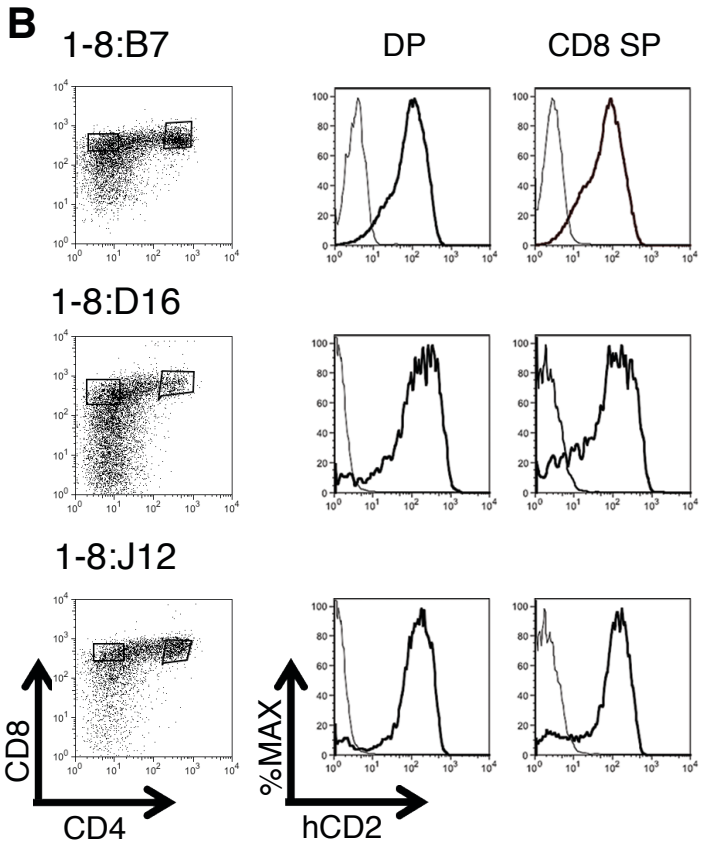

B
1-8:B7
DP
CD8 SP
1-8:D16
1-8:J12
CD8
CD4
%MAX
hCD2

FIGURE 7 TCRα LCR-dependent hCD2 reporter protein expression in T cells differentiated from ESC clones.

(A) Flow cytometry analyses of hCD2 expression in three representative independent clones of ESC transfected with hCD2DT (ΔT) and subsequently differentiated into CD4-CD8 DP and CD8 SP T cells. None of the clones produce T cells that are hCD2 positive (n = 6). (B) DP and CD8 SP T cells derived from three representative independent ESC clones transfected with hCD2:1-8 are positive for hCD2 reporter protein. All hCD2:1-8 ESC clones (1-8) produce T cells that express the hCD2 protein on their cell surface (n>12). Cell population gates are shown at left. hCD2 expression in transfected (dark curve) and non-transfected (light curve) gated cells is shown at right[76]. *Copyright 2013. The American Association of Immunologists, Inc.*

4.2 TCRα LCR Does Not Drive Consistent, High Level hCD2:1-8 Transgene Expression in Non-T Lineage Cell Types Derived *in vitro*.

In transgenic mouse models, a reporter gene linked to the TCRα LCR is highly expressed only in lymphoid tissues. To determine the cell type restriction on hCD2:1-8 reporter gene expression, we generated hCD2:1-8 transfected ESC clones into various hematopoietic cell types. Analysis of $TER119^{+}$ cells (erythroid lineage) or $CD11b^{hi}$ cells (monocytic lineage) revealed very little to no expression of hCD2 **(Figure 8)**. Five ESC clones were differentiated into B cells. Three of the five clones expressed hCD2 transgene at variable levels. However, similar observations were made in transgenic mice bearing the same hCD2:1-8 transgenic construct.

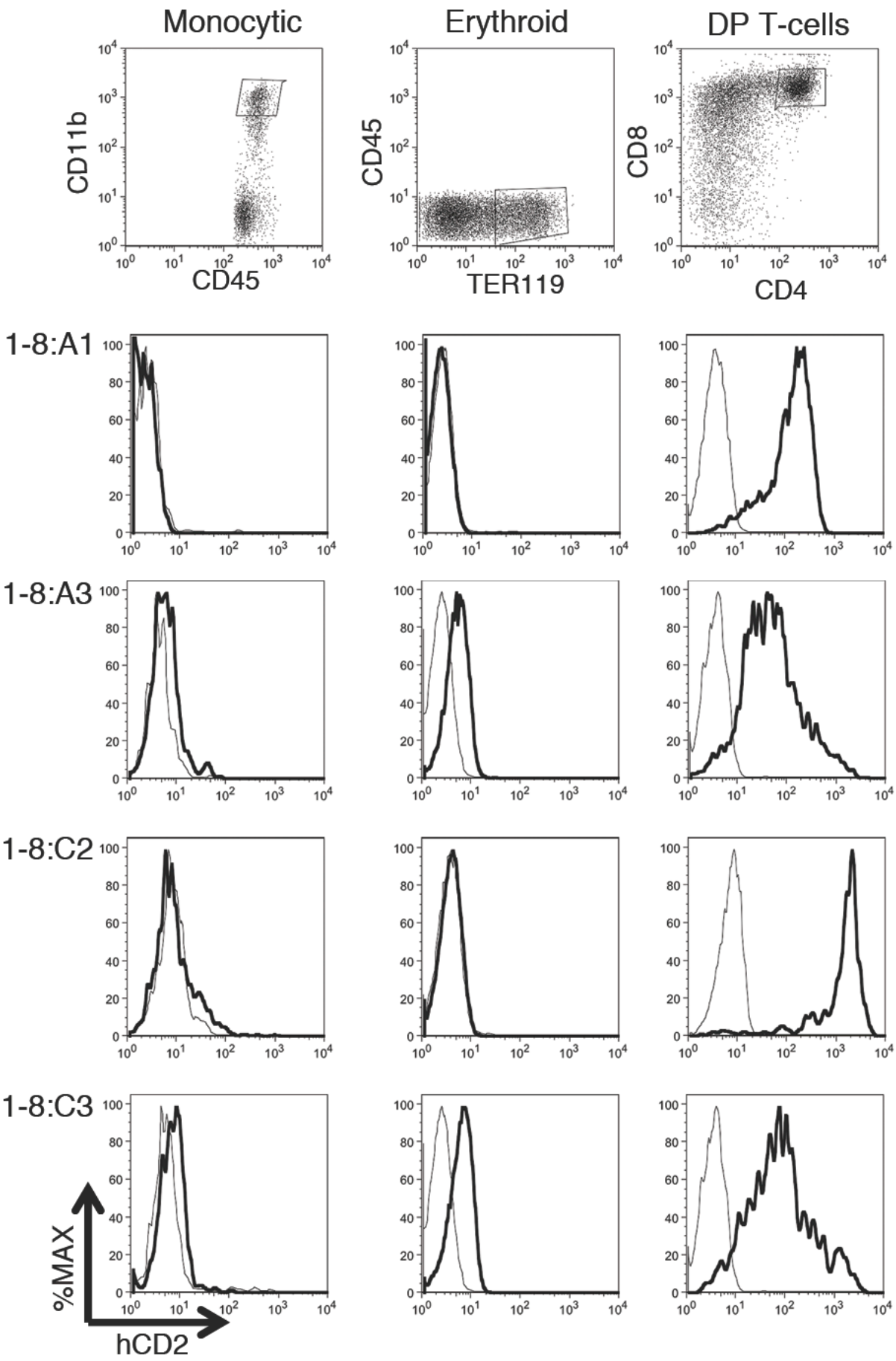

Monocytic
Erythroid
DP T-cells
CD11b
CD45
CD45
TER119
CD8
CD4
1-8:A1
1-8:A3
1-8:C2
1-8:C3
%MAX
hCD2

FIGURE 8 Cell-type restriction on high-level hCD2:1-8 reporter transgene expression during *in vitro* hematopoiesis from ESCs.

Flow cytometric analysis of *in vitro* differentiated hematopoietic progeny of four representative hCD2:1-8ESC clones (1-8). Monocytic ($CD11b^{hi}CD45^{+}$) and Erythroid ($TER119^{+}CD45^{-}$) cells were harvested on day 12 of OP9 plus ESC co-culture and were low to negative for hCD2. DP T cells ($CD4^{+}$ $CD8^{+}$) were harvested on day 18 of OP9DL1 plus ESC co-culture and were strongly positive for hCD2. Representative target cell population gates are shown at top. hCD2 expression in gated transfected (dark curve) and non-transfected (light curve) cells is shown below in each column[76]. *Copyright 2013. The American Association of Immunologists, Inc.*

4.3 TCRα LCR linked reporter gene is expressed with endogenous TCRα gene-like kinetics during T cell development *in vitro*.

In the ESC-OP9 co-culture system we can track T cell differentiation through their various stages of development via flow cytometry. The onset of hCD2 expression under the regulation of the TCRα LCR in the differentiated ESC clones is at the DN3 to DP stage transition **(Figure 9)**. These kinetics are similar to that of endogenous TCRα gene and transgenic mice bearing the same hCD2:1-8 construct[72].

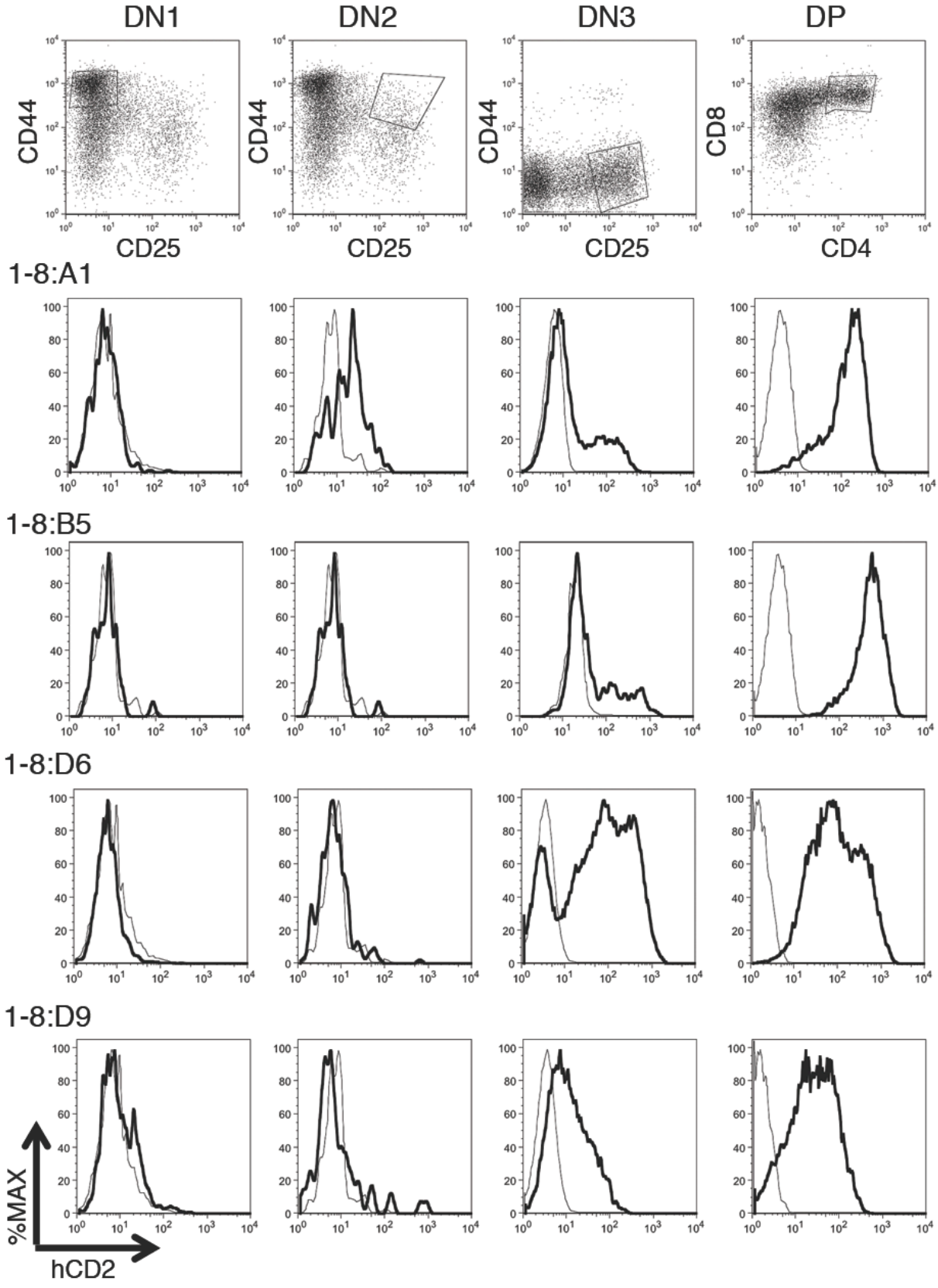

DN1
DN2
DN3
DP
CD44
CD25
CD8
CD4
1-8:A1
1-8:B5
1-8:D6
1-8:D9
%MAX
hCD2

FIGURE 9 Appropriate upregulation of the hCD2:1-8 reporter gene at DN3 or during the DN3 to DP stage transition of *in vitro* T cell development.

Flow cytometric analysis of T cell development from four representative hCD2:1-8ESC clones (1-8) differentiated in the OP9DL1 co-culture system. Cells were harvested on day 12 of co-culture to detect DN1 ($CD44^+CD25^-$) and DN2 ($CD44^+CD25^+$) stage T cells, or day 18 to examine DN3 ($CD44^-CD25^+$) and DP ($CD4^+CD8^+$) cells. Note that clone 1-8:A1 also appears in Fig. 8. Representative target cell population gates are shown at top. hCD2 expression in gated transfected (dark curve) and non-transfected (light curve) cells is shown below in each column[76]. *Copyright 2013. The American Association of Immunologists, Inc.*

4.4 hCD2:1-8 transgene linked to TCRα LCR is expressed in a copy number dependent manner by T cells derived *in vitro*.

The final hallmark of TCRα LCR activity is the ability to drive the mRNA expression levels of a linked transgene in a copy number dependent manner. To test for copy number dependence in this model we extracted RNA from day 20 ESC-OP9-DL1 co-cultures. On this day, the co-culture consists predominantly of DP and CD8 SP T cells. This RNA was analyzed for expression of hCD2 via real-time qRT-PCR. To normalize for the heterogeneous co-culture we normalized with TCRα primers. Both representative groups of independently derived clones that were tested expressed the hCD2 reporter gene in a narrow 1.6 fold range with a strong and significant correlation, which is indicative of full LCR activity[34,77] (**Figure 10**).

These data indicate that this *in vitro* differentiation system supports the properties that the TCRα LCR confers on a linked transgene in a transgenic mouse model.

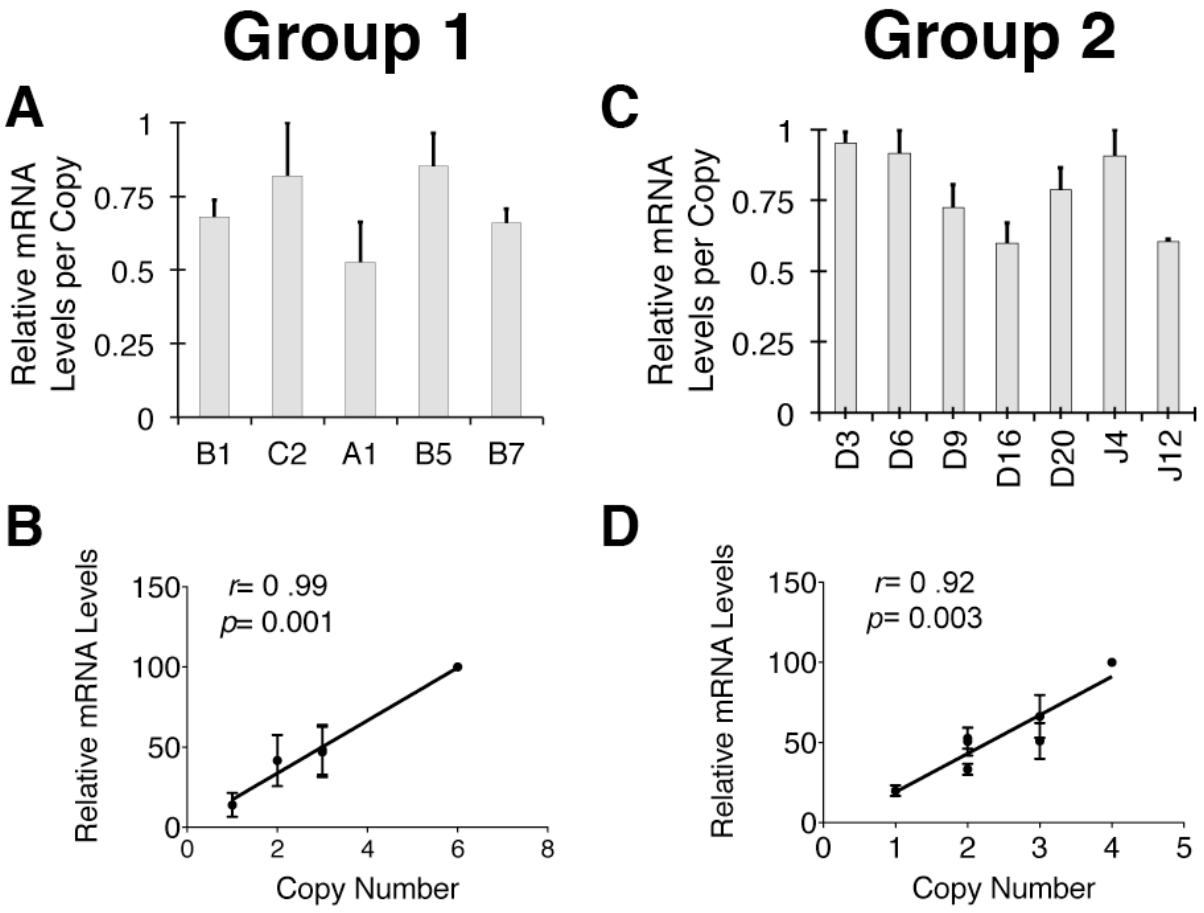


FIGURE 10 TCRαLCR drives copy number–related hCD2 reporter mRNA levels in T cells derived *in vitro* from ESCs.

qRT-PCR analysis of *in vitro* differentiated T cell progeny of representative hCD2:1-8 ESC clones. Cells were harvested on day 20 of OP9DL1-ESC co-culture. (A) qRT-PCR results from T cells derived from five individual, independent hCD2:1-8 ESC clones generated from an ESC transfection (Group 1). Copy number estimates are (left to right) 3, 6, 1, 2, and 3. Individual samples were run in triplicate in the qRT-PCR experiments. The y-axis signifies the relative mRNA levels detected in a given experiment with the highest level observed designated as 1.0. Bars represent averages of three independent experiments (see Materials and Methods). The range of mRNA levels per transgene copy is 1.6-fold. (B) Graph of the correlation between relative mRNA level and transgene copy number. Prism (Graphpad) software was used to calculate x-y value correlation (r) and significance (p) noted on the graphs. The p value was derived from an F test on the linear relationship between x and y values. (C) qRT-PCR results (analyzed and depicted as in A) from T cells derived from seven additional individual, independent hCD2:1-8 ESC clones generated from an additional, independent ESC transfection (Group 2). Copy number estimates are (left to right) 2, 4, 1, 3, 3, 2, and 2. The range of mRNA levels per transgene copy is 1.6-fold. (D) Graph of correlation between relative mRNA level and transgene copy number (analyzed and depicted as in B)[76]. *Copyright 2013. The American Association of Immunologists, Inc.*

4.5 Incomplete TCRα LCR activity after *de novo* introduction into lineage-committed T cell lines

The *in vitro* differentiation model was the first cell culture model to fully support LCR activity. Previous studies suggested that the LCR DNA elements would be required to be present in a cell before its lineage commitment. We tested this hypothesis by introducing the aforementioned hCD2:1-8 constructs into two different T cell lines. The VL3-3M2 is a DP stage T cell line[78] whereas the C6VLB cell line is SP CD8 T cells[79].

qRT-PCR data measuring the hCD2 mRNA expression levels per copy of clones generated from the VL3 or C6VLB cell lines is shown in (**Figure 11)**. This data represents, in triplicates, the mRNA levels of the hCD2 reporter gene, per intact copy, over three independent experiments. The expression levels of the reporter mRNA seen in the VL3 and C6VLB cell lines were both outside of the acceptable range of copy number dependence for a transgene linked to an LCR (VL3-3M2 were in a 5.5 fold range and C6VLB clones in a 83.5 fold range)[34,77]. Of special note is the transgene silencing observed in the single copy-bearing C6VLB clone CL6. Two single copy clones generated from ESCs, A1 and D9, expressed the hCD2 reporter gene normally. Single copy transgenes are known to have a difficult time overcoming position effects[77]. It is also important to note, that even without the inclusion of CL6 the range of hCD2 mRNA expression within the C6VLB clones is 5 fold, which is still outside of full LCR activity This, however, is indicative of partial LCR activity[34,77].

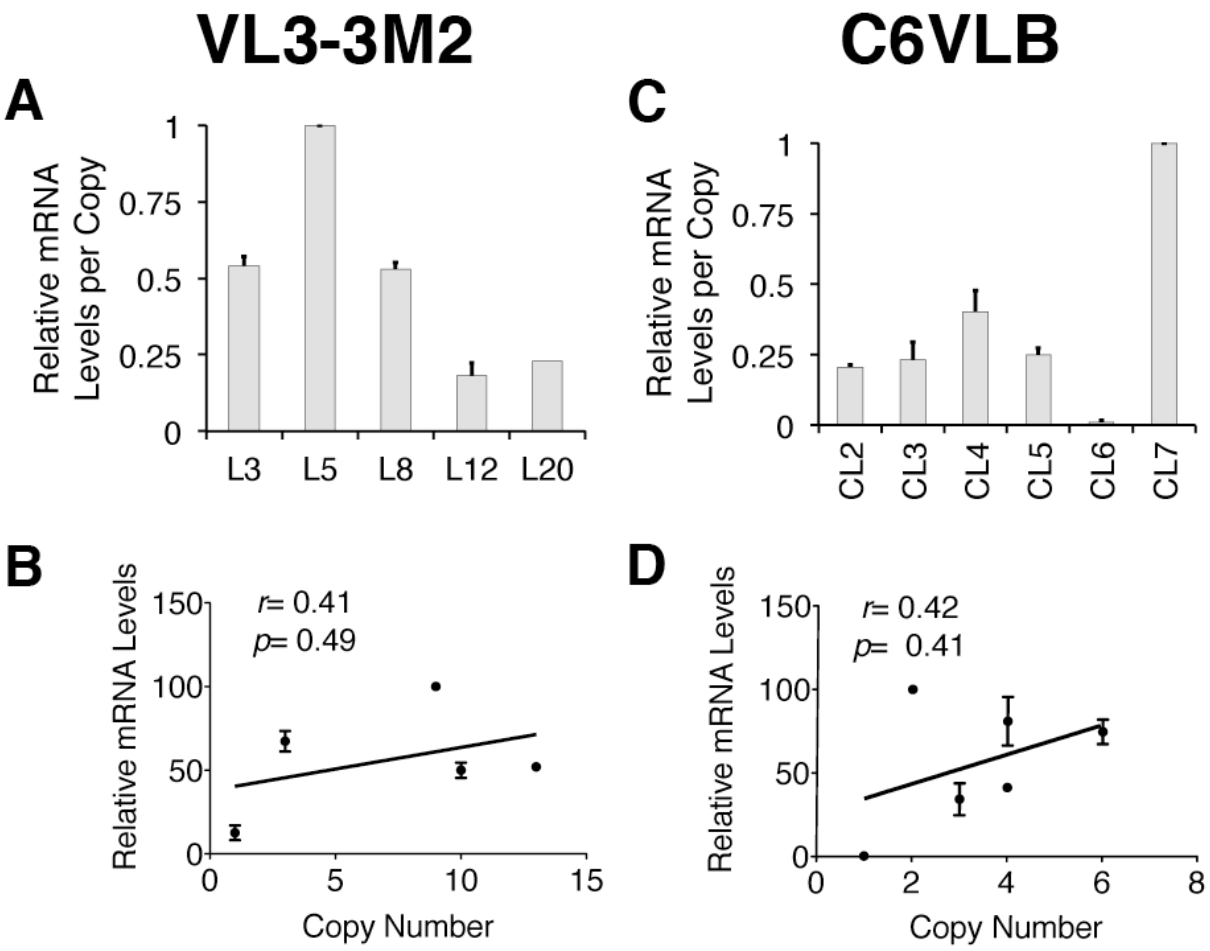


FIGURE 11 TCRα LCR-driven reporter mRNA levels are not copy number–related after *de novo* transfection into established T cell lines.

(A) qRT-PCR analysis of clones of a CD4, CD8 DP T cell line, VL3-3M2, transfected with the hCD2:1-8 reporter gene. Copy number estimates are (left to right) 8, 3, 1, 13, and 10. Individual samples were run in triplicate in the qRT-PCR experiments. The y-axis shows relative mRNA levels detected in a given experiment with the highest level observed designated as 1.0. Bars represent averages of three independent experiments (see Materials and Methods). The range of mRNA levels per copy is 5.5-fold. (B) Graph of the correlation between relative mRNA level and transgene copy number (analyzed and depicted as in Fig. 10B, 10D). (C) qRT-PCR results (analyzed and depicted as in A) of clones of C6VLB (a CD4+ T cell line) transfected with the hCD2:1-8 reporter gene. Copy number estimates are (left to right) 4, 3, 4, 6, 1, and 2. The range of mRNA levels per transgene copy is greater than 80-fold. (D) Graph of the correlation be- tween relative mRNA level and transgene copy number (analyzed and depicted as in Fig. 10B, 10D)[76]. *Copyright 2013. The American Association of Immunologists, Inc.*

4.6 Significance

The predominant model used to study LCRs has been transgenic mice. LCRs can consist of many hypersensitive sites, each of which may contain unique functions. Furthermore, these HS can cooperate to contribute to the overall properties of an LCR. To study all these intricacies would require the generation of many different transgenic mice, which is both costly and time consuming. To circumvent these issues we and others have attempted to establish cell culture models for the study of LCRs. Initial attempts in trying to establish a cell culture model using the β-globin LCR revealed the possible requirements for this LCR to be present in an uncommitted environment before it can establish its complete activity upon cellular differentiation[34,39,40,77]. Our work has also supported this theory. We have shown that the *in vitro* ESC differentiation model fully supports TCRα LCR activity. The hCD2 reporter expression in this model resembles what we observe in the transgenic mouse, including tissue specificity, copy number dependence and developmental timing. This is in contrast with the same reporter gene being introduced into two different committed T cell lines. The expression of hCD2 in these two models did not correlate with number of integrated reporter constructs. Together, these two results give further credence to the necessity for the presence of the LCR prior to lineage commitment. This hypothesis is supported by recent findings that "pioneering" factors may prime a gene's locus in pluripotent cells for correct expression and regulation later in the committed cell[80,81].

The ability of the *in vitro* ESC differentiation model to fully support TCRα LCR activity is extremely beneficial to the field. This model can be used to rapidly test expression of a reporter gene linked to modified versions of the TCRα LCR, in order to further characterize the functional regions of the LCR. Another advantage of the *in vitro* differentiation model is its ability to give rise to various hematopoietic cell types. There are multiple active LCRs in the cells of the hematopoietic system that can be easily studied using this model. With this knowledge, it would be interesting to investigate the possibility of using an amalgam of different DNA regulatory elements from other genes/LCRs and construct a "synthetic LCR". In principle, such a synthetic LCR can be designed to support the tissue specificity and developmental timing characteristics of one's choosing. This model will now make these types of exploratory studies more feasible.

Chapter 5: TCRα LCR Can Potentially Improve Gene Therapy Vectors

5.1 Background

We now have established an *in vitro* model for quick screening experiments aimed to study the activity of TCRα LCR[76]. We can use this model, for instance, to test expression of a reporter gene linked to miniaturized versions of the TCR α LCR containing only characterized functional regions. Mini LCRs that can fully recapitulate the activity of the full length LCR, as seen in the ESC *in vitro* differentiation model, can potentially be included in gene therapy vectors.

One current approach to treating patients with mutations in genes expressed in T cells is to introduce a therapeutic gene into patient-derived T cells and reintroduce these cells back into the patient[82]. However, this approach has certain limitations. First, the probability of the therapeutic gene integrating in a euchromatic region proximal to an active gene is low. Second, integration within these area may interfere the regulation of the endogenous gene that can lead to cell death or cancer[83-86]. Finally, given the likely finite life span of virally transduced *ex vivo* T cells, these patients may have to repeat the gene therapy process indefinitely.

Long term HSCs are self-renewing cells that can be driven to differentiate in to multiple lineages, including to T cells. Therefore, HSCs represent an ideal candidate for introducing a therapeutic gene for T cell gene deficiencies. A drawback to this approach is that the therapeutic gene can be ectopically expressed in any or all of the types of HSC progeny and

can be expressed at an inappropriate stage of development. For example, the premature expression of TCRαβ on T cells can lead to premature cell expansion[87]. By linking the therapeutic gene to a T cell specific LCR, we can circumvent these potential limitations by directing the expression of the therapeutic gene at physiological levels in T cells. In addition, this therapeutic gene should be expressed in a similar spatiotemporal pattern as *TCRα*.

Introduction of a transgene into HSCs is not straightforward. They are not easily transfectable since they do not divide often[82]. Therefore, to introduce a transgene into HSCs, a common method is viral transduction using lentiviruses[82,88,89]. Lentiviruses, such as Human Immunodeficiency Virus (HIV-1), can infect non-dividing or slowly dividing cells and have much higher exogenous DNA capacity than other types of viruses. These characteristics make lentiviruses the preferred viral backbone in which to incorporate the TCRα LCR in gene therapy vectors.

5.2 Establishing Minimal TCRα DNA Sequence For Full Activity

Although lentiviral vectors have a high capacity for exogenous DNA, it is still beneficial to integrate the smallest amounts of DNA as possible in order to limit any interference with viral packaging[90-92]. Researchers utilizing the β-globin LCR in the context of gene therapy have taken a similar approach. The β-globin LCR spans 20kb of DNA in its locus[11,23]. Through the identification of the functional regions, researchers were able to determine a 1kb core consisting of three HS[93]. However, upon further research, it was determined that this functional core was not sufficient for consistent high-level β-globin RNA

expression[93,94]. Inclusion of additional flanking DNA from each HS markedly improved LCR activity and expression of β-globin RNA[95].

Therefore, we took a similar approach as did researchers studying the β-globin LCR, by designing vectors to ascertain the minimal amount of LCR DNA sequence that can still confer its full activity. In previously published work, we identified four functional HS regions containing at least five distinct elements. We used this information to propose three different mini versions of TCRα LCR, consisting of varying lengths and combinations of LCR functional components (**Figure 12**).

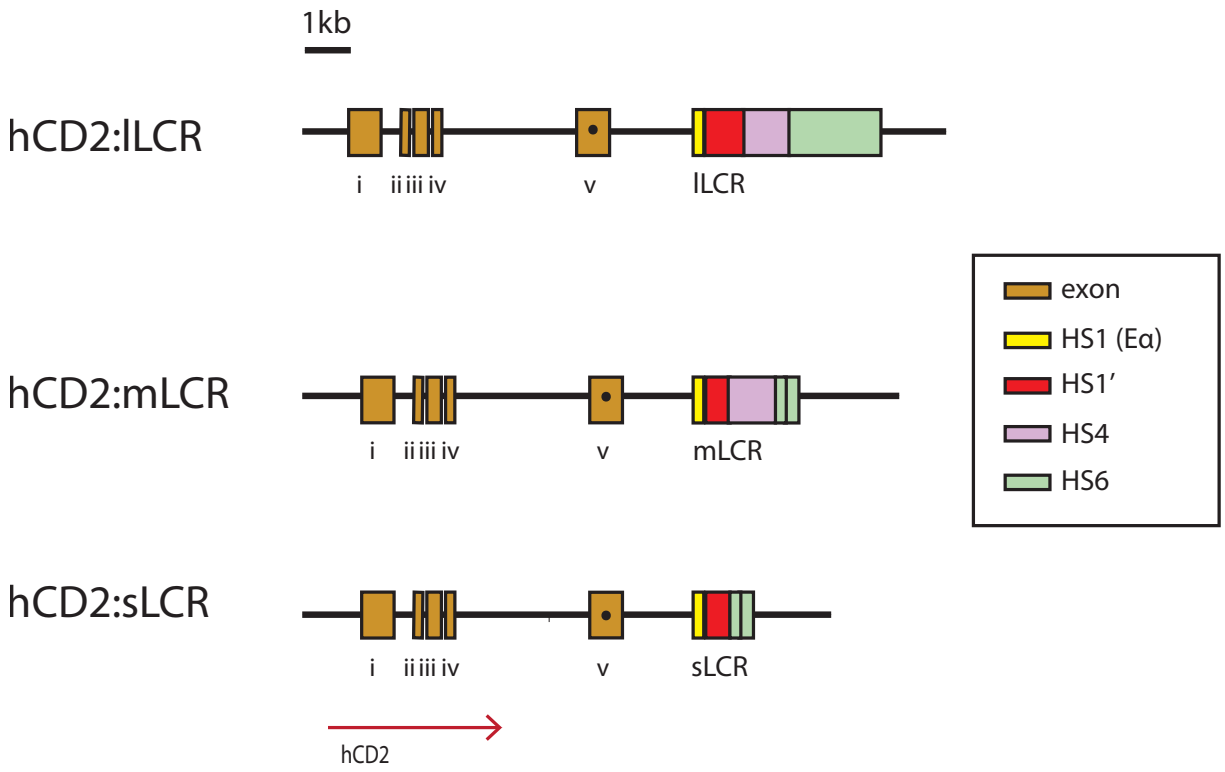


FIGURE 12 Mini LCR-linked reporter transgenes.

Depiction of transgenes containing the hCD2 reporter gene linked to either the large, medium or small TCRα LCR. These constructs contain the same hCD2 reporter as described in Figure 6. The mini LCR is located at the 3' end of each construct and the figure key indicates the color corresponding each HS. The large LCR (lLCR) contains the complete regions of HS1, HS1', HS4 and HS6. The medium (mLCR) and small (sLCR) LCRs contain, HS1, the identified functional regions of HS1', and TF-1-2-3 and HS6-316 regions of HS6, in order. The mLCR also contains the complete HS4.

The first of the three mini LCR constructs created, Large LCR (lLCR), includes HS1 and 1' in addition to HS4 and HS6, which would be 4kb in length. However, we also proposed two other versions. The activity of HS6 so far has been mapped to two regions: a 238bp factor binding region (called TF123) and a 316bp region at the 3' end. In addition, the activity of

HS1 and 1' can be mapped to a 744bp region. Therefore, we can further reduce the size of the mini LCRs by including only these core regions of HS 1, 1', 4 and 6, to make a construct that is 2.4kb in length (medium LCR (mLCR)). Finally, we proposed the smallest mini LCR, which contains the functional regions of HS1, 1' and 6, but does not include HS4 (small LCR (sLCR)). Although the deletion of HS4 from the LCR led to low levels of reporter gene expression in mice, the aim of this construct was to test if the combination of the other three elements might yield sufficient activity in gene therapy models. To test the activity of these mini LCR constructions, we linked them to the previously described hCD2 reporter.

We have generated stable transfected ESC clones, with either the large, medium or small mini LCRs and determined copy numbers and intactness using Southern blots. We will compare the reporter mRNA expression levels driven by these mini LCR vectors to those observed from full length LCR containing reporter constructs in T cells derived *in vitro*, as described in Chapters 3 and 4. Preliminary data indicates that the largest mini LCR maintains high-level reporter expression that is restricted to T cells **(Figure 13).** These clones are now being differentiated and analyzed further by J. Lovett as part of her thesis project.

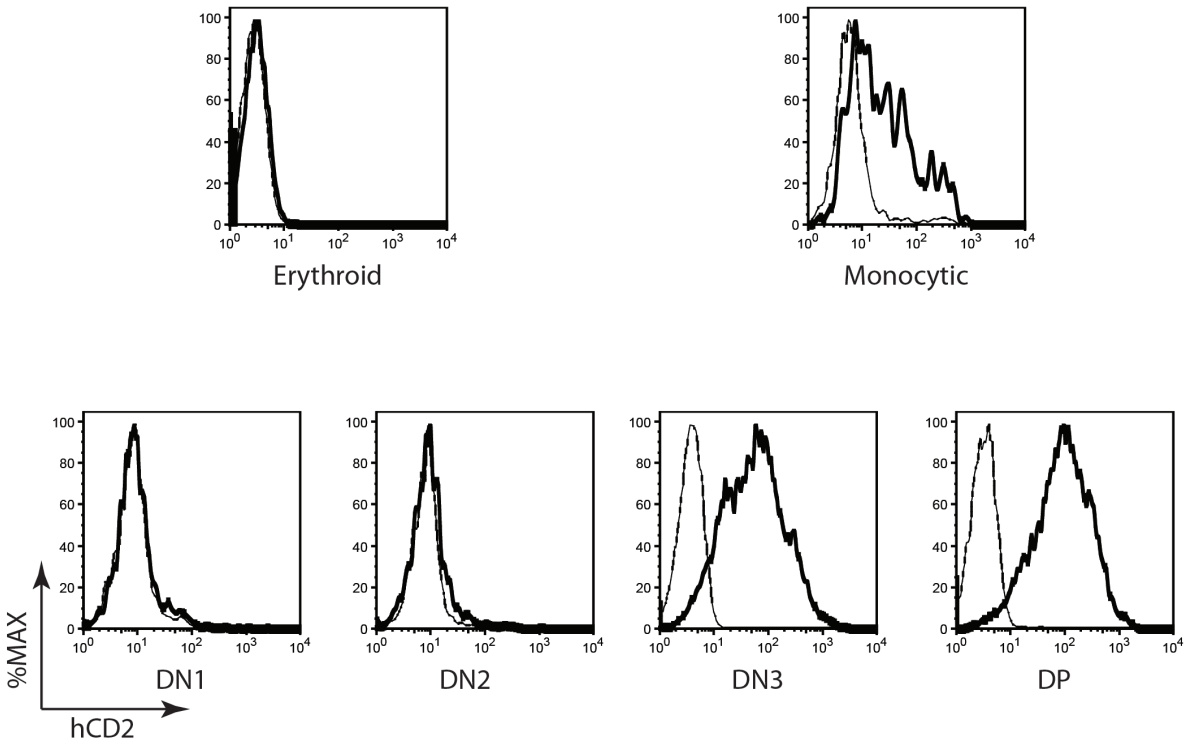


FIGURE 13 Analysis of a single ESC clone transfected with hCD2:lLCR transgene.

An ESC clone, L1, transfected cells were harvested on day 12 of co-culture to detect DN1 (CD44+CD25-) and DN2 (CD44+CD25+) stage T cells, or day 20 to examine DN3 (CD44-CD25+) and DP (CD4+CD8+) cells. Erythroid cells were negative for hCD2 expression. Minimal expression of hCD2 was detected in monocytic cells. Expression of the hCD2 protein in the L1 clone was upregulated at DN3 and persisted in DP T cells. hCD2 expression in gated transfected (dark curve) and non-transfected (light curve) cells is shown.

5.4 HIV-1 Vector Plasmid Analysis

As stated previously, one of the goals of our lab is to initiate studies to assess the potential for harnessing the properties of the TCRα LCR to improve upon the efficacy and expression of gene therapy vectors. This would first require the investigation into whether the TCRα LCR will interfere with the production of virion. As a first step in accomplishing this, we made three different constructs containing one of our previously described mini LCRs linked to a reporter gene. The viral components of these constructs include a transduction marker, a central polypruine tract (cPPT), and a woodchuck hepatitis virus post-transcriptional regulatory element (WPRE). cPPT increases transduction efficiency[96,97] and the WPRE is an element used for increased transgene expression[98,99]. Flanking this viral construct are *cis* acting regulatory elements called long terminal repeats (LTR), which regulate the expression of the gene and interact with host machinery[100].

We modified a viral vector, containing a viral transduction marker, a modified non-signaling Low-Affinity Nerve Growth Factor Receptor (LNGFR) protein called NTP[101], driven by the human Phosphoglycerate Kinase 1 (hPGK) promoter. A Vα17 driven YFP reporter gene linked to either the sLCR, mLCR, or lLCR was placed 5' to the hPGK promoter **(Figure 14)**. This allowed us to confirm whether or not the TCRα LCR components would interfere with the function of other components within this transgene. C6VLB cells were transfected, stable clones were generated for each construct, and analyzed via flow cytometry for the expression of the NTP transduction marker and YFP reporter gene. Results indicated that both the reporter gene and transduction markers are expressed on cells. This confirmed that the TCRα LCR components do not interfere with the expression of

NTP **(Figure 15)**. This, however, only confirms that these components do not interfere with the transcription/translation of each of the reporter genes. We still need to investigate the possibility of TCRα LCR interfering with the packaging or the transduction of the virus.

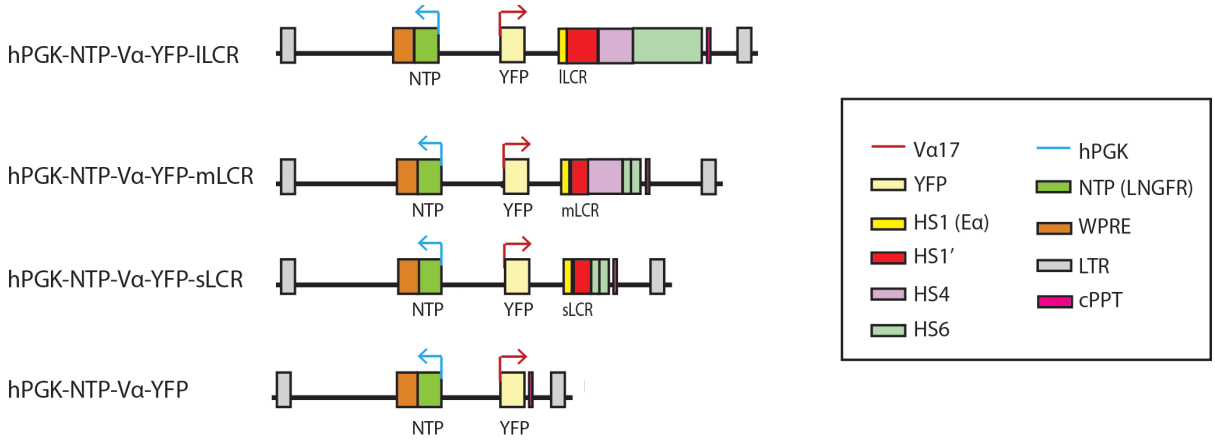


FIGURE 14 Mini LCR reporter and viral transduction vector transgenes.

Depiction of transgenes containing the YFP reporter gene driven by a Vα17 promoter linked to either the large, medium, small, or no TCRα LCR. Figure key indicates the color corresponding to each element. These constructs also contain a viral transduction marker (NTP), cPPT (to increase transduction efficiency), and a WPRE element (to increase expression of transduction marker). The mini LCR colors are depicted in the figure legend.

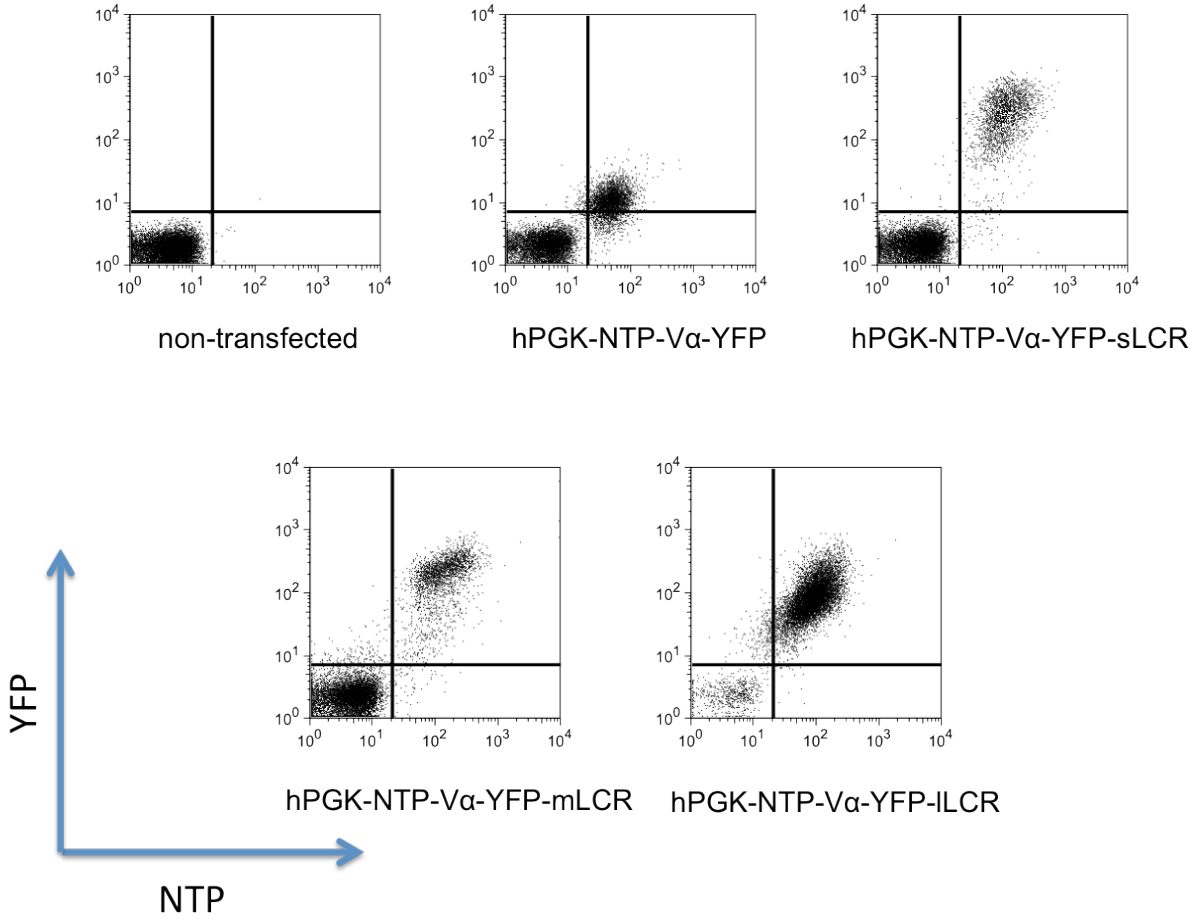


FIGURE 15 The TCRα mini LCRs do not interfere with the expression of the viral transduction marker NTP (LNGFR).

Representative C6VLB clones transfected with one of the constructs in Figure 14. Both the viral transduction marker (NTP) and YFP reporter genes were expressed. Moreover, both of these were still expressed when the YFP reporter gene was linked to one of the mini LCRs.

5.5 Significance

One of the current treatment options for patients with gene product deficiency disorders, such as β-thalassemia[95], sickle cell anemia[102] and SCID X1[103], is to introduce a therapeutic gene via viral vectors into patient-derived cells. Through this gene therapy, one can harvest a patient's cells, transduce them with the "corrected" gene, and reintroduce these cells into the patient. One cannot, however, direct the integration of these transgenes into cells. Furthermore, the targeted patient cells can have a limited lifespan, which will necessitate multiple transduction events of patient-derived cells. Transgenes can integrate in heterochromatic or euchromatic regions. Since heterochromatin is more prevalent than euchromatin, it is more likely for a transgene to integrate into silencing regions. Using retroviral vectors, however, favors integration into active sites[104-106]. Despite the fact that these therapeutic genes might not be silenced, this can be problematic because it can interfere with the proximal genes, such as proto-oncogenes[86,107,108]

Recent studies have addressed some of the aforementioned limitations. The therapeutic gene can be targeted to multipotent adult stem cells, such as self-renewing HSCs. These cells provide a continuous source of terminally differentiated cells, which can express the therapeutic gene product. These differentiated cells can span multiple lineages. If one wishes to direct the expression of the gene product, however, into a single lineage, tissue specific promoters or enhancers can be used. Linking the therapeutic gene just to promoters and enhancers, though, can render it susceptible to position effects or integration at sites proximal to active genes.

Locus Control Regions can modulate the expression of a linked transgene to produce physiological expression levels and patterns that mirror the LCR's locus of origin. These properties can possibly be exploited to improve on current gene therapy vectors.

Retroviral vectors have been used to deliver a β-globin gene linked to portions of its LCR into patients with β-thalassemia. Although HS2-4 of the β-globin LCR do not confer full LCR activity[4], they do drive expression of the β-globin gene sufficiently for clinical applications[95]. Furthermore, clones with favorable integrations sites can be selected. These "safe harbor" sites, briefly, are distant from transcription sites and regulatory regions, such that the integrated transgenes are less prone to interfere with these regions[84].

Applying the knowledge from the β-globin LCR to other LCRs, including the TCRα LCR can be pivotal to the field of gene therapy. We have taken a similar approach in designing our mini LCRs, which include the functional HS. Data from our mini LCR constructs will serve two purposes: from a basic science point of view, it will allow us to determine if included functional regions are sufficient in driving the linked reporter gene in the correct spatiotemporal and copy number dependent manner, irrespective of the site of integration. From a translational perspective, the studies will serve to ascertain the capability of the mini LCRs to generate safer and more efficacious gene therapy vectors. If successful, it would fortify the use of TCRα LCR and other LCRs in improving the field of gene therapy vectors.

Chapter 6: The TCRα LCR Displays Barrier Insulator-Like Activity

6.1 The Possible Existence of Barrier Insulator Activity in the TCRα LCR

Chapter 4 described differentiated T cell lines (VL3-3M2 and C6VLB) transfected with the hCD2 reporter gene linked to the TCRα LCR. The clones from this experiment did not express hCD2 mRNA levels in a copy number dependent manner. In spite of these results, nearly all the clones still expressed the hCD2 reporter gene. These observations indicated that these two T cell lines at least partially support the TCRα LCR's ability to inhibit position effects. Position effects can be manifested as the complete silencing of a transgene due to its integration in or to the spreading of a heterochromatic region in the genome. One of the key properties of an LCR is its ability to suppress position effects. Though the mechanism by which this is accomplished remains unknown we have hypothesized that this suppression can be mediated in part by barrier insulator-like elements present within the LCR. Barrier insulators protect a transgene from the lateral spreading of surrounding heterochromatic regions.

6.2 Barrier Insulators

One mechanism for regulating gene expression in a cell is the targeted formation of heterochromatin and/or euchromatin at a gene locus. The regulation of these types of chromatin structures will be important in loci that contain differentially expressed genes, such as the *TCRα* and *Dad1* locus. Since the LCR is situated in between these two genes, it

may play a role in their differential expression via the modification of the surrounding chromatin.

The euchromatic region usually surrounding an active gene can be encroached upon by the spreading of heterochromatin of a nearby silenced gene locus. This progression of heterochromatic region into regions of active chromatin that is thought to cause position effect variegation of transgenes. One way an active gene can be protected from this phenomenon is through barrier insulators[2,109]. These elements prevent the intrusion of regions of closed chromatin by recruiting histone modifying enzyme for the unwinding of nucleosomes[110].

The first discovered barrier insulators were the specialized chromatin structures (scs and scs') in the *Drosophila* hsp70 gene locus[111-113]. These elements flank this gene, forming boundaries to inhibit the formation/encroachment of heterochromatin. In vertebrates, HS4 of the chicken β-globin (cHS4) locus has been identified as supporting a barrier insulator function[114,115]. A protein called USF1 has been implicated in this barrier activity [116,117]. USF1 contributes to the establishment of euchromatic regions through its interaction with histone methyltransferases and histone acetyltransferases.

The differential expression patterns of the genes located within the *TCRα*/*Dad1* locus brings to mind the possibility of a barrier insulator like activity existing in this locus that may partake in the regulation of their expression. This hypothesis is further supported by our observations that all but one of the previously mentioned VL3 and C6VLB clones containing hCD2 reporter gene linked to the TCRα LCR were protected from silencing.

6.3 Regions of the TCRα LCR Display Barrier Insulator-Like Activities

Barrier insulators should stop the encroachment of heterochromatic regions surrounding a gene, in turn inhibiting the silencing of a nearby gene. Including an insulator within a reporter transgene should, thereby, prevent the silencing of the reporter. To test for barrier insulator activity, we designed an *in vitro* assay in which we linked a promoter for the TCRα gene, (Vα17) to a YFP reporter gene. We then linked this reporter gene to either HS1-6 ($V\alpha_{17}$-YFP-HS1-6, fLCR), which as previously described, contains full LCR activity[27], or only HS1 that contains the Eα enhancer (Va_{17}-YFP-HS1) **(Figure 16A)**. If there exists an insulator within the HS1-6 regions of the TCRα LCR, T cell clones bearing this reporter construct should express the YFP reporter gene, regardless of the site of integration in the genome. These constructs also contained a SV40 promoter driven Neomycin-G418 resistance gene that enabled selection for stable transfectants. These constructs were transfected into the VL3-3M2 cell line[78]. Generated clones were analyzed for intact transgene integration and incorporated transgene copy number.

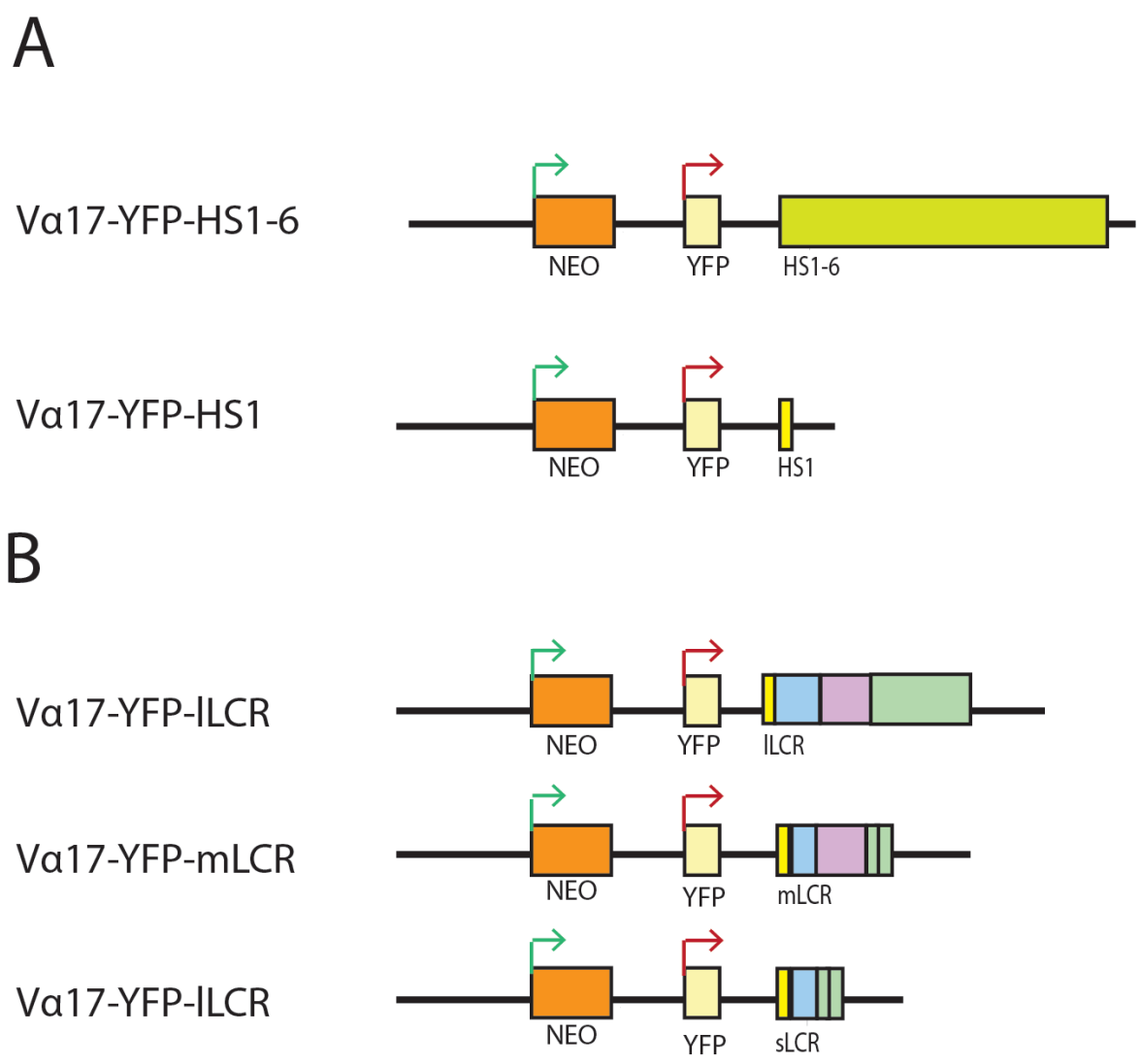


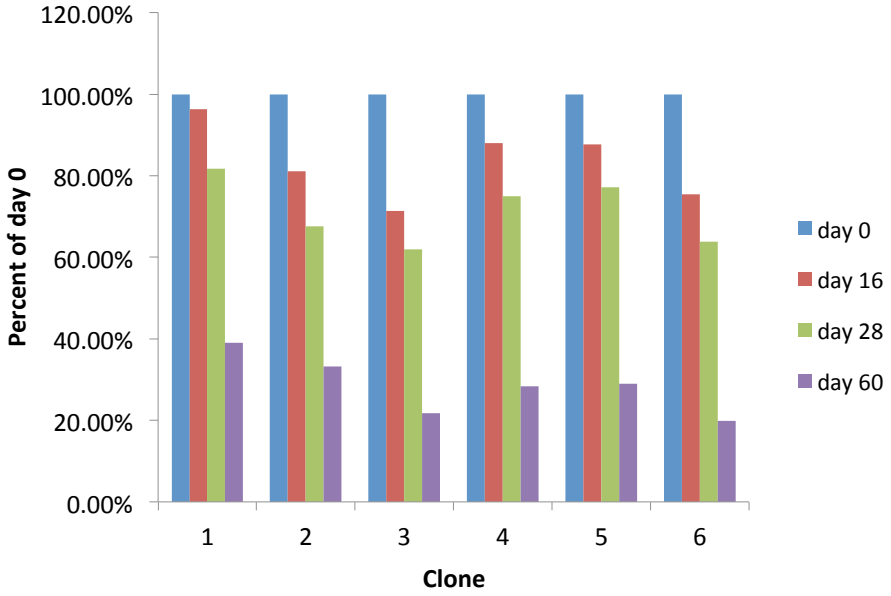
FIGURE 16 YFP reporter constructs linked to various components of the TCRα LCR used for barrier insulator assays.

(A) Depiction of transgenes containing a YFP reporter gene driven by a Vα17 promoter linked to either HS1-6 of the TCRα LCR or just to the classical enhancer, HS1. Both constructs also contain an SV40 Neomycin resistance gene. (B) Depiction of transgenes containing a YFP reporter gene driven by a Vα17 promoter linked to either large, medium, or small TCRα mini LCRs. All constructs also contain an SV40 Neomycin resistance gene.

VL3-3M2 clones transfected with Va_{17}-YFP-HS1 all expressed the YFP gene. That all clones expressed the YFP reporter gene was puzzling because one might not expect the inclusion of just the enhancer to overcome position effects. This, however, can be attributed to the proximity of the Va_{17}-YFP-HS1 to the Neomycin resistance gene. Since these clones are under selective pressure of G418, those that silence the Neomycin resistance gene would not survive. Therefore, if the YFP is silenced in these clones, then this may also silence the G418-resistance gene, thereby selecting against their survival. The removal of drug selection from these cultures would allow the clones that are silencing either gene to propagate. Conversely, if there exists an insulator within TCRα LCR then clones transfected with Va17-YFP-HS1-6 should not exhibit silencing of the YFP gene after removal of the selection drug.

After removal of G418 from the culture of clones bearing the Va17-YFP-HS1 constructs, we observed the gradual loss of YFP expression over the course of 60 days **(Figure 17A)**. In contrast, the clones bearing the Va17-YFP-HS1-6 constructs maintained the YFP expression over the same period of time **(Figure 17B)**. The maintenance of YFP expression of the Va17-YFP-HS1-6 indicated the possible existence of a barrier insulator within the HS1-6 region of the TCRα LCR. We thus sought to design an experiment that would directly test this hypothesis.

A


120.00%
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
Percent of day 0
1
2
3
4
5
6
Clone
day 0
day 16
day 28
day 60

B

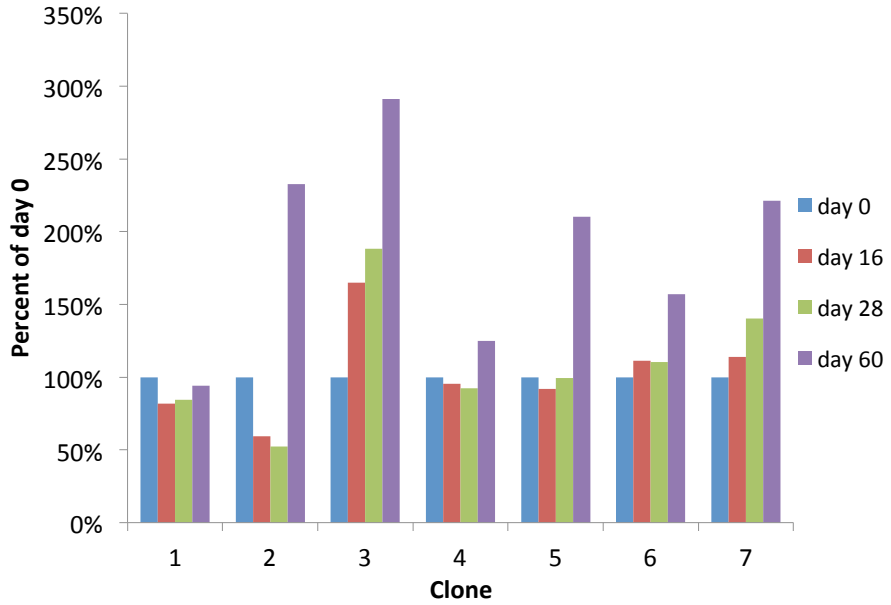

350%
300%
250%
200%
150%
100%
50%
0%
Percent of day 0
1
2
3
4
5
6
7
Clone
day 0
day 16
day 28
day 60

FIGURE 17 HS1-6 of the TCRα LCR displays barrier insulator like activity.

(A) VL3-3M2 cells were transfected with Vα17-YFP-HS1. After stable clones were generated, Neomycin drug selection was removed and YFP expression was analyzed by flow cytometry on day 0, day 16, day 28 and day 60. YFP expression in the clones was gradually lost over a 60 day time period in all six clones. This is in contrast to (B) VL3-3M2 cells transfected with Vα17-YFP-HS1-6. After removal of the Neomycin resistance gene, all of these clones gained or maintained YFP expression clones over the same 60 day time period.

6.4 An Optimized Model to Isolate Barrier Insulator Activity

To improve upon the described barrier insulator assay, we transfected VL3-3M2 cells with either Va17-YFP-HS1 or Va17-YFP-HS1-6 **(Figure 16A)** and generated a new set of clones. We then divided each clone into two groups. One set was maintained in drug selection, while the other set was maintained in drug free media. Both sets were cultured side by side. During each analysis time point, the mean fluorescence intensity (MFI) of YFP expression in a given clone propagated in the drug free media was directly compared to the same clone cultured in drug. The maintenance of YFP expression was calculated by comparing the ratio of YFP expression in drug free clones to their counterparts (maintained in selection media) over a period of 90 days.

Maintenance of YFP expression in Va17-YFP-HS1 clones was significantly diminished over this time period. This is in contrast to clones bearing the Va17-YFP-HS1-6 construct, which maintained the YFP expression over the same time. This maintenance of YFP expression was significant when compared to the transgenes with just the enhancer (**Figure 18)**.

Having identified possible barrier insulator activity within the HS1-6 region of the TCRα LCR, we attempted to narrow this region by linking the YFP reporter gene to the mini LCRs

that were described previously in Chapter 5 **(Figure 16B)**. The expression of the YFP reporter gene was maintained when linked to any of the three mini LCR constructs **(Figure 19)**. The existence of barrier insulator like activity in our construct containing just the functional regions allows us to narrow down the location of the functional regions contained in this small mini LCR construct.

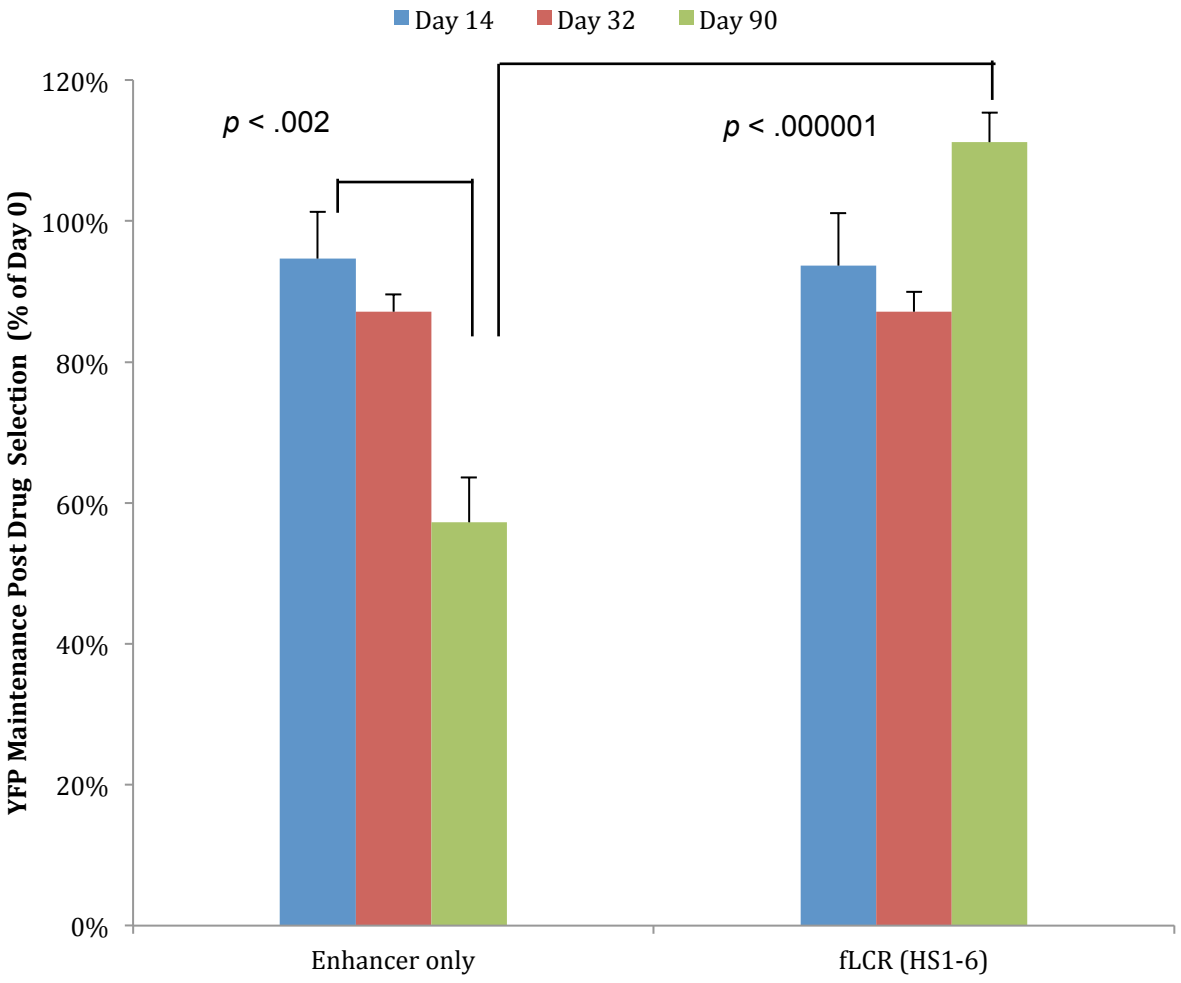


FIGURE 18 HS1-6 of the TCRα LCR displays barrier insulator like activity in a more stringent model.

VL3-3M2 clones were transfected with either Vα17-YFP-HS1 (Enhancer Only) or Vα17-YFP-HS1-6 (fLCR). After stable clones were generated, each clone was divided into two wells. One was maintained in the presence of drug (Neomycin) and the other was maintained in drug-free media for the duration of the experiment. YFP expression was analyzed by flow cytometry on days 14, day 32, and day 90. The YFP expression of each clone in the drug-free media was divided by the YFP expression in the clones that were maintained in drug to calculate the maintenance of YFP expression. Clones transfected with Vα17-YFP-HS1 showed significant gradual decrease of YFP maintenance, on average, when compared to their counterparts (the same clone maintained in drug selection media). These results are in contrast to VL3-3M2 cells transfected with Vα17-YFP-HS1-6. The Vα17-YFP-HS1-6 clones cultured in drug free media maintained or gained, on average, YFP expression over the same time period. The maintenance of YFP expression in these clones is significantly higher when compared to clones of Vα17-YFP-HS1. p values were calculated by Student's t-test.

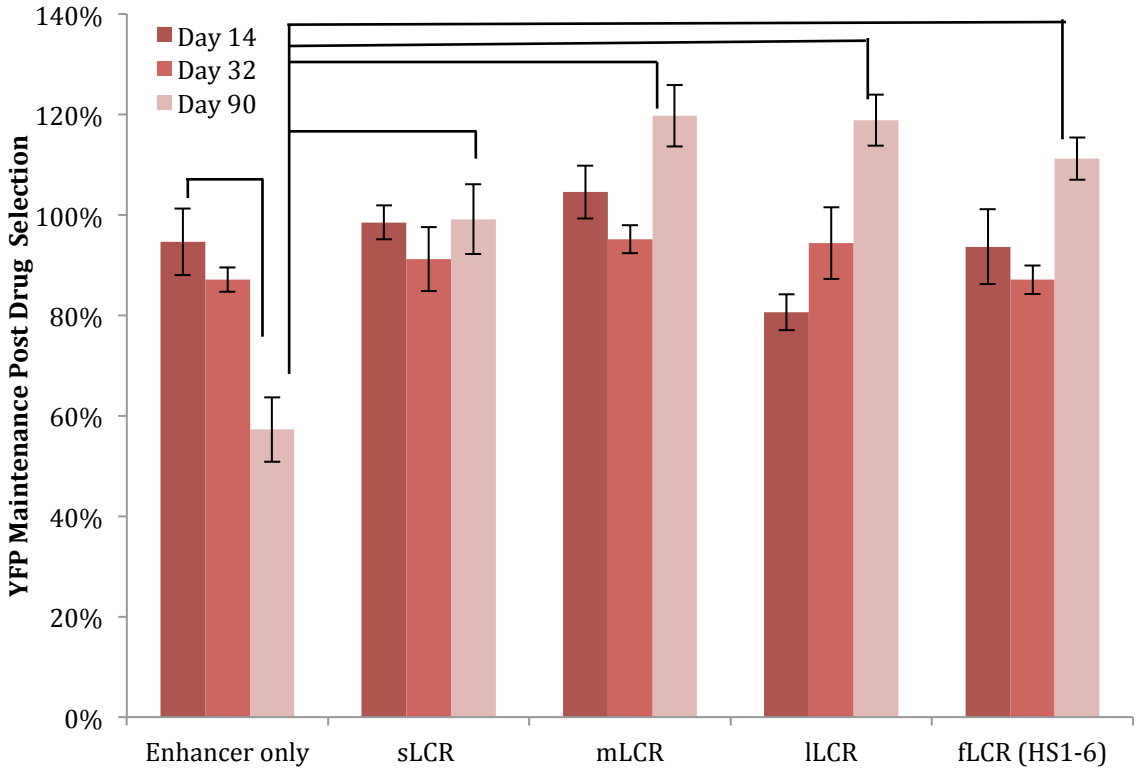


FIGURE 19 TCRα LCR's barrier insulator-like activity is located within the functional regions of HS1, 1', and 6 (sLCR).

VL3-3M2 clones transfected with Vα17-YFP linked to either the Eα enhancer, sLCR, mLCR, lLCR or fLCR of the TCRα LCR. Each clone was divided into two wells under the same conditions as described in Figure 18. YFP expression was analyzed by flow cytometry on days 14, 32, and 90. The YFP expression of each clone in the drug free media was divided by the YFP expression in clones maintained in drug to calculate the maintenance of YFP expression. All clones with the Vα17-YFP transgene linked to either the sLCR, mLCR, lLCR, or fLCR maintained or gained YFP expression, on average, over a period of 90 days. This maintenance of YFP expression was also significant when compared to clones with just Vα17-YFP-HS1. p values were calculated by Student's t-test. Note: the data from the enhancer and fLCR groups is the same as Figure 18. Bracketed bars indicate statistically significant differences (*p* value ranging from .002 to 0000001).

6.5 Significance

LCRs consist of multiple DNase I hypersensitive sites, each of which supports a distinct set of properties contributing to overall LCR function. The functional interactions of these HS can be complex and challenging to characterize. They ultimately synergize to produce the unique properties that distinguish LCR activity from that of other types of *cis*-acting elements. Investigating the individual properties of the LCR and their interactions using transgenic mice can be time consuming. Thus, the advent of new models to study the TCRα LCR, such as the described VL3-3M2 and C6VLB models as well as the aforementioned ESC *in vitro* differentiation model (Chapters 3 & 4) are beneficial to the field.

Furthermore, the data obtained from the transfected VL3-3M2 and C6VLB clones provide strong evidence of potential barrier insulator activity in the TCRα LCR. The possible existence of a barrier insulator will also further expand our knowledge of the TCRα LCR and provide additional evidence that this LCR functions by altering the state of chromatin. That TCRα is expressed only in T cells, whereas, Dad1 is expressed ubiquitously can be attributed to the possible existence of an insulator in between these two genes contributing to their differential expression.

Studies in *Drosophila* have revealed a number of insulators each interacting with their respective proteins that can contribute to their activity. However, there is a paucity of known vertebrate insulators. The finding of another such element will be a significant contribution to the field and will allow for further research into the mechanisms of insulators and the proteins involved in their function. To date, the best-characterized vertebrate insulator has been the chicken β-globin locus derived cHS4 element and it

requires the USF and VEZF1 proteins for its function[118]. USF and VEZF1 mediate the modification of histones and CpG methylation respectively[109,117,118]. USF proteins have also been implicated in barrier like activity within the human ankyrin and α-spectrin genes[109,119,120]. It would be interesting to investigate whether USF or VEZF bind to areas within the TCRα LCR. It would also be interesting to explore whether the protein repertoire associated with vertebrate insulators is limited to VEZF1 or USF. The identification and characterization of possible barrier activity within the TCRα LCR will help us expand our knowledge about the proteins involved and their possible mechanisms.

Chapter 7: Discussion

Establishing a rapid and robust *in vitro* differentiation model that can support the complex properties of the LCR seen in transgenic mice is a significant contribution to the field. Until this work, attempts to create an *in vitro* cell culture model for the study of locus control regions have not been completely successful. These previous investigations argued the necessity for an LCR-regulated transgene to be present in the genome prior to lineage commitment. The establishment of an *in vitro* differentiation system that fully supports TCRα LCR activity has provided new and clear evidence for this requirement. Further data supporting this hypothesis stems from observations that terminally differentiated cell lines bearing the same transgenes supports only partial LCR activity.

The current prevalent model used to study the TCRα LCR is the transgenic mouse. Although powerful, studying the elaborate properties supported by this LCR using this model can be time consuming. The stem cell differentiation model allows us to devise experiments that are quicker and less resource intensive. For example, work has already begun on testing three different mini TCRα LCR reporter constructs. Creating the numerous transgenic mouse lines needed to study the activity of each of these constructs could extend over a period of six months and bear significant costs. The ESC clones, however, were all generated within a month at a fraction of the cost.

There are still many questions about the TCRα LCR that remain unanswered. For example, are HS1, 1', 4 and 6 the only functional regions within this LCR? Testing the large mini LCR

is the first step in addressing the issue of whether or not the regions of HS2, 3 or 5, or the space they occupy are necessary for complete activity of the LCR.

If the large mini LCR is found to support complete TCRα LCR activity, then the results from clones bearing the small and medium LCRs will test whether we can further narrow down the activity of the TCRα LCR. Studies from the β-globin LCR revealed that the core functioning HS of that LCR were inadequate for LCR activity. In order to establish full LCR activity, these constructs had to include the sequences flanking the core HS[95]. It is possible that similar requirements apply to the TCRα LCR. If so, we would expect the small and medium mini LCRs to show deficiencies if LCR activity compared to the large mini LCRs. In that event, the core regions of HS1, 1' 4 and 6 might require their flanking sequences or yet other undiscovered functional regions within these HS.

Understanding the sequences that are required for TCRα LCR activity can allow us to harness their properties for more efficacious vectors in gene therapy. We have taken the first steps towards constructing viral vectors that include components of the TCRα LCR. Preliminary functional tests of these constructs were carried out in transfected T cell lines. These tests revealed that both the viral transduction marker gene and the LCR driven reporter gene are both functional and do not appear to interfere with each other. However, further testing of these vectors is needed. We still have yet to confirm whether or not the components of the TCRα LCR interfere with the assembly of the virus.

The target cells for gene therapy vectors are also of importance. Full function of the β-globin and now, the TCRα LCRs seem more clearly dependent on their presence in the genome before cell lineage commitment, however, at what point is unknown. Would the

introduction of the TCRα LCR into HSCs yield full LCR activity upon differentiation into T cells? The answer to this question will determine if TCRα-bearing gene therapy vectors introduced into HSCs will be successful.

Although, the TCRα LCR did not confer copy number dependence upon our reporter gene in committed T cell lines, it did display partial activity. Preliminary experiments show this activity could be barrier insulator-like. Experiments are underway to confirm this activity molecularly. There is a correlation between silencing of a gene and the modifications of histones and DNA molecules surrounding the gene, in particular the hypoacetylation of H3 and H4 histones, methylation of H3K9 and the hypermethylation of cytosine residues[121]. As gene silencing progresses, the first set of modifications are the removal of acetyl groups from H3 and H4 and the loss of tri- and di- methylation of H3K4[121]. This is followed by an upregulation of methylation of the H3K9 residue[121] and the methylation of DNA molecules at CpG sequences, which is considered to be the final step for silencing genes[121]. The lack of any CG residues surrounding our promoter might explain the absence of complete silencing of our reporter gene in this system.

If there are, indeed, barrier insulator elements within the TCRα LCR, the identification of more precise regions bearing this activity will be necessary. One clue as to the location of this element comes from the fact that some barrier insulators (such as scs, scs'[112,113] and cHS4[114,122]) have also been shown to posses the enhancer blocking functions[2,109]. These elements differ from barrier insulators, in that they are found in between an enhancer of one gene and a promoter of another gene, such that they inhibit an enhancer of one

promoter from interfering or activating another nearby promoter. The TCRα LCR contains two known regions of enhancer blocking insulators, HS1'[29] and HS2-6[123].

The novel identification of a barrier insulator within the TCRα will not only be useful from a basic science point of view, but also from a translational science perspective. Upon integration in a cell's genome, gene therapy vectors are susceptible to position effects. A transgene integrated into heterochromatin is susceptible to silencing. Insulators flanking a transgene integrated into such sites can be used to reduce position effects by inhibiting the encroachment of the surround heterochromatin[124-126]. This property makes it an intriguing candidate for incorporation into gene therapy vectors. The cHS4 insulator has been used to increase the expression of retrovirally-introduced transgenes[127]. However, this approach has met with mixed results, as there has been variable expression of these flanked transgene[127-129], which have been attributed to the cHS4 possibly being a weak barrier insulator. Thus, the finding of stronger insulators might yield more consistent results[90,127,129].

The finding of barrier insulator-like activity in the TCRα LCR, along with our work to create mini LCR constructs gives us a two-prong approach to generate safer and more effective gene therapy vectors. Introduction of TCRα LCR constructs into patient specific cells should direct TCRα-like transgene expression that is not subject to position effects. This work also opens the possibility to incorporating other LCRs in gene therapy vectors specific to patient diseases. Moreover, the finding of an alternative vertebrate barrier insulator element can also be exploited for the creation of gene therapy vectors. Together, the incorporation of these elements in gene therapy vectors, can direct expression of a linked transgene in a

spatiotemporal manner. The work presented here lays the groundwork for further research in the possibilities of designing better gene therapy vectors.

Chapter 8: Perspectives

The TCRα locus control region is a unique regulatory element, which has been classically studied in context of a transgene residing at an ectopic site in the genome. Elements of the endogenous TCRα LCR are required for the proper recombination and expression of the *TCRα* gene. A knockout of the Eα enhancer (HS1) along with HS1' resulted in failure of *TCRα* to rearrange and low level expression of TCRα[130,131]. Attempts to study the downstream regions of the LCR, HS2-6, have also been made[24]. However, the knockout of these regions also removed the third exon of the neighboring *Dad1* gene, rendering homozygous knockouts embryonic lethal. Therefore, all attempts to study the HS2-6 regions were done in heterozygous knockout mice. These mice showed no adverse affects on the rearrangement of *TCRα*. However, there was a decrease αβ expression in peripheral T cells[24].

Here, we have shown that the TCRα LCR may also possess barrier-insulator like activity, *in vitro*. The presence of a barrier-insulator in the native locus may be important for the differential expression of the *TCRα* and *Dad1* genes flanking the LCR in vivo. Dad1 is ubiquitously expressed, while the TCRα is expressed only in a subset of T cells. In non-T cells, any silencing mechanisms that may cause the restriction of the TCRα gene to T cells could inappropriately impact *Dad1* gene expression.

The existence of a barrier-insulator within the TCRα LCR might also be a distinguishing factor that elevates the function of an LCR beyond that of an enhancer. Enhancers not only function to recruit transcription factors to drive high level transcription, they have also been shown to recruit histone-modifying enzymes to allow for permissive chromatin. The

presence of an insulator, which can inhibit the encroachment of heterochromatin, goes beyond the enhancer capabilities. An LCR incorporates both of these functions, in addition to other possibly distinct regulatory activities. These can include the classical properties that are attributed to an LCR such as the spatiotemporal specificities and position independence, which can direct copy number dependence. It is the cooperation of all these properties that contribute to LCR activity and set it apart from enhancers.

We have shown that the ESC-OP9 co-culture model can support both the position independence and copy number dependence properties of the LCR, while the committed T cell line models (VL3-3M2 and C6VLB) support only the ability to overcome position dependence (**Figures 10 and 11**). These data point to the possibility that the properties of integration site independence and copy number dependence may be mechanistically different. A semi-synergistic mechanism, in which copy number dependence is reliant on the ability to overcome position effect (but not vice versa) may be plausible. The fact that the mRNA of an LCR-linked reporter gene is expressed in a copy number dependent manner when this transgene is introduced in a non-committed cell line, seems to point to the fact that factors present within this environment might be essential for this property. These factors may include one or more pioneer factors that function by priming and interacting with the surrounding chromatin to allow for a more permissive environment. These factors may allow for the DNA to be in a proper configuration within chromatin allowing for interactions that are conducive to proper expression patterns. The early priming of a gene locus has been shown to be relevant to the proper expression and regulation at subsequent developmental time points for certain genes. Thus, the action of pioneer factors in uncommitted cells may be directly linked to the subsequent ability of the

LCR to support copy number dependence. If this were true, it would explain why differentiated T cell lines, where these early priming factors are absent, failed to support the copy number dependence of LCR activity *de novo*.

Chapter 9: Materials and Methods

9.1 Reporter Constructs

All three versions of the mini LCRs were constructed from portions of the LCR and first assembled in pBlueScript SK I (pBSK) and then placed downstream of a modified pEYFP-C1 construct (Clontech).

pS-HS11' PvuII/Sca I

The smallest version of the mini TCRα LCR contain regions of HS1,HS1' and HS6. To attain the HS1 and 1' region a PCR scheme was developed that contained the *Pvu* II site located in HS1 and ended at the *Sca* I site located in HS1'[27]. This region contains the HS1 enhancer and ends right after the identified CTCF sites within HS1'[29,30]. The forward primer included a *Xho* I site on the 5' end and reverse primer included a *Cla* I site on the 3' end. The sequence for the forward primer was: 5'-GACCTGCTCGAGCTGCACCCTGAAATGGT-3'. The reverse primer sequence was: 5'-CGAGTCATCGATACTGCTGCCCCCGTGTG-3'. The insertion of these two unique enzymes enabled us to digest the PCR product an insert it into the *Xho* I and *Cla* I sites of pBSK. Once assembled, this pBSK plasmid contained the functional regions of the enhancer and HS1' and was termed pS-HS11' PvuII/Sac I.

pS-LCR

To complete the construction of the small LCR, a *Bcl* I to *Bgl* II HS6 fragment was placed into the BamHI site of psHS11' PvuII/Sca I[27,34,35]. Since digestions of DNA with either *Bcl* I and *Bgl* II leave the ends compatible to *Bam* HI, the resulting ligation had to be analyzed for

correct orientation. This was confirmed by analysis restriction analysis and the plasmid with correct orientation was selected. The resulting plasmid was termed pSLCR.

pHS1

To isolate the TCRα enhancer, *pS-HS11'* was digested with *Xho* I and *Bgl* II. The resulting fragment contains the TCRα enhancer[26-28]. This *Xho* I to *Bgl* II fragment from *pS-HS11'* was placed into restriction site *Xho* I and *Bam* HI of PBSK.

pM-LCR

The medium version of the mini TCRα LCR has the same regions HS1, HS1' and HS6 as the small LCR, but, with the inclusion of region containing HS4. An *ApaL* I to *Dra* I region of HS4[26,30] was blunted and ligated into a unique *EcoR* V site within pSLCR.

pL-HS11' PvuII/BamHI

To attain the enhancer (HS1) and complete HS1' region[27] of the TCRα LCR, a PCR primer pair was designed to span this region from a *Pvu* II site located in HS1 to a *Bam* HI site located in HS1'. The forward primer included a non-endogenous *Xho* I site on the 5' end and reverse primer included a non-endogenous *Cla* I site on the 3' end. The sequence for the forward primer was: 5'-GACCTGCTCGAGCTGCACCCTGAAATGGT-3', while the sequence for the reverse primer was: 5'-CGAGTCATCGATACTGCTGCCCCCGTGTGG-3'. The PCR product was digested with *Xho* I and *Cla* I and placed in to their corresponding sites on PBSK. Once assembled, this PBSK plasmid contained the complete HS and HS1' region and was termed pL-HS11' PvuII/BamH I.

pL-HS11'/L-HS6

We excised a *Mfe* I to *Ecl36* II fragment containing the complete HS6 site. We placed this fragment into the *EcoR* I site and *Sma* I site of pL-HS11' PvuII/BamH I.

pL-LCR

An *ApaL* I to *Dra* I fragment containing HS4 was blunted and inserted into the *EcoR* V site of pL-HS11'/L-HS6 to complete pL-LCR.

p-eYFP-VαXS ΔMCS KPN

The multiple cloning site for p-EYFP-C1 (Clontech) was deleted and the resulting plasmid further modified by replacing the CMV promoter with a 2.2 kb TCRα promoter, $V\alpha_{17}$ (D. Sant'Angelo) by J. Gomos-Klein. Truncation mutations were made to the resulting plasmid, p-eYFP-Vα Δmcs, to attain the smallest $V\alpha_{17}$ fragment that still maintained promoter functionality. It was determined that a *Xba* I to *Sma* I deletion, resulting in 324bp $V\alpha_{17}$ fragment still resulted in YFP expression. This plasmid, now being driven by a 324bp $V\alpha_{17}$ fragment was called p-eYFP-Vα17XS.

This plasmid was further modified to aid in cloning by inserting unique enzymes 5' to the promoter and 3' to the Poly A. 5' to $V\alpha_{17}$ promoter, *Kpn* I, *Pac* I, and *Sal* I were introduced in that order (5'→ 3'). 3' to the Poly A enzyme sites for *Xho* I, *Cla* I, *Not* I, *Sac* II and *Kpn* I were introduced, in that order.

p-eYFP-VαXSHS1, p-eYFP-VαXS SLCR, p-eYFP-VαXS MLCR, and p-eYFP-VαXS LLCR.

The mini LCR reporter plasmids were constructed using the aforementioned mini LCR in PBSK and the p-eYFP-VαXS ΔMCS KPN. *Xho* I to *Not* I fragments from p-S-LCR, p-M-LCR,

and p-L-LCR were inserted into *Xho* I and *Not* I sites of p-eYFP-VαXS ΔMCS KPN to generate three mini LCR reporter constructs, p-eYFP-HS1, p-eYFP-VαXS SLCR, p-eYFP-VαXS MLCR, and p-eYFP-VαXS LLCR.

p-eYFP-HS1-6

A *Xho* I to *Cla* I fragment containing HS1-6 region of the LCR[27] was inserted into p-eYFP-VαXS ΔMCS KPN into the same respective enzyme restriction site.

hCD2 Reporter Constructs and Neomycin resistance gene

The hCD2ΔT transgene[73] was excised from the pBluescript SK vector using *Sal* I and *Bam* HI. The hCD2:1-8 transgene[72] was excised using *Kpn* I and *Not* I. The SV40 promoter-driven Neomycin-G418 resistance cassette was excised from the pEYFP-C1 vector (Clontech) using *Ssp*I and *Eco*O109I enzymes.

The Vα17 driven YFP reporter gene, was next replaced by a hCD2 reporter gene. A *Sal* I to *Bam* HI fragment containing hCD2 and its endogenous promoter were excised from hCD2ΔT. This was placed in the *Sal* I and *Bam* HI sites of either p-eYFP-HS1 p-eYFP-VαXS SLCR, p-eYFP-VαXS MLCR, p-eYFP-VαXS LLCR, or p-eYFP-VαXS HS1-6. This replaced the $V\alpha_{17}$ driven YFP reporter gene with the hCD2 reporter gene linked to the respective mini LCR, thereby generating hCD2 HS1, hCD2 SLCR, hCD2 MLCR, hCD2 LLCR, and hCD2:HS1-6.

Lentiviral Vector Reporter Constructs

pRRL sin cPPT hPGK eGFP WPRE and pRRL sin cPPT NTP eGFP WPRE vector were kindly provided by M. Sadelain. A 1kb *Xho* I (blunt) to *Not* I (fragment from pRRL sin cPPT hPGK eGFP WPRE was extracted containing the cPPT region of the vector. This fragment was

cloned into *Xba* I (blunt) and *Not* I of either p-eYFP-VαXS, p-eYFP-VαXS SLCR, p-eYFP-VαXS MLCR, or p-eYFP-VαXS LLCR. The resulting plasmid were called p-eYFP-VαXS cPPT, p-eYFP-VαXS SLCR cPPT, p-eYFP-VαXS MLCR cPPT, or p-eYFP-VαXS LLCR cPPT respectively.

A *Sal* I (blunt) to *Not* I fragment from either p-eYFP-VαXS cPPT, p-eYFP-VαXS SLCR cPPT, p-eYFP-VαXS MLCR cPPT, or p-eYFP-VαXS LLCR cPPT was then ligated into the *Cla* I (blunt) or *Not* I site of the pRRL sin cPPT NTP eGFP WPRE vector. The resulting vectors contained a hPGK driven NTP viral transduction marker and a Vα17 driven YFP reporter gene alone or linked to either the small, medium or large TCRα mini LCRs.

9.2 Culturing T Cell Lines and Their Transfection

T cell lines VL3-3M2[78] and C6VLB[79] were cultured in RPMI 1640 with 5% FBS and 10% FBS respectively, supplemented with 1% Penicillin-Streptomycin (Cellgro), 1% Glutagro (Cellgro) and 54μM β-mercaptoethanol (Sigma). Cells were transfected using a BioRad Gene Pulser (Approximately 1 $x10^7$ cells were re-suspended in 0.5 ml of Electroporation Buffer (Millipore) with 10 μg of hCD2:1-8 transgene fragment, or 5 μg of hCD2ΔT fragment. An equimolar amount of a Neomycin G418 resistance cassette was co- transfected with the reporter transgene. 24 hours post-transfection, Neomycin-G418 was added at a concentration of 0.4 mg/ml for VL3-3M2 and 0.35 mg/ml for C6VLB. Individual clones were obtained by serial dilution.

For the barrier assay, T cell line VL3-3M2 were transfected with either p-eYFP-VαXS HS1 p-eYFP-VαXS SLCR, p-eYFP-VαXS MLCR, p-eYFP-VαXS LLCR or p-eYFP-VαXS HS1-6 via nucleofection (Amaxa) according to manufacture protocol using program, G16. 48 hours

post-transfection Neomycin-G418 was added at a concentration of 0.4 mg/ml for approximately ten days. Individual clones were obtained by serial dilution.

9.3 Embryonic Stem Cell (ESC) Culture and Transfection

The mouse ESR1 cell line was co-cultured with Mitomycin C arrested Mouse Embryonic Fibroblasts (MEFs) (Millipore) in Dulbeco's Modified Eagle Media (DMEM) (Cellgro) supplemented with 20% FBS (Gemini), 1% Glutagro (Cellgro), 1% Penicillin/Streptomycin (Cellgro), 1% HEPES (Millipore), 1% Non Essential Amino Acids (Millipore), 0.1%

Gentamycin (Life Technologies), 0.1% β-mercaptoethanol (Life Technologies), and 10 ng/ml of Leukemia Inhibitory Factor (LIF) (Millipore). Cells were transfected by BioRad Gene Pulser (0.24kV and 950μF). Approximately 1 x10^7 ESCs were re-suspended in 500 ul of Electroporation Buffer (Millipore) with 15 mg of hCD2:1-8 transgene fragment, or 7.5 mg of hCD2ΔT fragment. An equimolar amount of a Neomycin-G418 resistance cassette was co- transfected with the reporter transgene. For complete selection protocol please refer to Chapter 3.

Both ESC and T cell transfectant clones were initially screened for transgene integration by PCR using primers complimentary to the hCD2 promoter region (*Forward*: 5'-GAGGAAAC CAACCCCTAAGATGAG-3' *Reverse*: 5'-CGTAATCTCTTTGGAGACTGCACC-3'). Intact transgene copy number was subsequently determined via Southern blot using an 800 bp *Bgl* II probe from the HS6 region of the TCRα LCR, as described previously[72]. Copy number estimates were determined by PhosphorImager analyses of at least three Southern blots for each set of clones. All clones directly compared in assays were analyzed for relative copy number on

the same Southern blots. Enzymes and probes chosen enabled simultaneous detection of distinct sized fragments from both the endogenous TCRα locus and the transgene. Transgene signals were normalized to the endogenous signal within each sample.

9.4 *In vitro* ESC Differentiation

The protocol for *in vitro* derivation of T cells, and other hematopoietic cell types, from mouse ESCs was carried out as previously described[69]. For our optimizations to this model please refer to Chapter 3.

9.5 Flow Cytometry

Analysis of cells derived from the *in vitro* Differentiation System

FACSCalibur and FACSVantage devices were used for FACS analysis. Antibodies used include Fluorescein isothiocyanate (FITC) conjugated, anti-hCD2 (clone S5.2), R-Phycoerythrin/cyanine dye 7 (PE- Cy7) conjugated anti-CD45 (Clone 30F-11), Allophycocyanin (APC) conjugated anti-CD44 (Clone IM7), PE-conjugated anti-CD25 (Clone 3C7), PE conjugated anti-CD8 (Clone 53-6.7), Anti CD16/32 (Clone 2.4G2), (all from BD Biosciences) and CD4 APC (Clone RM4-5) (Life Technologies). Dead Cell Discriminator (DCD) or 4', 6-diamidino-2-phenylindole (DAPI) (Life Technologies) was used to label non-viable cells. Before staining, cells were pre-treated with anti CD16/32 antibody (to block Fc receptors). Cells were stained with fluorochrome-conjugated antibodies for 20 minutes and washed three times. The media for staining and washes is 1% HEPES, 3% FBS in RPMI. For analyses, live cells were gated based on forward and side scatter and lack of DAPI or DCD

signal. CD45 was additionally used to gate on white blood cell types derived in co-culture. FlowJo (Tree Star, Inc.) software was used for data analyses.

Analysis of VL3 cells Transfected with Lentivirus Reporter Constructs

FACSCalibur device was used. Antibody used: CD271 Alexa Fluor 647 (BD). Cell staining was done as above and Flowjo was used to analyze data.

9.6 Quantitative Real-time PCR

On day 20 of ESC-OP9-DL1 co-culture, total RNA was extracted from one well of a 6 well plate per clone (Qiagen RNeasy micro kit). RNA for VL3-3M2 and C6VLB clones was extracted from 1 $x10^7$ cells (Qiagen RNeasy mini kit). cDNA was synthesized from 6µg of each of these RNAsamples (NEB Protoscript cDNA synthesis kit). qRT-PCR was performed using an Applied Biosystems 7500 device. Samples were prepared using the Dynamo SYBR green qPCR kit, (New England Biolabs). TCRα primers[60] were used to normalize for both loading and T cell content of the co-cultures. hCD2 primers were used to detect reporter gene expression (Forward: 5'-CCTTTCTGCTGGTGAACTTGTG-3' Reverse: 5'-TCAACACAACCCTGACCTGTG-3'). Relative levels of hCD2 gene products were calculated as follows: Non-transfected ESR1 C_t (cycle threshold) value was used as the "calibrator". C_t values observed using TCRα primers were used as our normalizing control, as TCRα is the "reference gene" in this assay. C_t values observed with hCD2 primers were all subtracted by the C_t values obtained for TCRα to obtain the normalized hCD2 ΔC_t for each co-culture. Next, all ΔC_t values were then subtracted by ΔC_t of the calibrator to obtain $\Delta\Delta C_t$. Normalized relative hCD2 expression in T cells derived from each clone, was calculated by using the

$2^{-\Delta\Delta CT}$ method. The resulting values were then divided by the transgene copy number determined for each clone.

9.7 Barrier Assay

Each clones generated from the transfection of VL3-3M2 (see section 8.2 in this Chapter) was separated into two populations and cultured in the presence (.4 mg/ml) or absence of Neomycin-G418 for up to 90 days to monitor the loss of YFP expression in the set maintained in drug free condition relative to the set maintained with drug. FACS was done using a BD Biosciences FACSCalibur.

Chapter 10: References

1 Richards, E. J. & Elgin, S. C. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* **108**, 489-500 (2002).

2 Gaszner, M. & Felsenfeld, G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature reviews. Genetics* **7**, 703-713, doi:10.1038/nrg1925 (2006).

3 West, A. G., Gaszner, M. & Felsenfeld, G. Insulators: many functions, many mechanisms. *Genes & development* **16**, 271-288, doi:10.1101/gad.954702 (2002).

4 Li, Q., Peterson, K. R., Fang, X. & Stamatoyannopoulos, G. Locus control regions. *Blood* **100**, 3077-3086, doi:10.1182/blood-2002-04-1104 (2002).

5 Li, Q., Harju, S. & Peterson, K. R. Locus control regions: coming of age at a decade plus. *Trends in genetics : TIG* **15**, 403-408 (1999).

6 Fraser, P. & Grosveld, F. Locus control regions, chromatin activation and transcription. *Current opinion in cell biology* **10**, 361-365 (1998).

7 Grosveld, F. Activation by locus control regions? *Current opinion in genetics & development* **9**, 152-157, doi:10.1016/S0959-437X(99)80023-9 (1999).

8 Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447**, 407-412, doi:10.1038/nature05915 (2007).

9 Engel, J. D. & Tanimoto, K. Looping, linking, and chromatin activity: new insights into beta-globin locus regulation. *Cell* **100**, 499-502 (2000).

10 Bulger, M. & Groudine, M. Looping versus linking: toward a model for long-distance gene activation. *Genes & development* **13**, 2465-2477 (1999).

11 Grosveld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51**, 975-985 (1987).

12 Jones, B. K., Monks, B. R., Liebhaber, S. A. & Cooke, N. E. The human growth hormone gene is regulated by a multicomponent locus control region. *Molecular and cellular biology* **15**, 7010-7021 (1995).

13 Kushida, M. M. *et al.* A 150-base pair 5' region of the MHC class I HLA-B7 gene is sufficient to direct tissue-specific expression and locus control region activity: the alpha site determines efficient expression and in vivo occupancy at multiple cis-active sites throughout this region. *J Immunol* **159**, 4913-4929 (1997).

14 Greaves, D. R., Wilson, F. D., Lang, G. & Kioussis, D. Human CD2 3'-flanking sequences confer high-level, T cell-specific, position-independent gene expression in transgenic mice. *Cell* **56**, 979-986 (1989).

15 Ortiz, B. D., Cado, D., Chen, V., Diaz, P. W. & Winoto, A. Adjacent DNA elements dominantly restrict the ubiquitous activity of a novel chromatin-opening region to specific tissues. *The EMBO journal* **16**, 5037-5045, doi:10.1093/emboj/16.16.5037 (1997).

16 Forrester, W. C. *et al.* A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. *Genes & development* **4**, 1637-1649 (1990).

17 Levings, P. P. & Bungert, J. The human beta-globin locus control region. *European journal of biochemistry / FEBS* **269**, 1589-1599 (2002).

18 Kioussis, D., Vanin, E., deLange, T., Flavell, R. A. & Grosveld, F. G. Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature* **306**, 662-666 (1983).

19 Driscoll, M. C., Dobkin, C. S. & Alter, B. P. Gamma delta beta-thalassemia due to a de novo mutation deleting the 5' beta-globin gene activation-region hypersensitive sites. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 7470-7474 (1989).

20 Magram, J., Chada, K. & Costantini, F. Developmental regulation of a cloned adult beta-globin gene in transgenic mice. *Nature* **315**, 338-340 (1985).

21 Festenstein, R. *et al.* Locus control region function and heterochromatin-induced position effect variegation. *Science* **271**, 1123-1125 (1996).

22 Milot, E. *et al.* Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* **87**, 105-114 (1996).

23 Blom van Assendelft, G., Hanscombe, O., Grosveld, F. & Greaves, D. R. The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell* **56**, 969-977 (1989).

24 Hong, N. A. *et al.* A targeted mutation at the T-cell receptor alpha/delta locus impairs T-cell development and reveals the presence of the nearby antiapoptosis gene Dad1. *Molecular and cellular biology* **17**, 2151-2157 (1997).

25 Hong, N. A., Kabra, N. H., Hsieh, S. N., Cado, D. & Winoto, A. In vivo overexpression of Dad1, the defender against apoptotic death-1, enhances T cell proliferation but does not protect against apoptosis. *J Immunol* **163**, 1888-1893 (1999).

26 Diaz, P., Cado, D. & Winoto, A. A locus control region in the T cell receptor alpha/delta locus. *Immunity* **1**, 207-217 (1994).

27 Ortiz, B. D., Cado, D. & Winoto, A. A new element within the T-cell receptor alpha locus required for tissue-specific locus control region activity. *Molecular and cellular biology* **19**, 1901-1909 (1999).

28 Winoto, A. & Baltimore, D. A novel, inducible and T cell-specific enhancer located at the 3' end of the T cell receptor alpha locus. *The EMBO journal* **8**, 729-733 (1989).

29 Magdinier, F., Yusufzai, T. M. & Felsenfeld, G. Both CTCF-dependent and -independent insulators are found between the mouse T cell receptor alpha and Dad1 genes. *The Journal of biological chemistry* **279**, 25381-25389, doi:10.1074/jbc.M403121200 (2004).

30 Gomos-Klein, J., Harrow, F., Alarcon, J. & Ortiz, B. D. CTCF-independent, but not CTCF-dependent, elements significantly contribute to TCR-alpha locus control region activity. *J Immunol* **179**, 1088-1095 (2007).

31 Bell, A. C., West, A. G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387-396 (1999).

32 Santoso, B., Ortiz, B. D. & Winoto, A. Control of organ-specific demethylation by an element of the T-cell receptor-alpha locus control region. *The Journal of biological chemistry* **275**, 1952-1958 (2000).

33 Palmiter, R. D. & Brinster, R. L. Germ-line transformation of mice. *Annual review of genetics* **20**, 465-499, doi:10.1146/annurev.ge.20.120186.002341 (1986).

34 Ortiz, B. D., Harrow, F., Cado, D., Santoso, B. & Winoto, A. Function and factor interactions of a locus control region element in the mouse T cell receptor-alpha/Dad1 gene locus. *J Immunol* **167**, 3836-3845 (2001).

35 Harrow, F., Amuta, J. U., Hutchinson, S. R., Akwaa, F. & Ortiz, B. D. Factors binding a non-classical Cis-element prevent heterochromatin effects on locus control region activity. *The Journal of biological chemistry* **279**, 17842-17849, doi:10.1074/jbc.M401258200 (2004).

36 Anderson, M. K., Hernandez-Hoyos, G., Diamond, R. A. & Rothenberg, E. V. Precise developmental regulation of Ets family transcription factors during specification and commitment to the T cell lineage. *Development* **126**, 3131-3148 (1999).

37 Taniuchi, S. *et al.* Differential diagnosis and clinical course of autoimmune neutropenia in infancy: comparison with congenital neutropenia. *Acta Paediatr* **91**, 1179-1182 (2002).

38 Westendorf, J. J. & Hiebert, S. W. Mammalian runt-domain proteins and their roles in hematopoiesis, osteogenesis, and leukemia. *Journal of cellular biochemistry* **Suppl 32-33**, 51-58 (1999).

39 Skarpidi, E., Vassilopoulos, G., Stamatoyannopoulos, G. & Li, Q. Comparison of expression of human globin genes transferred into mouse erythroleukemia cells and in transgenic mice. *Blood* **92**, 3416-3421 (1998).

40 Vassilopoulos, G. *et al.* Correct function of the locus control region may require passage through a nonerythroid cellular environment. *Blood* **93**, 703-712 (1999).

41 Nakano, T. Lymphohematopoietic development from embryonic stem cells in vitro. *Seminars in immunology* **7**, 197-203 (1995).

42 Wiles, M. V. & Keller, G. Multiple hematopoietic lineages develop from embryonic stem (ES) cells in culture. *Development* **111**, 259-267 (1991).

43 Schmitt, T. M. *et al.* Induction of T cell development and establishment of T cell competence from embryonic stem cells differentiated in vitro. *Nature immunology* **5**, 410-417, doi:10.1038/ni1055 (2004).

44 Martinez-Agosto, J. A., Mikkola, H. K., Hartenstein, V. & Banerjee, U. The hematopoietic stem cell and its niche: a comparative view. *Genes & development* **21**, 3044-3060, doi:10.1101/gad.1602607 (2007).

45 Mikkola, H. K. & Orkin, S. H. The journey of developing hematopoietic stem cells. *Development* **133**, 3733-3744, doi:10.1242/dev.02568 (2006).

46 Yang, Q., Jeremiah Bell, J. & Bhandoola, A. T-cell lineage determination. *Immunological reviews* **238**, 12-22, doi:10.1111/j.1600-065X.2010.00956.x (2010).

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

48 Ogawa, M., Porter, P. N. & Nakahata, T. Renewal and commitment to differentiation of hemopoietic stem cells (an interpretive review). *Blood* **61**, 823-829 (1983).

49 Olsen, A. L., Stachura, D. L. & Weiss, M. J. Designer blood: creating hematopoietic lineages from embryonic stem cells. *Blood* **107**, 1265-1275, doi:10.1182/blood-2005-09-3621 (2006).

50 Nakano, T., Kodama, H. & Honjo, T. Generation of lymphohematopoietic cells from embryonic stem cells in culture. *Science* **265**, 1098-1101 (1994).

51 Billips, L. G. *et al.* Differential roles of stromal cells, interleukin-7, and kit-ligand in the regulation of B lymphopoiesis. *Blood* **79**, 1185-1192 (1992).

52 Fahlman, C., Blomhoff, H. K., Veiby, O. P., McNiece, I. K. & Jacobsen, S. E. Stem cell factor and interleukin-7 synergize to enhance early myelopoiesis in vitro. *Blood* **84**, 1450-1456 (1994).

53 McNiece, I. K., Langley, K. E. & Zsebo, K. M. The role of recombinant stem cell factor in early B cell development. Synergistic interaction with IL-7. *J Immunol* **146**, 3785-3790 (1991).

54 Rusten, L. S., Lyman, S. D., Veiby, O. P. & Jacobsen, S. E. The FLT3 ligand is a direct and potent stimulator of the growth of primitive and committed human CD34+ bone marrow progenitor cells in vitro. *Blood* **87**, 1317-1325 (1996).

55 Veiby, O. P., Lyman, S. D. & Jacobsen, S. E. Combined signaling through interleukin-7 receptors and flt3 but not c-kit potently and selectively promotes B-cell commitment and differentiation from uncommitted murine bone marrow progenitor cells. *Blood* **88**, 1256-1265 (1996).

56 Radtke, F. *et al.* Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity* **10**, 547-558 (1999).

57 Pui, J. C. *et al.* Notch1 expression in early lymphopoiesis influences B versus T lineage determination. *Immunity* **11**, 299-308 (1999).

58 Thompson, P. K. & Zuniga-Pflucker, J. C. On becoming a T cell, a convergence of factors kick it up a Notch along the way. *Seminars in immunology* **23**, 350-359, doi:10.1016/j.smim.2011.08.007 (2011).

59 MacDonald, H. R., Wilson, A. & Radtke, F. Notch1 and T-cell development: insights from conditional knockout mice. *Trends in immunology* **22**, 155-160 (2001).

60 Knirr, S. *et al.* Ectopic T cell receptor-alpha locus control region activity in B cells is suppressed by direct linkage to two flanking genes at once. *PloS one* **5**, e15527, doi:10.1371/journal.pone.0015527 (2010).

61 Kodama, H., Nose, M., Niida, S. & Yamasaki, A. Essential role of macrophage colony-stimulating factor in the osteoclast differentiation supported by stromal cells. *The Journal of experimental medicine* **173**, 1291-1294 (1991).

62 Cho, S. K. *et al.* Functional characterization of B lymphocytes generated in vitro from embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 9797-9802 (1999).

63 de Pooter, R. F., Cho, S. K., Carlyle, J. R. & Zuniga-Pflucker, J. C. In vitro generation of T lymphocytes from embryonic stem cell-derived prehematopoietic progenitors. *Blood* **102**, 1649-1653, doi:10.1182/blood-2003-01-0224 (2003).

64 Schmitt, T. M. & Zuniga-Pflucker, J. C. Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro. *Immunity* **17**, 749-756 (2002).

65 Schmitt, T. M. & Zuniga-Pflucker, J. C. T-cell development, doing it in a dish. *Immunological reviews* **209**, 95-102, doi:10.1111/j.0105-2896.2006.00353.x (2006).

66 Kaye, J. *et al.* Selective development of CD4+ T cells in transgenic mice expressing a class II MHC-restricted antigen receptor. *Nature* **341**, 746-749, doi:10.1038/341746a0 (1989).

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

68 Cho, S. K., Bourdeau, A., Letarte, M. & Zuniga-Pflucker, J. C. Expression and function of CD105 during the onset of hematopoiesis from Flk1(+) precursors. *Blood* **98**, 3635-3642 (2001).

69 Holmes, R. & Zuniga-Pflucker, J. C. The OP9-DL1 system: generation of T-lymphocytes from embryonic or hematopoietic stem cells in vitro. *Cold Spring Harbor protocols* **2009**, pdb prot5156, doi:10.1101/pdb.prot5156 (2009).

70 Wachter, R. M., Elsliger, M. A., Kallio, K., Hanson, G. T. & Remington, S. J. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure* **6**, 1267-1277 (1998).

71 Brinster, R. L., Allen, J. M., Behringer, R. R., Gelinas, R. E. & Palmiter, R. D. Introns increase transcriptional efficiency in transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 836-840 (1988).

72 Harrow, F. & Ortiz, B. D. The TCRalpha locus control region specifies thymic, but not peripheral, patterns of TCRalpha gene expression. *J Immunol* **175**, 6659-6667 (2005).

73 Melton, E. *et al.* Transgene-encoded human CD2 acts in a dominant negative fashion to modify thymocyte selection signals in mice. *European journal of immunology* **26**, 2952-2963, doi:10.1002/eji.1830261222 (1996).

74 Lang, G., Mamalaki, C., Greenberg, D., Yannoutsos, N. & Kioussis, D. Deletion analysis of the human CD2 gene locus control region in transgenic mice. *Nucleic acids research* **19**, 5851-5856 (1991).

75 Rostovskaya, M. *et al.* Transposon-mediated BAC transgenesis in human ES cells. *Nucleic acids research* **40**, e150, doi:10.1093/nar/gks643 (2012).

76 Lahiji, A. *et al.* Complete TCR-alpha Gene Locus Control Region Activity in T Cells Derived In Vitro from Embryonic Stem Cells. *J Immunol* **191**, 472-479, doi:10.4049/jimmunol.1300521 (2013).

77 Ellis, J. *et al.* A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region. *The EMBO journal* **15**, 562-568 (1996).

78 Groves, T. *et al.* In vitro maturation of clonal CD4+CD8+ cell lines in response to TCR engagement. *J Immunol* **154**, 5011-5022 (1995).

79 Allison, J. P., McIntyre, B. W. & Bloch, D. Tumor-specific antigen of murine T-lymphoma defined with monoclonal antibody. *J Immunol* **129**, 2293-2300 (1982).

80 Smale, S. T. Pioneer factors in embryonic stem cells and differentiation. *Current opinion in genetics & development* **20**, 519-526, doi:10.1016/j.gde.2010.06.010 (2010).

81 Xu, J. *et al.* Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes & development* **23**, 2824-2838, doi:10.1101/gad.1861209 (2009).

82 Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nature biotechnology* **15**, 871-875, doi:10.1038/nbt0997-871 (1997).

83 Schroder, A. R. *et al.* HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521-529 (2002).

84 Papapetrou, E. P. *et al.* Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells. *Nature biotechnology* **29**, 73-78, doi:10.1038/nbt.1717 (2011).

85 Howe, S. J. *et al.* Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *The Journal of clinical investigation* **118**, 3143-3150, doi:10.1172/JCI35798 (2008).

86 Hacein-Bey-Abina, S. *et al.* LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415-419, doi:10.1126/science.1088547 (2003).

87 Lacorazza, H. D., Tucek-Szabo, C., Vasovic, L. V., Remus, K. & Nikolich-Zugich, J. Premature TCR alpha beta expression and signaling in early thymocytes impair thymocyte expansion and partially block their development. *J Immunol* **166**, 3184-3193 (2001).

88 Amado, R. G. & Chen, I. S. Lentiviral vectors--the promise of gene therapy within reach? *Science* **285**, 674-676 (1999).

89 Buchschacher, G. L., Jr. & Wong-Staal, F. Development of lentiviral vectors for gene therapy for human diseases. *Blood* **95**, 2499-2504 (2000).

90 Chang, A. H. & Sadelain, M. The genetic engineering of hematopoietic stem cells: the rise of lentiviral vectors, the conundrum of the ltr, and the promise of lineage-restricted vectors. *Molecular therapy : the journal of the American Society of Gene Therapy* **15**, 445-456, doi:10.1038/sj.mt.6300060 (2007).

91 Bett, A. J., Prevec, L. & Graham, F. L. Packaging capacity and stability of human adenovirus type 5 vectors. *Journal of virology* **67**, 5911-5921 (1993).

92 Thomas, C. E., Ehrhardt, A. & Kay, M. A. Progress and problems with the use of viral vectors for gene therapy. *Nature reviews. Genetics* **4**, 346-358, doi:10.1038/nrg1066 (2003).

93 Sadelain, M., Wang, C. H., Antoniou, M., Grosveld, F. & Mulligan, R. C. Generation of a high-titer retroviral vector capable of expressing high levels of the human beta-globin gene. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 6728-6732 (1995).

94 Sadelain, M. Genetic treatment of the haemoglobinopathies: recombinations and new combinations. *British journal of haematology* **98**, 247-253 (1997).

95 May, C. *et al.* Therapeutic haemoglobin synthesis in beta-thalassaemic mice expressing lentivirus-encoded human beta-globin. *Nature* **406**, 82-86, doi:10.1038/35017565 (2000).

96 Sirven, A. *et al.* The human immunodeficiency virus type-1 central DNA flap is a crucial determinant for lentiviral vector nuclear import and gene transduction of human hematopoietic stem cells. *Blood* **96**, 4103-4110 (2000).

97 Zennou, V. *et al.* HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell* **101**, 173-185, doi:10.1016/S0092-8674(00)80828-4 (2000).

98 Huang, Z. M. & Yen, T. S. Hepatitis B virus RNA element that facilitates accumulation of surface gene transcripts in the cytoplasm. *Journal of virology* **68**, 3193-3199 (1994).

99 Donello, J. E., Loeb, J. E. & Hope, T. J. Woodchuck hepatitis virus contains a tripartite posttranscriptional regulatory element. *Journal of virology* **72**, 5085-5092 (1998).

100 Krebs, F. C., Hogan, T. H., Quiterio, S., Gartner, S., & Wigdahl, B. Lentiviral LTR-directed expression, sequence variation, and disease pathogenesis. *HIV sequence compendium*, 29-70 (2001).

101 Gallardo, H. F., Tan, C., Ory, D. & Sadelain, M. Recombinant retroviruses pseudotyped with the vesicular stomatitis virus G glycoprotein mediate both stable gene transfer and pseudotransduction in human peripheral blood lymphocytes. *Blood* **90**, 952-957 (1997).

102 Pawliuk, R. *et al.* Correction of sickle cell disease in transgenic mouse models by gene therapy. *Science* **294**, 2368-2371, doi:10.1126/science.1065806 (2001).

103 Hacein-Bey, H. *et al.* gamma-c gene transfer into SCID X1 patients' B-cell lines restores normal high-affinity interleukin-2 receptor expression and function. *Blood* **87**, 3108-3116 (1996).

104 Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R. & Breindl, M. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *Journal of virology* **61**, 336-343 (1987).

105 Mitchell, R. S. *et al.* Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS biology* **2**, E234, doi:10.1371/journal.pbio.0020234 (2004).

106 Schambach, A. *et al.* Overcoming promoter competition in packaging cells improves production of self-inactivating retroviral vectors. *Gene therapy* **13**, 1524-1533, doi:10.1038/sj.gt.3302807 (2006).

107 Kohn, D. B., Sadelain, M. & Glorioso, J. C. Occurrence of leukaemia following gene therapy of X-linked SCID. *Nature reviews. Cancer* **3**, 477-488, doi:10.1038/nrc1122 (2003).

108 McCormack, M. P. *et al.* The Lmo2 oncogene initiates leukemia in mice by inducing thymocyte self-renewal. *Science* **327**, 879-883, doi:10.1126/science.1182378 (2010).

109 Barkess, G. & West, A. G. Chromatin insulator elements: establishing barriers to set heterochromatin boundaries. *Epigenomics* **4**, 67-80, doi:10.2217/epi.11.112 (2012).

110 Oki, M., Valenzuela, L., Chiba, T., Ito, T. & Kamakaka, R. T. Barrier proteins remodel and modify chromatin to restrict silenced domains. *Molecular and cellular biology* **24**, 1956-1967 (2004).

111 Udvardy, A., Maine, E. & Schedl, P. The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *Journal of molecular biology* **185**, 341-358 (1985).

112 Kellum, R. & Schedl, P. A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64**, 941-950 (1991).

113 Kellum, R. & Schedl, P. A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Molecular and cellular biology* **12**, 2424-2431 (1992).

114 Chung, J. H., Whiteley, M. & Felsenfeld, G. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell* **74**, 505-514 (1993).

115 Chung, J. H., Bell, A. C. & Felsenfeld, G. Characterization of the chicken beta-globin insulator. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 575-580 (1997).

116 West, A. G., Huang, S., Gaszner, M., Litt, M. D. & Felsenfeld, G. Recruitment of histone modifications by USF proteins at a vertebrate barrier element. *Molecular cell* **16**, 453-463, doi:10.1016/j.molcel.2004.10.005 (2004).

117 Huang, S., Li, X., Yusufzai, T. M., Qiu, Y. & Felsenfeld, G. USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier. *Molecular and cellular biology* **27**, 7991-8002, doi:10.1128/MCB.01326-07 (2007).

118 Dickson, J. *et al.* VEZF1 elements mediate protection from DNA methylation. *PLoS genetics* **6**, e1000804, doi:10.1371/journal.pgen.1000804 (2010).

119 Gallagher, P. G. *et al.* An insulator with barrier-element activity promotes alpha-spectrin gene expression in erythroid cells. *Blood* **113**, 1547-1554, doi:10.1182/blood-2008-06-164954 (2009).

120 Gallagher, P. G. *et al.* Mutation of a barrier insulator in the human ankyrin-1 gene is associated with hereditary spherocytosis. *The Journal of clinical investigation* **120**, 4453-4465, doi:10.1172/JCI42240 (2010).

121 Mutskov, V. & Felsenfeld, G. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *The EMBO journal* **23**, 138-149, doi:10.1038/sj.emboj.7600013 (2004).

122 Pikaart, M. J., Recillas-Targa, F. & Felsenfeld, G. Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes & development* **12**, 2852-2862 (1998).

123 Zhong, X. P. & Krangel, M. S. Enhancer-blocking activity within the DNase I hypersensitive site 2 to 6 region between the TCR alpha and Dad1 genes. *J Immunol* **163**, 295-300 (1999).

124 Molto, E., Fernandez, A. & Montoliu, L. Boundaries in vertebrate genomes: different solutions to adequately insulate gene expression domains. *Briefings in functional genomics & proteomics* **8**, 283-296, doi:10.1093/bfgp/elp031 (2009).

125 Bell, A. C. & Felsenfeld, G. Stopped at the border: boundaries and insulators. *Current opinion in genetics & development* **9**, 191-198, doi:10.1016/S0959-437X(99)80029-X (1999).

126 Gerasimova, T. I. & Corces, V. G. Boundary and insulator elements in chromosomes. *Current opinion in genetics & development* **6**, 185-192 (1996).

127 Rivella, S., Callegari, J. A., May, C., Tan, C. W. & Sadelain, M. The cHS4 insulator increases the probability of retroviral expression at random chromosomal integration sites. *Journal of virology* **74**, 4679-4687 (2000).

128 Emery, D. W. *et al.* Development of virus vectors for gene therapy of beta chain hemoglobinopathies: flanking with a chromatin insulator reduces gamma-globin gene silencing in vivo. *Blood* **100**, 2012-2019, doi:10.1182/blood-2002-01-0219 (2002).

129 Ramezani, A., Hawley, T. S. & Hawley, R. G. Performance- and safety-enhanced lentiviral vectors containing the human interferon-beta scaffold attachment region and the chicken beta-globin insulator. *Blood* **101**, 4717-4724, doi:10.1182/blood-2002-09-2991 (2003).

130 Bassing, C. H. *et al.* T cell receptor (TCR) alpha/delta locus enhancer identity and position are critical for the assembly of TCR delta and alpha variable region genes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 2598-2603, doi:10.1073/pnas.0437943100 (2003).

131 Sleckman, B. P., Bardon, C. G., Ferrini, R., Davidson, L. & Alt, F. W. Function of the TCR alpha enhancer in alphabeta and gammadelta T cells. *Immunity* **7**, 505-515 (1997).