NOTE TO USERS

This reproduction is the best copy available.

UMI®

uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Zeshawn Awan
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Cellular and Molecular Medicine)
GRADE / DEGREE

School of Cellular and Molecular Medicine
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Investigating the Role of MarfK Subnuclear Relocalization During Erythroid Differentiation

TITRE DE LA THÈSE / TITLE OF THESIS

Marjorie Brand
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

Robin Parks

Michaël Rudnicki

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Investigating the Role of MafK Subnuclear Relocalization During Erythroid Differentiation

By
Zeshawn Awan, B.Sc.

This thesis is submitted as a partial fulfillment of the M.Sc. program in Cellular and Molecular Medicine.

May 13, 2009

Faculty of Medicine
Faculty of Graduate and Postdoctoral Studies
University of Ottawa
Ottawa, Ontario, Canada

©Zeshawn Awan, Ottawa, Canada, 2009

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-61337-5
Our file Notre référence
ISBN: 978-0-494-61337-5

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

ABSTRACT

MafK, a transcription factor that regulates β-globin gene expression, undergoes subnuclear relocalization from heterochromatic to euchromatic regions during erythroid differentiation. As a first step to deciphering the role of MafK subnuclear relocalization, we investigated the characteristics of MafK heterochromatic localization before differentiation. Immunoprecipitation and mass spectrometry experiments indicate that MafK exists as a homodimer in the heterochromatic fraction, distinguishing it from MafK at the β-globin locus. Immunofluorescence microscopy results demonstrate that a functional DNA-binding domain is necessary for MafK heterochromatic targeting. Native chromatin immunoprecipitation and quantitative PCR results suggest that MafK localizes to heterochromatin by binding pericentromeric major satellite DNA. A bioinformatics search for known Maf recognition elements indicates that pericentromeric heterochromatin contains novel MafK binding site(s). Our results support a role for subnuclear relocalization in regulating the spatial distribution and local concentration of MafK, presumably as an additional level of control over β-globin gene expression, during erythroid differentiation.

TABLE OF CONTENTS

Chapter 1: Introduction 1-19

1.1 Overview 2
1.2 Erythropoiesis 3
1.3 β-globin 6
1.4 MafK 11
1.5 Heterochromatin 15
1.6 Statement of Intent 19
1.7 Hypothesis 19

Chapter 2: Materials and Methods 20-31

2.1 Cell Culture 21
2.2 Protein Extraction 21
2.3 Silver Stain Analysis 23
2.4 Immunoprecipitation of MafK 23
2.5 Western Blot Analysis 24
2.6 LC-ESI-MS/MS 25
2.7 Cloning, Constructs and Transfection 26
2.8 Co-immunoprecipitation of Flag/Myc-MafK 28
2.9 Immunofluorescence Microscopy 28
2.10 Chromatin Immunoprecipitation (ChIP) 29
2.11 Quantitative PCR 30

Chapter 3: Results 32-58

3.1 MafK relocalizes away from heterochromatin upon differentiation. 33
3.2 Heterochromatic MafK does not interact with known heterodimerization partners. 37
3.3 MafK homodimers localize preferentially to the heterochromatic fraction. 41
3.4 A functional DNA-binding domain is critical to MafK heterochromatic localization. 47
3.5 MafK interacts with pericentromeric heterochromatin major satellite repeats. 50
3.6 Heterochromatic satellite DNA appears to contain a novel MafK binding site. 55

Chapter 4: Discussion 59-68

References 69

Appendix 79

LIST OF TABLES

Table 1. Characteristics of Murine Constitutive Heterochromatin. **17**

Table 2. Tandem mass spectrometry (LC-ESI-MS/MS) identification of proteins immunoprecipitated from the heterochromatic extract of undifferentiated MEL cell nuclei using MafK-specific antibody. **44**

Table 3. Tandem mass spectrometry (LC-ESI-MS/MS) identification of proteins immunoprecipitated from the heterochromatic extract of undifferentiated MEL cell nuclei using normal rabbit IgG. **45**

Table 4. Maf recognition element (MARE) single nucleotide substitution variants. **58**

Table 5. Results of search for known Maf recognition element (MARE) variants in murine heterochromatic satellite DNA using Findmatches.pl. **58**

List of Figures

Figure 1. Organization of the human β-globin locus. 10

Figure 2. Examining MafK protein-protein interactions in heterochromatin. 35

Figure 3. MafK relocalizes away from heterochromatin upon differentiation. 36

Figure 4. MafK in heterochromatin does not interact with known dimerization partners. 40

Figure 5. MafK homodimers are preferentially associated with the heterochromatic fraction. 46

Figure 6. MafK DNA-binding domain is required for heterochromatic localization. 49

Figure 7. MafK interacts strongly with heterochromatin. 53

Figure 8. MafK interacts directly with pericentromeric chromatin. 54

Figure 9. Possible roles of MafK subnuclear relocalization in β-globin silencing before erythroid differentiation. 68

Figure S1. Findmatches.pl script. 80

Figure S2. Murine Satellite DNA sequences. 81

LIST OF ABBREVIATIONS

Bach1	BTB and CNC Homology 1
BFU-E	Blast Forming Unit - Erythroid
BLAST	Basic Local Alignment Search Tool
bZIP	Basic Leucine Zipper
cAMP	Cyclic Adenosine Monophosphate
CFU-E	Colony Forming Unit - Erythroid
CFU-GEMM	Colony Forming Unit - Granulocyte, Erythroid, Macrophage, Megakaryocyte
CFU-S	Colony Forming Unit - Spleen
ChIP	Chromatin Immunoprecipitation
CNC	Cap-N-Collar
CRE	cAMP Responsive Element
DAPI	4′,6-diamidino-2-phenylindole
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNase 1	Deoxyribonuclease 1
DTT	1,4-Dithithreitol
EDTA	Ethylenediaminetetraacetic Acid
EE	Euchromatic Extract
EKLF	Erythroid Krüppel-like Factor
HE	Heterochromatic Extract
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HP1α	Heterochromatin Protein 1α
HPLC	High-performance liquid chromatography
HS	Hypersensitive Site
HSC	Hematopoietic Stem Cell
ICAT	Isotope Coded Affinity Tags
IgG	Immunoglobulin G
IP	Immunoprecipitation
L2PM4P	Mutation of Leucine at Position 2 and Methionine at Position 4 to Proline
LC-ESI-MS/MS	Liquid Chromatography – Electrospray Ionization – Tandem Mass Spectrometry
LCR	Locus Control Region
Maf	Musculoaponeurotic fibrosarcoma
MARE	Maf Recognition Element
MEL	Murine Erythroleukemia
MNase	Micrococcal Nuclease
NCBI	National Center for Biotechnology Information
NF-E2p45	Nuclear Factor – Erythroid 2 Subunit p45
NP	Nuclear Pellet
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PCV	Packed Cell Volume
PEV	Position Effect Variegation
PIC	Protease Inhibitor Cocktail

R22E	Mutation of Arginine at Position 22 to Glutamate
RNA	Ribonucleic Acid
RNase A	Ribonuclease A
SCL	Stem Cell Leukemia (transcription factor)
SDS-PAGE	Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis
TBP	TATA-Binding Protein
TEMED	Tetramethylethylenediamine
TPA	12-O-tetradecanoylphorbol 13-acetate
TRE	TPA Responsive Element

ACKNOWLEDGEMENTS

The process of undertaking graduate research has been one of the most rewarding experiences of my life. The late nights in the lab, working alone at 2am to try and finish the three western blots running simultaneously; the hours of pacing back and forth, pondering the results of an experiment, or how to optimize it, or what direction to take next; reading and assessing studies conducted by countless other scientists from around the world in the hopes that some pattern will emerge, some hint will be found that will shed light on the phenomenon I'm exploring; and finally the excitement of reaching some conclusions by our own experiments and analysis: These are the realities of research and the fond memories that I will take with me from this experience.

Many people supported me, guided me and taught me along this journey, without whom this thesis would not be. I would like to begin by thanking my supervisor Dr. Marjorie Brand. Being one of the first members to join the Brand lab, I was lucky to have learned directly from Marjorie as she was always in the lab. As I grew into my research, so grew the lab around me: new people, new equipment, new building; yet Marjorie was always available to me, showing me how to plan and conduct research, ready to answer any questions that I had, and taking time to discuss any idea that may have occurred to me. Marjorie always treated me more like a friend than a student, encouraged me to give my best during the most frustrating times, and recognized all of my achievements. Thank you Marjorie for the opportunity and the guidance; I sincerely appreciate it.

I would also like to thank Dr. Jeff Dilworth as well as members of my Advisory Committee, Dr. Michael Rudnicki and Dr. Bernard Jasmin, for their guidance, advice, critiques and challenges that shaped the direction of my project and taught me critical analysis of scientific data.

To my colleagues and friends in Marjorie's lab, particularly Celina Demers and Thanh Le, who were there with me every day through the daily grind of scientific research; your support and friendship were invaluable.

To my peers who went through the graduate program with me and all of my friends that supported me during this time, who kept me company during those late nights of researching and writing, encouraged me to finish the thesis, and cheered me on when I completed the degree; thank you for your kindness, your confidence, and your applause.

Finally, I would like to thank my family, my father Shamim Awan, my sisters Farheen and Sabeen, and my brother Rizwan, for the support, the laughs, the motivation and the inspiration. In particular, I would like to thank my mother, Feroza Naheed; it is not possible to put into words how she has contributed to every aspect of my life. In my every success is her hand, in every achievement her reflection; to her I dedicate this thesis.

STATEMENT OF CONTRIBUTIONS

All studies appearing in this manuscript were performed by Zeshawn Awan under the supervision of Dr. Marjorie Brand with the exception of the following:

- The Perl script "FindMatches.pl" was written by C. Porter and M. Huska of the Ottawa Health Research Institute Bioinformatics Group.

Additionally, the following reagents were obtained as indicated:

- MafK/NF-E2p18 cDNA cloned into the mammalian expression vector pCS2-MT in-frame with six Myc epitope tags in the N terminus was kindly provided by Dr. Mark Groudine (Francastel et al. 2001).

- MafK and point mutants MafK-R22E and MafK-L2PM4P cDNA cloned into the mammalian expression vector pEF-BOS were provided courtesy of Dr. Kohsuke Kataoka (Kataoka et al. 1995).

CHAPTER 1

INTRODUCTION

1.1 Overview

Over the last decade, stem cells have risen to the forefront of biomedical research. With their ability for self-renewal as well as giving rise to a diverse range of specialized cell types, both adult and embryonic stem cells have the potential to significantly impact the way in which a variety of human diseases and disorders are addressed. Stem cells can and are being used by researchers as models for investigating the molecular mechanisms underlying diseases such as cancer, with the hopes of identifying new therapeutic targets and strategies (Wade 2006). Clinically, the regenerative ability of stem cells holds the promise for effective therapeutic interventions in conditions wherein specialized cells have been damaged, such as spinal cord injury (Barnabe-Heider and Frisen 2008), are degenerating, as with Parkinson's disease (Deierborg et al. 2008) and muscular dystrophies (Boldrin and Morgan 2007), or are non-functional or improperly functioning, as with diabetes mellitus (Voltarelli et al. 2008). Currently, one of the most common and well-established stem cell treatments is the bone marrow transplant, which involves the transplantation of hematopoietic stem cells (HSC) for people with diseases of blood and/or bone marrow. HSCs located in the adult bone marrow give rise to all the different types of blood cells via a differentiation process termed hematopoiesis (Morrison et al. 1995; Mikkola and Orkin 2006). Multipotent HSCs differentiate successively into oligopotent progenitors, committed precursors, and fully differentiated blood cells, including erythrocytes (red blood cells). Each of these stages has an increasingly restricted potential

which is established by a unique and highly regulated pattern of gene expression.

Transcription factors play an important role in this regulation, allowing the cell to respond to physiological and environmental signals for differentiation by enhancing or repressing the expression of particular genes. A loss in the ability of hematopoietic progenitors or precursors to further differentiate can lead to leukemogenesis (Tenen 2003), and considerable evidence points to aberrant regulation of both ubiquitous and lineage-restricted transcription factors as a possible cause (Moreau-Gachelin et al. 1996; Hu et al. 1997; Zhang et al. 1999; Cantor and Orkin 2001; Rosenbauer et al. 2004). Additionally, a number of disorders are attributed to disruptions in gene expression and differentiation which adversely affect the quantity and/or functionality of specialized cells such as red blood cells. Thus, research that elucidates the molecular basis of transcriptional regulation during differentiation is not only crucial to unlocking the full medical potential of stem cells but could also potentially shed light on the causes of (and potential treatments for) a variety of human diseases and disorders. This type of research may also have relevance in fields such as gene therapy, where an understanding of how gene expression is regulated is a prerequisite to manipulating it for clinical purposes (Persons and Tisdale 2004).

1.2 Erythropoiesis

Erythrocytes, or red blood cells, are the epitome of specialized cells, functioning to carry life-giving oxygen to all internal tissues of aerobic organisms

and carrying away the waste-product of aerobic respiration, carbon dioxide. Erythrocyte function is mediated by hemoglobin, the iron-containing oxygen-binding protein that gives blood its red color, which represents approximately 95% of total proteins in the mature red blood cell (Hoffman et al. 2008). In adults, functional erythrocytes have a lifetime of about four months and must be continually replenished as aging cells are removed from the circulation. Like all blood cells, erythrocytes derive from HSCs; the differentiation program is called erythropoiesis.

Erythropoiesis takes place through a series of distinct cell intermediates that can be identified initially by *ex vivo* functional assays but later on by morphological characteristics. The first of these intermediates is termed the CFU-S (colony forming unit – spleen), which retains the multipotency of HSCs but loses long-term repopulation capability. Further differentiation leads to loss of lymphoid potential resulting in CFU-GEMM (colony forming unit – granulocyte, erythroid, macrophage, megakaryocyte) that give rise to the erythroid-lineage restricted progenitors BFU-E (burst forming unit – erythroid) and then CFU-E (colony forming unit – erythroid), which are characterized by *ex vivo* colony size and responsiveness to various cytokines (Johnson and Metcalf 1977; Krantz et al. 1987; Sawada et al. 1987; Sawada et al. 1990; Sawada et al. 1991).

Restriction of proliferative capability gives rise to the first cells morphologically recognizable as erythroid, called proerythroblasts. The cells are smaller than in previous stages of differentiation, the chromatin has a granular appearance, and nucleoli are visible. As development progresses, the cell size

decreases, the chromatin appears more condensed, and the increased abundance of ribosomes causes the cytoplasm to become more basophilic (i.e. stained blue with the basic dye haematoxylin); thus the cells are referred to as basophilic erythroblasts. As mitotic divisions continue and the cells begin to produce hemoglobin, the cytoplasm becomes both basophilic and eosinophilic (i.e. stained pink with the acidic dye eosin Y) and are called polychromatophilic erythroblasts. Accumulation of hemoglobin results in the cells eventually becoming more eosinophilic, now termed acidophilic orthochromatic erythroblasts. Throughout these last few stages of erythropoiesis, the cell size continues to decrease and chromatin continues to become more and more condensed, until finally the inactive nucleus is expelled. The resultant cells, which contain reticular networks of polyribosomes, are termed reticulocytes. Reticulocytes lose their polyribosomes and enter the circulation as mature erythrocytes (Alter 1994; Hoffman et al. 2008).

Murine erythroleukemia (MEL) cells provide a controllable system to study the molecular mechanisms involved in erythroid differentiation as well as leukemogenesis. Isolated from mice infected by the Friend virus as first described by Charlotte Friend (Friend 1957), MEL cells are committed to the erythroid lineage, but are blocked at the proerythroblast stage. These cells continue to proliferate in suspension culture indefinitely; however, the main advantage of MEL cells is that they can be consistently induced to terminally differentiate to red blood cells. Exposure of MEL cells to dimethyl sulfoxide (DMSO) treatment induces changes in gene expression that appear to overcome

the leukemic phenotype and reinstate normal erythroid differentiation. While protein kinase C has been implicated as an important early step, the exact mechanism(s) by which DMSO induces differentiation remains unclear (Chakravarthy et al. 1992; Durkin et al. 1992). Over a DMSO treatment period of four days, MEL cells display characteristics of mature erythrocytes: the cells become red in color, stain positive for benzidine, and there are increased levels of globin mRNA, globin synthesis and heme synthesis, as well as characteristic changes in the activity of various enzyme markers (Friend et al. 1971; Boyer et al. 1972; Ross et al. 1972; Reem and Friend 1975).

1.3 β-globin

Mammalian hemoglobin, the protein responsible for carrying oxygen in erythrocytes, is a tetrameric metalloprotein consisting of two α-like and two β-like globin polypeptide subunits with a quaternary structure of $\alpha_2\beta_2$ (two αβ dimers) in adults. Each globin chain contains an iron ion (Fe^{2+})-containing heme group to which oxygen reversibly binds, altering the entire hemoglobin tetramer structure in the process (Berg et al. 2002). Disruptions in α- or β-globin expression results in hemoglobinopathies; these are amongst the most prevalent of all human genetic disorders, affecting 300,000 newborn infants worldwide each year (World Health Organization 2006). This includes α- and β-thalassemia, which are caused by disrupted gene expression and decreased levels of α- and β-globin, respectively. The resulting imbalance of globins leads to decreased functional hemoglobin, degradation of erythrocytes and results in mild to lethal anemia.

The disturbed gene expression can be due to deletions or mutations in the globin genes and/or related regulatory elements, the nature of which affects the severity of the resultant anemia (Bank et al. 1980; Ramirez et al. 1980; Kioussis et al. 1983; Collins and Weissman 1984).

Another example of a common hemoglobinopathy is sickle cell disease, which is caused by a single base pair mutation in the β-globin gene resulting in a glutamine-to-valine substitution at position six of β-globin (Ingram 1956). This single substitution significantly reduces the solubility of the deoxygenated hemoglobin which then precipitates. This causes the red blood cells to take on a sickle shape and become rigid, which can then block blood flow in capillaries and cause organ damage. This quickly becomes a vicious cycle as blocked circulation leads to more deoxygenation, precipitation, sickling and ischemia. To this day, there is still no adequate treatment for sickle cell disease, though stem cell therapy (i.e. bone marrow transplant) and gene therapies appear to be promising (Persons and Tisdale 2004; Wu et al. 2005; Frenette and Atweh 2007). Interestingly, the prevalence of hemoglobinopathies appears to be related, at least in part, to a selective survival advantage conferred to affected individuals in certain parts of the world as the destruction of erythrocytes provides some measure of protection against malaria (Roberts and Williams 2003). Nevertheless, intensive research aimed at addressing these diseases that affect so many people has contributed to the accumulation of a wealth of knowledge regarding tissue-specific and stage-specific regulation of gene expression.

In particular, the β-globin locus has been extensively used as a model for studying transcriptional control of developmentally regulated gene expression. The β-globin locus consists of several β-like globin genes that are sequentially transcribed at specific developmental stages (Hardison et al. 1997). In humans, the five β-like globin genes ε, Gγ, Aγ, δ and β-globin are arranged in the β-globin locus from 5′ to 3′ in the same order as their developmental expression (**Figure 1**). In the earliest stages of development, hematopoiesis occurs in the embryonic yolk sac and the ε-globin gene is expressed. As development progresses to the fetal stage, the location of hematopoiesis then switches to the liver, with a corresponding switch in globin expression from ε-globin to the two γ-globins. These embryonic/fetal globins form a type of hemoglobin which has a higher affinity for oxygen than the adult type, thereby allowing efficient extraction of oxygen from the maternal blood (Purdie et al. 1983). The next switch in transcription from γ- to δ- and β-globin occurs during the postnatal period when bone marrow becomes the site of hematopoiesis. Both δ- and β-globin are expressed in adults, though δ-globin represents only a small fraction of total β-like globin gene expression; the majority of adult hemoglobin contains β-globin (Levings and Bungert 2002).

In addition to developmental gene expression, the β-globin locus also represents a paradigm for transcriptional regulation that drives cellular differentiation, since activation of high level β-globin gene expression and consequent hemoglobin synthesis are required for terminal erythroid differentiation in mammals. Genetic characterization of Dutch deletion type β-

thalassemia, in which the β-globin gene and its promoters are intact but there is a deletion of ~100 kb upstream of the β-globin locus, led to the discovery of a *cis*-regulatory element upstream called the locus control region (LCR) (Groudine et al. 1983; Kioussis et al. 1983; Tuan et al. 1985; Forrester et al. 1986; Forrester et al. 1987; Grosveld et al. 1987). Transgenic and knock-out studies have since demonstrated that high level transcription of the β-like globin genes is dependent on the LCR (Epner et al. 1998; Fraser and Grosveld 1998; Reik et al. 1998; Bender et al. 2000).

The LCR consists of five nuclease hypersensitive sites (HSs) dispersed over 20-30 kb of DNA located 10-60 kb upstream of the β-like globin genes (**Figure 1**). Among the five HSs, 5'HS2 and 5'HS3 have been shown to be particularly important for effective transcription of the β-globin genes, containing binding sites for various erythroid-specific and ubiquitous transcription factors such as NF-E2, GATA-1, EKLF, and SCL (Fraser et al. 1990; Philipsen et al. 1990; Morley et al. 1992; Orkin 1995; Ellis et al. 1996; Peterson et al. 1996; Hardison et al. 1997). It has been proposed that transcription factors bound to the HSs interact with each other to form a holocomplex which then facilitates/enhances transcriptional activation by recruiting chromatin remodeling proteins and interacting with β-globin promoters via looping of the intervening DNA, forming what is known as an Active Chromatin Hub (Tolhuis et al. 2002; de Laat and Grosveld 2003; Palstra et al. 2003; Patrinos et al. 2004). One of the key transcription factors involved in the process of regulating β-globin expression via the locus control region is MafK (Andrews et al. 1993a).

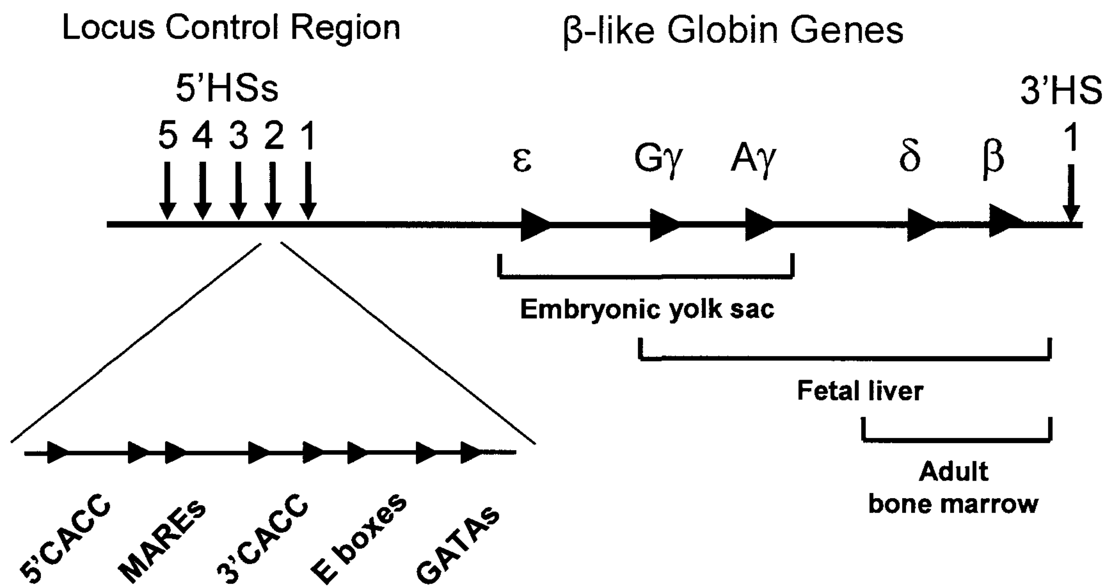


Figure 1: Organization of the human β-globin locus. The five β-like globin genes (green arrows) are arranged sequentially in the order of developmental expression. An upstream *cis*-regulatory element called the locus control region consists of 5 DNase1 hypersensitive sites (5′HSs) that contain a number of transcription factor binding sites (black arrows), including Maf recognition elements (MAREs).

1.4 MafK

Basic leucine zipper (bZIP) transcription factors play an important role in the regulation of transcription in eukaryotes. bZIP proteins contain both a basic region and a leucine zipper motif. The leucine zipper domain (an α-helix structure with leucine residues lined up on one side of the helix) mediates dimerization with other bZIP proteins by forming a coiled-coil structure via interactions between hydrophobic side-chains of parallel α-helices (Branden and Tooze 1999). The basic region mediates DNA-binding via hydrogen bonds and hydrophobic interactions between side-chains and DNA bases. Dimerization is a prerequisite for DNA-binding of bZIP proteins, while the identity of the dimerization partner affects DNA-binding specificity. This differential dimer-forming capability allows bZIP proteins to exert a more precise, combinatorial control of gene expression relative to regulation by individual proteins (Branden and Tooze 1999).

A variety of bZIP proteins are implicated in the transcriptional regulation of cellular differentiation, including the Maf family of proteins, which share relatively well-conserved bZIP motifs. The first Maf protein was identified as a component of avian musculoaponeurotic fibrosarcoma virus; this was called v-Maf, and its cellular homolog called c-Maf (Nishizawa et al. 1989; Kataoka et al. 1993). From these, the extended multigene family was discovered and its products classified as either large Maf proteins (c-Maf, MafB, NRL; 26-39 kDa) which contain a distinctive transactivation domain or small Maf proteins (MafK, MafF, MafG; 17-18 kDa) which lack this domain (Fujiwara et al. 1993; Kataoka et al. 1995). The

Mafs are able to form homodimers and can also heterodimerize with each other and a number of other bZIP transcription factors via LLLLYL (in large Mafs) or LLLMLL (in small Mafs) residues of the leucine zipper (Motohashi et al. 1997). DNA-binding studies have demonstrated that the Maf proteins preferably bind to sequences, known as Maf recognition elements (MAREs), that contain either a TPA (12-O-tetradecanoylphorbol 13-acetate)-responsive element (TRE; TGACTCA) or cAMP responsive element (CRE; TGACGTCA) flanked by conserved 5′-TGA and GCA-3′ sequences (Kataoka et al. 1994). Structural studies revealed that DNA-binding is mediated via NXXYAXXCR residues in the basic region interacting with the TRE or CRE core sequences, and that residues in an extended homology region outside of the basic domain recognize the flanking DNA-sequences for added DNA-binding affinity and/or specificity (Kusunoki et al. 2002).

Maf proteins are widely expressed throughout developmental stages and in a variety of tissues including neurons, the retina and hematopoietic cells, though with distinct spatial and temporal patterns of expression (Motohashi et al. 1997). A potential role for Mafs in β-globin expression and erythroid differentiation was first noted with the observation that the 5′HS2 of the β-globin LCR contained a sequence that resembles the TRE-type MARE (T-MARE) and was shown to be important in LCR function (Ney et al. 1990b; Talbot and Grosveld 1991). The sequence was known to bind a transcription activating factor, NF-E2 (nuclear factor – erythroid 2) (Ney et al. 1990a; Andrews et al. 1993b), which was shown by forced expression in NF-E2 null murine cells to be

essential for high level β-globin transcription and terminal erythropoiesis (Lu et al. 1994; Kotkow and Orkin 1995; Andrews 1998). Characterization of NF-E2 in murine cells demonstrated that functional NF-E2 is actually a dimer of the erythroid lineage-restricted 45 kDa bZIP factor NF-E2p45, a member of the cap-n-collar (CNC) family of transcription factors, interacting with MafK, also referred to as NF-E2p18 (Andrews et al. 1993b; Igarashi et al. 1994; Igarashi et al. 1995b; Blank et al. 1997). Reporter assays indicated that positive regulation of transcription by NF-E2 depends on the equilibrium concentrations of NF-E2p45 and MafK (Igarashi et al. 1994). Additional insights into the role of MafK were gained from the observation that forced expression of MafK (from stably transformed *mafK* under the control of an inducible promoter) resulted in the differentiation of MEL cells without any other inducers as well as accumulation of hemoglobin and globin gene transcripts, suggesting that quantitative control of *mafK* expression is important in β-globin gene expression and erythroid differentiation (Igarashi et al. 1995a; Francastel et al. 1997; Motohashi et al. 1997).

The role of MafK in β-globin transcriptional regulation was clarified in 2004 when Brand et al., using the ICAT quantitative proteomics strategy, defined MafK as a "dual-function" protein, switching from transcriptional repressor to activator by exchanging co-repressor complexes for co-activators at the β-globin LCR (Brand et al. 2004). This study demonstrated that prior to MEL cell differentiation, MafK interacts with the bZIP protein Bach1 (BTB and CNC homology 1), a member of the CNC family. The MafK:Bach1 heterodimer binds

to the MARE sequence at the β-globin LCR and recruits repressive cofactors, resulting in the active repression of β-globin gene activity. In response to DMSO-induced differentiation, MafK exchanges its dimerization partner from Bach1 to the erythroid specific NF-E2p45, forming the NF-E2 complex that is then able to activate β-globin transcription, hemoglobin synthesis and erythroid differentiation (Brand et al. 2004).

Another very interesting phenomenon regarding MafK biology was noted by immunofluorescence microscopy: There appears to be spatial reorganization of MafK within the nucleus during erythroid differentiation. Francastel et al. (2001) demonstrated that in undifferentiated MEL cells, a significant proportion of MafK associates with pericentromeric heterochromatin. Within 20h after induction of erythroid differentiation but before transcription of the β-globin gene, MafK is observed to be excluded from the pericentromeric heterochromatin. The position of Bach1 was not investigated, but immunostaining reveals that MafK and NF-E2p45 co-localize in the same subnuclear compartment only after differentiation, which is consistent with ICAT results showing them as dimerization partners after induction of differentiation (Brand et al. 2004). These results suggest that MafK moves from the heterochromatin to the euchromatin compartment once MEL cells are committed to erythroid differentiation, and the fact that it precedes β-globin gene activation suggests that subnuclear relocalization may represent an additional level of β-globin transcriptional regulation.

1.5 Heterochromatin

The DNA in the eukaryotic nucleus exists as a highly organized structure with varying degrees of organization (Becker et al. 1996). At the most basic level, DNA wraps around histone octamer consisting of two of each of the histones H2A, H2B, H3 and H4. About 147 bp of DNA wrap twice around each octamer, together called nucleosomes, with linker DNA between nucleosomes, resulting in a structure typically analogized as 'beads on a string'. The chromatin becomes more compact with the addition of histone H1 which brings nucleosomes close together to form a helical structure called a 30nm fibre. This is typically the structure of DNA that is transcriptionally active, referred to as euchromatin; it is an open-conformation of DNA that is accessible to the transcriptional machinery (Nemeth and Langst 2004). Greater compaction of the chromatin over various scaffolding proteins is associated with gene inactivation as it reduces the accessibility of the DNA, forming a transcriptionally less-permissive structure called heterochromatin. A variety of mechanisms regulate the degree to which genetic information is compacted (e.g. post-translational modification of histones); this represents yet another level of transcriptional regulation in eukaryotes (Nemeth and Langst 2004).

Heterochromatin can be categorized as either facultative or constitutive. Facultative heterochromatin refers to chromatin which is reversibly or impermanently condensed; it is a dynamic structure, differing between cell types and changing with developmental progression (Becker et al. 1996). It is the

heterochromatin normally associated with inactivation of genes by heterochromatinization. Constitutive heterochromatin, on the other hand, remains heterochromatin throughout development and is the same in all cell types of a given species. In mammals, the majority of constitutive heterochromatin is found at centromeres and surrounding areas (i.e. pericentromerically) and consists of genetically inactive, highly repetitive sequences of DNA called satellite repeats. The mouse genome contains at least two types of repetitive elements at centromeres, the major satellite and minor satellite repeats (**Table 1**) (Guenatri et al. 2004). While its compact nature plays a structural role in the process of sister chromatid segregation during mitosis, (peri)-centromeric heterochromatin remains condensed throughout the cell cycle and can be identified in murine nuclei by light microscopy as regions of deep DAPI staining (Guenatri et al. 2004). Constitutive heterochromatin is also known to play a role in regulating gene expression by a phenomenon called position-effect variegation (PEV), where rearrangements that place euchromatic DNA near/into the gene-poor, transcriptionally-repressive heterochromatic nuclear compartment lead to a variegated pattern of gene expression (Girton and Johansen 2008). Heterochromatin protein 1α (HP1α), a protein that associates with pericentromeric heterochromatin through interactions with methylated core histones, is thought to play an essential role in heterochromatin formation and PEV (Eissenberg et al. 1992; Nielsen et al. 2001).

Recent evidence indicates that there may be yet another way in which heterochromatin is involved in the regulation of gene expression, namely by regulating the spatial/local concentrations of transcription factors and/or their

Table 1 – Characteristics of Murine Constitutive Heterochromatin

	Major Satellite	Minor Satellite
Repeat Unit Length	234 bp	120 bp
Copy Number	>200,000	>50,000
Fraction of Genome	~3.00%	~0.5%
Chromosome Region	Pericentromeric	Centromeric

proximity to target genes. This arises from the observation that several transcription factors in various cell types have been identified as associating with the heterochromatic regions of the nucleus, an unexpected phenomenon given the gene-poor nature of constitutive heterochromatin. MafK is one such example as it associates with pericentromeric regions in erythroleukemia cells before, but not after, differentiation. Deletion of the leucine-zipper motif of MafK results in its delocalization from pericentromeric heterochromatin, implicating protein-protein interactions in its heterochromatic localization (Francastel et al. 2001). Another example is the bZIP transcription factor HBZ produced in HTLV-I-infected T-cell leukemia cell lines which was shown to localize to heterochromatic foci via three nuclear localization signal domains (Hivin et al. 2005). TLX1/HOX11, a homeodomain transcription factor essential for spleen

development and implicated in the development of T-cell acute lymphoblastic leukemia, demonstrates heterochromatic localization by direct binding to satellite DNA sequences that resemble the known TLX1/HOX11 binding motif (Heidari et al. 2006). Ikaros, a member of the Kruppel-like zinc finger family of transcription factors that plays an important role early hematopoietic and lymphoid differentiation, also localizes to pericentromeric regions in immature lymphocytes where it may play a role in gene silencing via PEV (Brown et al. 1997; Hahm et al. 1998; Klug et al. 1998). Much like TLX1/HOX11, Ikaros binds directly to pericentromeric DNA sequences (Cobb et al. 2000).

Yet another interesting example is the transcriptional intermediary factor 1β (TIF1β) which was shown to become localized to pericentromeric heterochromatin via interaction with HP1α upon induction of early (embryonic) endodermal differentiation (Cammas et al. 2002). The transcription factor CCAAT-enhancer-binding protein (C/EBPβ) also acquires pericentromeric targeting upon induction of differentiation of preadipocytes (Piwien Pilipuk et al. 2003). Other examples of heterochromatic localization include immunoglobulins IgH and Igκ in lymphocyte development (Kosak et al. 2002) and high-mobility group (HMG) proteins I/Y in rat thyroid cancer cells (Martelli et al. 1998). It is interesting to note that most of these observations were made in cells of the hematopoietic lineage, indicating a possible link; nevertheless, very little is understood about the purpose of heterochromatic localization.

1.6 Statement of Intent

The spatial organization and/or local concentration of transcription factors within the nucleus relative to target genes may represent an additional level of transcriptional regulation during the process of stem cell differentiation. The observed subnuclear relocalization of MafK from heterochromatic to euchromatic regions provides an ideal context in which to explore this phenomenon as it coincides with the exchange of MafK's dimerization partner, activation of β-globin gene expression and erythroid differentiation. As a first step in clarifying the role of subnuclear relocalization in regulating β-globin gene expression during erythropoiesis, we chose to investigate the characteristics of MafK localization to regions of pericentromeric heterochromatin in undifferentiated erythroleukemia cells.

1.7 Hypothesis

Given that fact that MafK represses β-globin activity, is known to interact with Bach1 in undifferentiated proerythroblasts, and requires an intact leucine-zipper domain for pericentromeric targeting, we hypothesized that the MafK:Bach1 heterodimer at the β-globin LCR is retained to heterochromatin through protein-protein interactions and that this localization contributes to β-globin silencing before differentiation.

CHAPTER 2

MATERIALS AND METHODS

2.1 Cell Culture

Murine erythroleukemia (MEL) strain 745A cells were maintained at 37°C and 5% CO_2 in a suspension culture using spinner flasks (Bellco) with RPMI-1640 medium (Sigma R8758 containing L-glutamine) containing 10% fetal bovine serum (Hyclone), and 1% penicillin/streptomycin (Hyclone). MEL cells, which are halted in the undifferentiated proerythroblast stage, were harvested by centrifugation at 800 x g for 10 minutes at room temperature. Thereafter, cells diluted to a concentration of 0.2 x 10^6 cells/mL were induced to differentiate by 2% DMSO treatment (Friend et al. 1971). Differentiated cells were harvested four days after induction of differentiation.

2.2 Protein Extraction

Protein extraction consisted of the non-denaturing extraction of soluble nuclear proteins, termed the euchromatic extract (EE), followed by the non-denaturing extraction of proteins from the remaining insoluble nuclear fraction (nuclear pellet NP), resulting in the heterochromatic extract (HE).

To prepare euchromatic extracts, harvested MEL packed cell volume (PCV) was noted and cells were washed once with 5X PCV of cold (4°C) phosphate-buffered saline (PBS), and allowed to swell in 5X PCV of cold Buffer A (10 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [HEPES; pH 7.9 at 25 °C], 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM 1,4-Dithithreitol [DTT], 1X PIC [Protease Inhibitor Cocktail, EDTA-free tablets; Roche]) (Dignam et al. 1983) for 10 minutes on ice. Cells were collected, resuspended in 5X PCV of cold Buffer A, and lysed

by dounce using type B pestle in 15 mL Kontes glass tube. After centrifugation (25000 x g), the nuclear pellet volume (NPV) was noted and resuspended in 1X NPV of cold Buffer C (20 mM HEPES [pH 7.9 at 25 °C], 25% (v/v) Glycerol, 1.5 mM $MgCl_2$, 600 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 1X PIC)(Dignam et al. 1983), and the nuclei homogenized by dounce. The nuclear suspension was stirred gently for 30 minutes on ice and then centrifuged at 25000 x g for 30 minutes at 4°C. The supernatant was kept on ice, and the extraction repeated two additional times using Buffer C′ (20 mM HEPES [pH 7.9 at 25°C], 25% (v/v) Glycerol, 1.5 mM $MgCl_2$, 300 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 1X PIC). The supernatants were combined and termed the euchromatic extract (EE). The volume of the EE was noted and the protein concentrations were determined by Bradford protein assay (Bio-Rad Laboratories).

The remaining nuclear pellet (NP) was resuspended in a volume of Buffer G (2 mM Triethanolamine-HCl [pH 7.6 at 25 °C], 0.05 mM $MgCl_2$, 150 mM KCl, 5 mM $CaCl_2$, 0.1% (v/v) NP40, 0.5 mM DTT, 1X PIC) equivalent to the EE volume. After homogenization using dounce, DNase 1 (1U/10ug DNA) (Roche) and RNase A (10ug/ul) (Sigma) were added and the mixture was transferred to siliconized tubes and incubated at 4 °C with constant rotation overnight. The mixture was centrifuged at 25000 x g for 30 minutes at 4°C and the supernatant was termed the heterochromatic extract (HE). The volume of the HE was equilibrated to the EE volume using Buffer G and protein concentrations were determined by Bradford protein assay (Bio-Rad Laboratories).

2.3 Silver Stain Analysis

Protein extracts were denatured at 95°C for 5 minutes in loading buffer (100 mM Tris-HCL pH 6.8, 4% SDS, 200 mM DTT, 20 % glycerol and 0.1% bromophenol blue) and resolved on 12% SDS-polyacrylamide gel (375 mM Tris-HCl [pH 8.8 at 25°C], 0.1% (v/v) SDS, 0.05 % (v/v) APS, 0.05% (v/v) TEMED) at a constant 30 mA. Precision-Plus protein standards (Bio-Rad Laboratories Inc.) were used to indicate molecular weight. After removing the stacking gel, the gels were fixed in 50% ethanol overnight. The gels were placed in staining solution (0.8 % (w/v) silver nitrate, 0.08% (v/v) sodium hydroxide, 1.6% (v/v) ammonia) for 15 minutes and then washed with ddH_2O for 5 minutes. Developing solution (0.25% (v/v) citric acid, 0.6% (v/v) formaldyhyde) was then added with gentle mixing until the desired level of stain was achieved, and the staining was halted with the stopping solution (50% (v/v) ethanol, 10% (v/v) acetic acid). The gels were stored by drying between cellophane sheets using the DryEase® Mini-Gel Drying System (Invitrogen) as per the manufacturers instructions.

2.4 Immunoprecipitation of MafK

MafK-specific rabbit polyclonal antibodies (Santa Cruz Biotechnology, sc-477) were crosslinked on a Dynabeads® Protein A resin (Invitrogen) with 20 mM (final concentration) dimethylpimelimidate (Sigma). Euchromatic extracts (EE) and heterochromatic extracts (HE) were pre-cleared with resin and then incubated with the MafK antibody–bound resin at 4°C for 12 h. Proteins bound to Dynabeads® Protein A resin–MafK antibody were then washed with IP buffer (25

mM Tris [pH 7.9 at 25°C], 5 mM $MgCl_2$, 100 mM KCl, 10% (v/v) glycerol, 0.1% (v/v) NP40, 0.3 mM DTT, 1X PIC) and antibody-bound proteins were eluted with either boiling SDS-PAGE loading buffer for subsequent silver stain and western blot analysis or elution buffer (0.1% (v/v) formic acid in 50% (v/v) acetonitrile) for downstream mass spectrometry analysis. A mock IP was done in parallel under the same conditions using normal rabbit IgG (Santa Cruz Biotechnology) instead of the MafK antibody.

2.5 Western Blot Analysis

Protein extracts as well as IP elutions were denatured at 95°C for 5 minutes in loading buffer and resolved on 12% SDS-polyacrylamide gel as previously described. Precision-Plus protein standards (Bio-Rad Laboratories Inc.) were used as molecular weight markers. Proteins were transferred to nitrocellulose membrane (Amersham Pharmacia Biotech) using the Mini-Protean II Electrophoresis Cell kit system (Bio-Rad Laboratories Inc.) at 150 V for 1 hour in cooled transfer buffer (0.06% (w/v) Tris base, 0.29% (w/v) glycine). Blots were incubated in primary antibody plus 5% skim milk powder in PBS for 1 hour at room temperature or overnight at 4°C on a rocking platform. Following 3 x 5 minute washings in 0.05% Tween in PBS, membranes were blotted with an appropriate anti-mouse or anti-rabbit secondary antibody conjugated to horseradish peroxidase (BioRad) for 1 hour at room temperature in PBS. The membranes were washed again in 0.05% Tween in PBS and the proteins were visualized by the HRP-chemiluminescent reaction using the Immobilin Western

ECL kit (Millipore, Billerica, Mass) and exposed on XAR film (Kodak). Antibodies against the following proteins were used for western blot analysis of extracts and IP elutions: MafK/NF-E2MafK (Santa Cruz Biotechnology, sc-477), Bach1 (A1-5, rabbit polyclonal antibody raised against a GST-Bach1 fusion protein residues 174−415, kind gift from Dr. K. Igarashi), HP1α (Abcam, ab9057), NF-E2p45 (Santa Cruz Biotechnology, sc-291), TBP (mouse monoclonal 2CI, kind gift from Dr. L. Tora), H2A/B (Cell Signaling Technology, #2578). Normal rabbit IgG (Santa Cruz Biotechnology, sc-2027) were used for controls where indicated.

2.6 LC-ESI-MS/MS

MafK-IP and mock-IP elutions in 50% (v/v) acetonitrile (ACN) + 0.1% (v/v) formic acid were assayed for protein concentration using the Bio-Rad protein assay kit (Bio-Rad Laboratories). Typically, eluates contained approximately 7.5 µg protein. The eluates were adjusted to pH 8.0 slowly with 50 mM ammonium bicarbonate (NH_4HCO_3). After adding DTT (200 mM in 50 mM NH_4HCO_3) to 10 mM final concentration, the samples were boiled for 10 minutes and incubated for 1 hour at room temperature to reduce disulfide bonds. For alkylation of cysteines, iodoacetamide (1 M in 50 mM NH_4HCO_3) was added to 40 mM final concentration, allowed to incubate for 30 minutes at room temperature in the dark, and then quenched for 30 minutes with additional DTT (1 M in 50 mM NH_4HCO_3) to a final concentration of 50 mM. The proteins were then proteolyzed using 1:15 (w/w) trypsin (Promega) in 50 mM NH_4HCO_3 overnight at 37°C. ZipTip™ (Millipore) containing C_{18} resin was then used to purify and

concentrate the peptides. Briefly, the sample was prepared by adding 0.1% (v/v) Trifluoroacetic acid (TFA) and the peptides were bound to the C_{18} by pipetting up and down 10 times slowly. The peptides were washed 5 times with 0.1% TFA, and the peptides then eluted using 10 µl of 50% ACN.

Peptides were then pressure loaded onto in-house prepared 10 cm x 75 µm fused silica microcapillary reversed phase columns equilibrated with HPLC Buffer A (0.1% (v/v) formic acid in water). The peptides were resolved by running 27 minute gradients from 2-40% HPLC Buffer B (0.1% (v/v) formic acid in ACN) at 0.3 µl/min and analyzed by electrospray ionization (ESI) tandem mass spectrometry (MS/MS) using an LTQ ion trap mass spectrometer (ThermoFinnigan). The mass spectrometer was set to scan from 400-1800 m/z followed by one data dependent MS/MS scan on the most abundant ion. Dynamic exclusion was set to exclude ions that had been selected for MS/MS analysis for 3 minutes with a mass window of 1 dalton. Peptides were identified by sequence database searching using the search algorithm SEQUEST (Eng 1994) and Mascot (Perkins et al. 1999). The significance of each peptide and protein identification from SEQUEST is estimated using the software tools Peptide Prophet (Keller et al. 2002) and Protein Prophet (Nesvizhskii et al. 2003).

2.7 Cloning, Constructs and Transfection

MafK/NF-E2p18 cDNA cloned into the mammalian expression vector pCS2-MT in-frame with six Myc epitope tags in the N terminus was kindly provided by Dr. Mark Groudine (Francastel et al. 2001). MafK and point mutants

MafK-R22E and MafK-L2PM4P cDNA cloned into the mammalian expression vector pEF-BOS (Mizushima and Nagata 1990) were provided courtesy of Dr. Kohsuke Kataoka (Kataoka et al. 1995). To construct a mammalian expression plasmid for the expression of FLAG-fused MafK variants, the coding sequences were first amplified from the pEF-BOS plasmid template by PCR using the Phusion DNA polymerase (New England Biolabs). The primers used for amplification had the following sequence: 5′-CGTAAGCTTATGACGACTAATCC CAAACCG-3′ (forward) and 5′-CGAGAATTCCTAGGATGCTGCTGAGAACGG-3′ (reverse). PCR was done at 94°C for 5 min, followed by 30 cycles at 94°C for 1 min, 55°C for 1 min and 72°C for 1 min, followed by a final elongation at 72°C for 10 min. After ensuring amplification of an appropriate sized fragment (0.45 kb) by 1% agarose gel electrophoresis, the inserts were ligated to the multiple cloning region of the p3XFLAG-CMV™-10 mammalian expression vector (Sigma-Aldrich) at the *Hind*III and *Eco*RI sites. The expression vectors were amplified by transforming *E. coli* strain DH5α and selecting for colonies by ampicillin resistance, followed by purification using GenElute™ HP Plasmid Maxiprep Kits (Sigma-Aldrich) and sequencing (StemCore Laboratories). MEL cells were electroporated in 4 mm cuvettes using the Bio-Rad Laboratories Inc. GenePulser XCell (260 V; 960 μF; Ω = ∞), transiently transfected using 10 μg of expression vector for ~3 x 10^7 cells. The cells were cultured as described above and harvested after 3 days for analysis by co-immunoprecipitation or immunofluorescence microscopy.

2.8 Co-immunoprecipitation of Flag/Myc-MafK

Euchromatic extracts and heterochromatic extracts were prepared as above using MEL cells transiently expressing either FLAG-tagged MafK, Myc-tagged MafK, or both; untransfected MEL cells were used as a control. The extracts were incubated with monoclonal anti-FLAG®M2 affinity gel (Sigma-Aldrich) at 4°C for 12 h. Proteins bound to the anti-FLAG®M2 antibody were then washed first with 0.3 mM KCl, and then with 0.1 mM KCl, in Buffer H (25 mM Hepes-KOH [pH 7.6 at 25°C], 0.1 mM EDTA, 0.1 mM EGTA, 2 mM $MgCl_2$, 20% (v/v) glycerol, 0. 02% (v/v) NP40, 1 mM DTT, 1X PIC). The antibody-bound proteins were eluted with boiling SDS-PAGE loading buffer for use in subsequent western blot analysis using the following antibodies: Monoclonal mouse anti-FLAG®M2 (Sigma-Aldrich); Polyclonal rabbit anti-c-Myc (Bethyl Laboratories Inc.); Polyclonal rabbit anti-MafK/NF-E2MafK (Santa Cruz Biotechnology sc-477).

2.9 Immunofluorescence Microscopy

Microscope slides (VWR Scientific, superfrost plus) were coated with poly-L-lysine (1 mg/mL in PBS) (Sigma-Aldrich) and allowed to dry. MEL cells transiently expressing FLAG-tagged MafK, MafK-R22E, or MafK-L2PM4P, were attached to the coated slides, rinsed gently with PBS, and fixed by incubation with 4% paraformaldehyde in PBS, 10 mM EDTA, and 2 mM $MgCl_2$ for 10 min at room temperature (RT). The slides were then washed twice in PBS and the cells were permeabilized with 0.5% Triton X-100 in PBS for 5 minutes. After washing again in PBS, the cells were blocked using 5% normal donkey serum (Jackson

ImmunoResearch Laboratories Inc.) in 0.1% (v/v) Tween 20 in PBS for 20 minutes at room temperature in a moist chamber. The cells were then incubated in a moist chamber overnight at 4°C with mouse monoclonal anti-FLAG M2 antibody (Sigma-Aldrich) in PBS/donkey serum at 1:1000. The slides were washed four times for 5 minutes each with 0.1% Triton X-100 in PBS and incubated with Texas Red® dye-conjugated donkey anti-mouse IgG (Jackson ImmunoResearch Laboratories Inc.) for 1 hour at room temperature in a moist chamber. After labeling, slides were washed with 0.1% Triton X-100 in PBS and mounted in harsdset antifade media (Vectashield, Vector laboratories) containing 1 μg/ml 4′,6-diamidino-2-phenylindole (DAPI). Data was collected from the fluorescence microscope using AxioVision digital image processing software (Carl Zeiss); representative results are shown in the figures.

2.10 Chromatin Immunoprecipitation (ChIP)

Heterochromatic extracts from MEL cell nuclei were prepared as previously described; however, the digestion was also conducted using MNase (1U/10μg DNA) instead of DNase 1 for varying lengths of time from 0.5 – 12 hours. HE termed 0h were prepared without DNA digestion. Note that ChIP experiments were conducted without formaldehyde-crosslinking of MEL cells and without sonication in order to preserve protein-DNA interactions in as near a native state as possible. MafK-specific rabbit polyclonal antibodies (Santa Cruz Biotechnology, sc-477) were bound to Dynabeads® Protein A resin (Invitrogen) without crosslinking, and resin-precleared HEs were incubated with the MafK

antibody−bound resin at 4°C for 12 h. DNA-bound MafK bound to Dynabeads® Protein A resin−MafK antibody were then washed with IP buffer and eluted using ChIP elution buffer (50 mM Tris [pH 8.0 at 25°C], 1 mM EDTA, 1% (v/v) SDS). A mock ChIP was done in parallel under the same conditions using normal rabbit IgG (Santa Cruz Biotechnology) instead of the MafK antibody. The DNA was extracted by phenol chloroform, precipitated with ethanol, and redissolved in 10 mM Tris [pH 7.4 at 25°C] for analysis by quantitative PCR.

2.11 Quantitative PCR

ChIP inputs and elutions for the varying lengths of MNase digestions of MEL cell nuclei were analyzed by real-time PCR using the ABI Prism 7500 sequence detector with TaqMan probes (Applied Biosystems). Each real-time PCR experiment was done in duplicate at 50 °C for 2 min and 94°C for 10 min, followed by 40 cycles at 94°C for 20s, 55°C for 20s and 72°C for 30s. Data were collected at 72 °C and analyzed using the 'standard curve' method (ABI). The standard curve was obtained using mouse genomic DNA. Primers and TaqMan fluorescent probes were selected using ABI Primer Express software and were all obtained from Integrated DNA Technology (IDT). The primers for major satellite DNA were 5'-TATGGCGAGGAAAACTGAAA-3' and 5'-TTCACGTCCTA AAGTGTGTAT-3' (Shestakova et al. 2004) and the probe was 5'-6FAM-ACA CTGAAGGACCTGGAATATGGCGA-BHQ-1. The generated 233-bp amplicon comprises the entire major satellite repeat unit less the final base pair. MafK binding to major satellite sequence was calculated by dividing the amount of the

sequence in the ChIP elution fraction by the amount in the input fraction. The value obtained for the mock ChIP was then subtracted from the enrichment value obtained with the MafK antibody in each experiment. The enrichments are expressed as a function of the highest enrichment obtained on the locus (set to 100%).

CHAPTER 3

RESULTS

While ICAT mass spectrometry studies indicated that MafK switches from a transcriptional repressor to transcriptional activator at the β-globin gene locus control region (LCR) (Brand et al. 2004) upon terminal erythroid differentiation, immunofluoresence studies observed a simultaneous relocalization of the majority of MafK from heterochromatic to euchromatic areas of the nucleus (Francastel et al. 2001). We chose to further investigate the subnuclear relocalization as it may represent an additional dimension of regulatory control over the expression of β-globin and subsequent erythroid differentiation. To this end, we conducted a series of experiments to further characterize the protein-protein and protein-DNA interactions involved in the localization of MafK to heterochromatin in undifferentiated murine erythrolukemia (MEL) cells.

3.1 MafK relocalizes away from heterochromatin upon differentiation.

In order to biochemically evaluate MafK protein-protein interactions in heterochromatin (as outlined in **Figure 2**), the first step was to devise a method of biochemically separating proteins that are enriched in euchromatin from those associated with heterochromatin. We hypothesized that these proteins could be roughly separated on the basis of solubility, as euchromatin is by definition in a more open conformation and proteins associated with euchromatic DNA are not as tightly bound, and thus more easily solubilized in high salt conditions, than those retained to the highly compact heterochromatin. In order to test this, we extracted the soluble nuclear proteins from both undifferentiated and differentiated MEL cells using standard high salt procedures for nuclear extract

preparation (Dignam et al. 1983); we called this soluble fraction the euchromatic extract (EE) and compared it with the remaining insoluble nuclear pellet (NP) by western blot analysis (**Figure 3**). TATA-binding protein (TBP), a general transcription factor that recognizes the repeated TATAAA-DNA motif which serves as an assembly signal for the basic transcription machinery, was enriched in the euchromatic extract both before and after differentiation. TBP can bind its recognition sequence only if the DNA is in a relatively open conformation, as in euchromatin but not heterochromatin. The histones H2A/H2B appeared only the insoluble nuclear pellet, indicating that this fraction was enriched in proteins tightly bound to DNA.

Western blot analysis demonstrated that MafK was enriched in the nuclear pellet before differentiation but not so after differentiation, where the majority of MafK appeared in the euchromatic extract. This move from the NP to the EE mirrored the immunostaining observations of Francastel et al. (2001) that MafK relocates from heterochromatin to euchromatin upon terminal erythroid differentiation. Thus, we were able to reproduce the subnuclear relocalization of MafK by separating proteins that are tightly bound to DNA (e.g. histones) from those that bind DNA in an open conformation and are more easily solubilized with high salt (e.g. TBP). This allowed us to roughly correlate the soluble nuclear fraction as enriched in euchromatic proteins and the insoluble nuclear fraction as enriched in heterochromatic proteins, thereby effectively separating euchromatic and heterochromatic proteins biochemically.

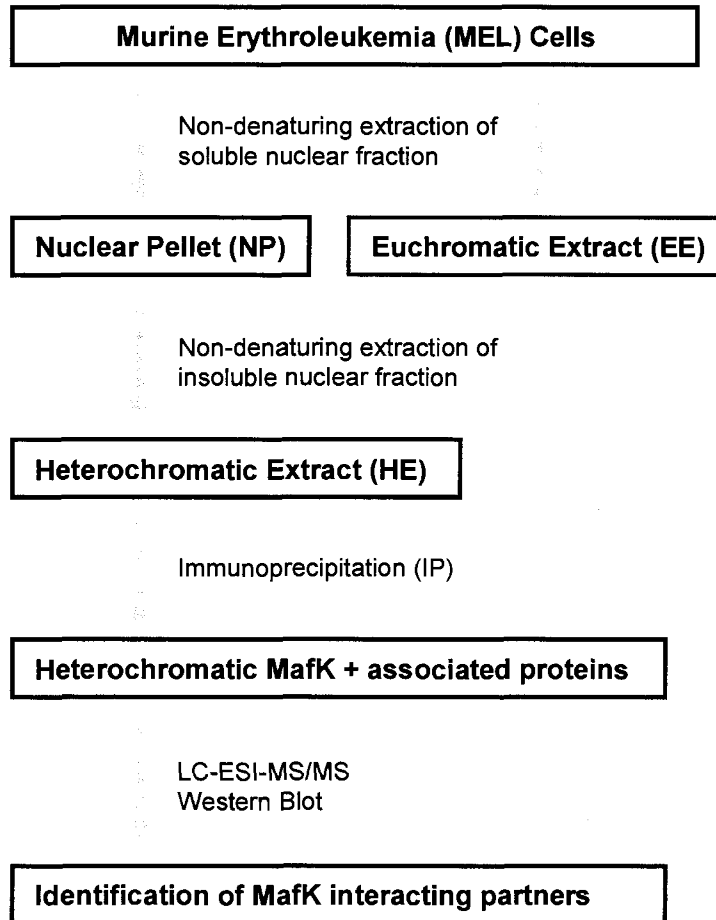


Figure 2. Examining MafK protein-protein interactions in heterochromatin. Flow chart outlining the approach taken for separating and extracting heterochromatic and euchromatic proteins from murine erythroleukemia (MEL) cell nuclei and comparing MafK interacting partners in each.

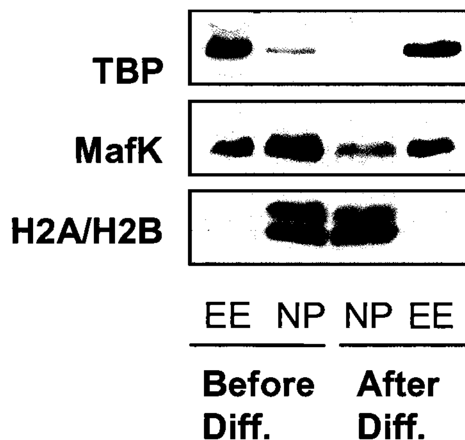


Figure 3. Specific decrease in heterochromatic MafK during erythroid differentiation. Suspension cultured MEL cells taken before and four days after DMSO-induced erythroid differentiation were lysed and nuclei extracted using standard non-denaturing protocols with 150 mM KCl (final); this was termed the euchromatic extract (EE). The EE and remaining insoluble nuclear pellet (NP) were equilibrated by volume, resolved using 15% SDS-PAGE and immunoblotted with antibodies specific for the proteins indicated on the left. TATA-binding protein (TBP) and histones H2A/H2B serve as markers for EE and NP, respectively. MafK is enriched in the NP before and the EE after differentiation. Thus, the separation of proteins tightly bound to DNA (e.g. H2A/H2B) from those less so (e.g. TBP) biochemically reproduced previously noted MafK subnuclear relocalization from heterochromatin to euchromatin upon erythroid differentiation.

3.2 Heterochromatic MafK does not interact with known heterodimerization partners.

After biochemically validating the separation method and immunostaining results, we then focused solely on undifferentiated MEL cell nuclei as our aim was to characterize the heterochromatic localization of MafK which occurs only before the induction of differentiation. To this end, we needed to extract proteins associated with heterochromatin from the insoluble nuclear pellet. As there were no well-established protocols for such an extraction, we devised a method to extract insoluble protein complexes from highly condensed heterochromatin into the soluble phase while preserving protein-protein interactions. Essentially, the nuclear pellet was homogenized by dounce in a buffer containing KCl and $CaCl_2$ and then treated with DNase 1 in order to release protein complexes from the compact heterochromatin. Final concentrations of 150mM KCl and 5mM $CaCl_2$ were determined to be optimal. After digestion and subsequent centrifugation, the supernatant containing solubilized proteins was termed the heterochromatic extract (HE).

Western blot analysis showed the MafK was effectively extracted from the nuclear pellet, and even though the HE contained approximately 1/10 of the total protein content of the corresponding EE (3.3 μg/μl in EE vs. 0.3 μg/μl in HE), the HE contains much more MafK than the EE (inputs in **Figure 4b**). The presence of heterochromatin-specific marker heterochromatic protein 1α (HP1α) in our heterochromatic extract, as well as the majority of transcription factors Bach1 and NF-E2p45 in the euchromatic extract, validated our protocol. However, it

seemed that there is some contamination of the heterochromatic extract with euchromatic proteins, given that there is some Bach1 and NF-E2p45 in the HE (inputs in **Figure 4b** - lanes 2 and 3). This is most likely due to the fact that after the euchromatic extract procedure, the nuclei are still structurally intact and therefore euchromatic proteins will not be extracted with complete efficiency.

Having successfully extracted heterochromatic and euchromatic proteins in non-denaturing conditions, we were then able to focus on investigating MafK protein-protein interactions in each. Brand et al. (2004) had identified MafK as interacting with Bach1 in undifferentiated MEL cells and NF-E2p45 after differentiation. Therefore, it was important to determine whether MafK in the heterochromatic fraction was associating with either of these proteins, as this would shed some light on its heterochromatic localization. To this end, we purified MafK and associated proteins in both the HE and EE by immunoprecipitation with MafK-specific antibody crosslinked to protein-A bound magnetic Dynabeads. Silver stained SDS-PAGE comparison of the MafK-IP and mock-IP elutions (**Figure 4a**) shows that the normal rabbit IgG did not pull-down any MafK, demonstrating the specificity of the MafK antibody (results confirmed by western blot analysis; data not shown). Analysis of the immunoprecipitation elutions by western blotting (**Figure 4b**) revealed that while MafK is associated with its predifferentiation β-globin repression partner Bach1 in the euchromatic fraction (EE MafK IP elution), as expected based on the findings of Brand et al. (2004), it is not associated with either Bach1 or with its post-differentiation β-globin activation partner NF-E2p45 in the heterochromatic fraction (HE MafK IP

elution). Thus it appears that heterochromatic MafK is in fact distinct from euchromatic MafK in terms of its protein-protein interactions.

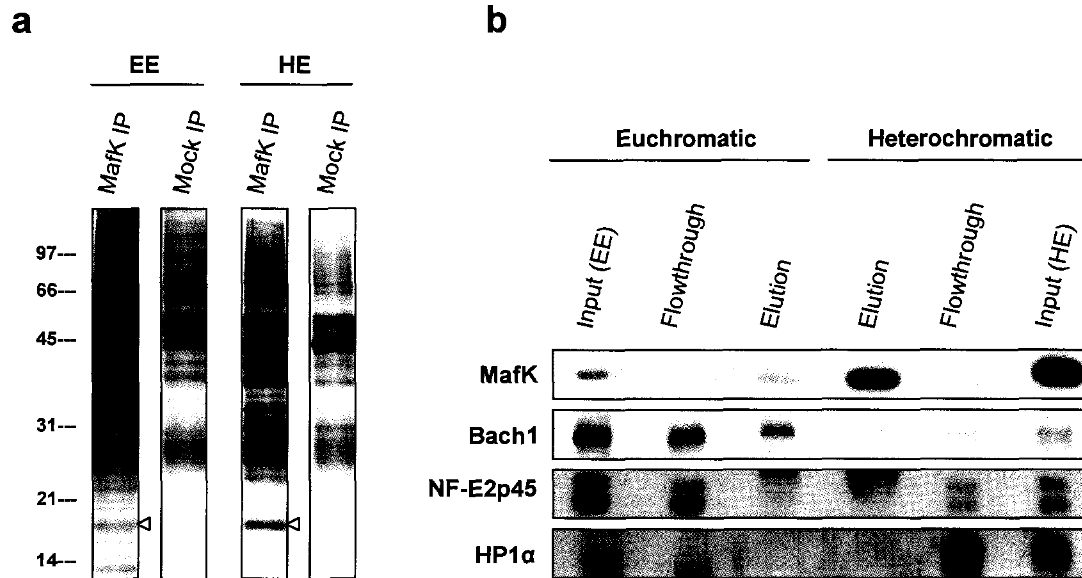


Figure 4. MafK in heterochromatin does not interact with known dimerization partners. Using undifferentiated MEL cells, heterochromatic extracts (HE) were prepared by DNase 1 digestion of the insoluble nuclear pellet (NP) that remained after euchromatic extract (EE) preparation. EE and HE were incubated with anti-MafK antibodies crosslinked to Protein-A dynabeads; normal rabbit IgG was used in parallel as a negative control. MafK and associated proteins were eluted by boiling in SDS buffer and resolved using 12% SDS-PAGE. (a) Silver stained SDS-PAGE gels of immunoprecipitation (IP) elutions show that MafK, indicated by the arrow (◅), was not eluted in the mock IP, demonstrating the specificity of the MafK antibody (confirmed by western blot analysis not shown). Molecular weight marker values are indicated on the right in kilodaltons. (b) Western blot analysis comparing euchromatic and heterochromatic MafK IP fractions using antibodies specific for proteins indicated on the left indicates that heterochromatic MafK does not co-precipitate known dimerization partners Bach1 and NF-E2p45. Heterochromatic protein 1α (HP1α) was used as a marker for heterochromatin.

3.3 MafK homodimers localize preferentially to the heterochromatic fraction.

Given the lack of binding to known partners, we considered the possibility that MafK found in heterochromatin is involved in novel protein-protein interactions. In an attempt to identify the MafK dimerization partner, we analysed the MafK immunoprecipitation elutions from heterochromatin using microcapillary reverse phase liquid chromatography (LC) electrospray ionization (ESI) tandem mass spectrometry (MS/MS). The MafK IP was compared to a mock IP conducted in parallel using normal rabbit IgG instead of the MafK antibody as a negative control to identify non-specific proteins in the IP elution (silver stained SDS-PAGE analysis of fractions is shown in the HE part of **Figure 4a**). The advantage of tandem mass spectrometry is that peptides separated by mass-to-charge ratio are isolated, individually fragmented and again scanned, giving mass spectra on individual peptides. Various algorithms are then used to cross correlate the observed tandem mass spectra with projected theoretical mass spectra from a database base search of candidate peptides in order to identify the sequence of each peptide.

Using SEQUEST software to search the murine peptide database, we identified two MafK peptides in the HE MafK IP elution spanning 17.3% of the MafK amino acid sequence, with cross correlation scores indicating a probability of identification of 0.999 as per Proteinprophet software analysis (**Table 2**). MafK was not identified in the HE mock IP elution, confirming the specificity of the MafK IP (**Table 3**). Separate analysis of the MS/MS data against the database of

rabbit peptides identified IgG gamma chain C in both MafK and mock IP elutions (data not shown), confirming the mass spectrometry identification as rabbit antibodies were used in both IPs. A number of additional proteins were also identified, such as histone H4 and nucleolar protein 1; however, these proteins were identified in both MafK and mock IP elutions and therefore likely represented contaminant proteins. While two peptides from histone H2A variant Z (H2A.Z) were identified only in the MafK IP elution using SEQUEST (**Table 2**),- analysis of the data using the Mascot search engine identified one of the same peptides (present in both H2A and H2A.Z) in the mock IP elution (AGLQFPVGR; see **Table 3**). Thus, while an interaction between MafK and H2A.Z cannot be ruled out, the identification of a histone H2A peptide in the mock IP elution, as well as the identification of another histone (H4) as a non-specific protein, suggested to us that the detected MafK-H2A.Z interaction may be non-specific. Ultimately, the LC-ESI-MS/MS analysis indicated that novel protein-protein interactions are not involved in the heterochromatic targeting of MafK.

Given the mass spectrometry results, it appeared that one of two things could be occurring: Either the heterochromatic extraction and MafK purification conditions caused denaturation and loss of protein-protein interactions, or MafK localizes to heterochromatin as a homodimer. In order to examine the possibilities, we devised a co-immunoprecipitation experiment to test for MafK homodimerization. Undifferentiated MEL cells were transiently transfected to express FLAG-tagged MafK as well as Myc-tagged MafK, in addition to endogenous MafK. Heterochromatic and euchromatic extracts were prepared

from these MEL cells as before, which were then used for immunoprecipitation with FLAG M2-specific monoclonal antibody. The results were visualized by western blot analysis (**Figure 5**). The identification of Myc-MafK in the HE FLAG IP elution from cells expressing both FLAG- and Myc-tagged MafK indicated that MafK is indeed present as a homodimer in the heterochromatic fraction. The elutions from cells expressing FLAG-MafK or Myc-MafK alone were used as controls to demonstrate the specificity of the antibodies used for immunoprecipitation and western blotting. Interestingly, the HE FLAG IP elutions from cells expressing FLAG-MafK (lanes marked FLAG-MafK and FLAG+Myc-MafK) contained endogenous non-tagged MafK (Native-MafK), which was identified by MafK-specific antibody because it migrated faster in the SDS-PAGE than either of the tagged MafK. This further demonstrated homodimerization of MafK in the heterochromatin. The relative absence of Myc-MafK and presence of Bach1 in the corresponding EE FLAG IP elutions demonstrated that the observed homodimerization occurs preferentially when MafK localizes to heterochromatin. Thus, the co-immunoprecipitation not only demonstrated that MafK is present in heterochromatin as a homodimer, but also validated the non-denaturing heterochromatic extract protocol as well as the LC-ESI-MS/MS results.

Table 2 – Tandem mass spectrometry (LC-ESI-MS/MS) identification of proteins immunoprecipitated from the heterochromatic extract of undifferentiated MEL cell nuclei using MafK-specific antibody.

Protein/peptide	Charge state (z)	Cross-correlation score (xc)	Total Coverage (%)	Probability of Protein Identification
Nucleolar Protein 1			19.9	1.000
- LAAAILGGVDQIHIKPGAK	2+	4.824		
- VLYLGAASGTTVSHVSDIVGP DGLVYAVEFSHR	3+	6.514		
- VSISEGDDKIEYR	2+	2.087		
Histone H4			28.8	1.000
- ISGLIYEETR	2+	2.987		
- VFLENVIR	2+	2.998		
- DNIQGITKPAIR	2+	2.972		
Histone H2A (variant Z)			20.3	1.000
- HLQLAIRGDEELDSLIK	3+	2.487		
- AGLQFPVGR	2+	2.492		
MafK			17.3	0.9999
- EAGENAPVLSDDELVSMSVR	2+	4.248		
- LELDALR	2+	1.899		
ATP synthase (subunit alpha), mitochondrial			2.7	0.9945
- ILGADTSVDLEETGR	2+	3.037		
ATP synthase (subunit beta), mitochondrial			3.6	0.7758
- VLDSGAPIKIPVGPETLGR	2+	2.689		

Table 3 – Tandem mass spectrometry (LC-ESI-MS/MS) identification of proteins immunoprecipitated from the heterochromatic extract of undifferentiated MEL cell nuclei using normal rabbit IgG.

Protein/peptide	Charge state (z)	Cross-correlation score (xc)	Total Coverage (%)	Probability of Protein Identification
Nucleolar Protein 1			14.1	1.000
- VLYLGAASGTTVSHVSDIVGP DGLVYAVEFSHR	3+	7.102		
- VSISEGDDKIEYR	2+	2.502		
ATP synthase (subunit O), mitochondrial			6.6	0.9987
- LVRPPVQVYGIEGR	3+	3.630		
Ki-67 cell proliferation antigen			0.6	0.9954
- ERPQSPGKQESPGITPPR	3+	3.411		
Ribosomal Protein L38			18.6	0.9834
- KIEEIKDFLLTAR	2+	3.753		
Histone H4			12.1	0.8888
- VFLENVIR	2+	2.323		
Histone H2A			7.1	---
- AGLQFPVGR **	2+	---		

**** Identified by Mascot online MS/MS ion search engine. Peptide score = 61; scores > 54 indicate significant ion match ($p<0.05$).**

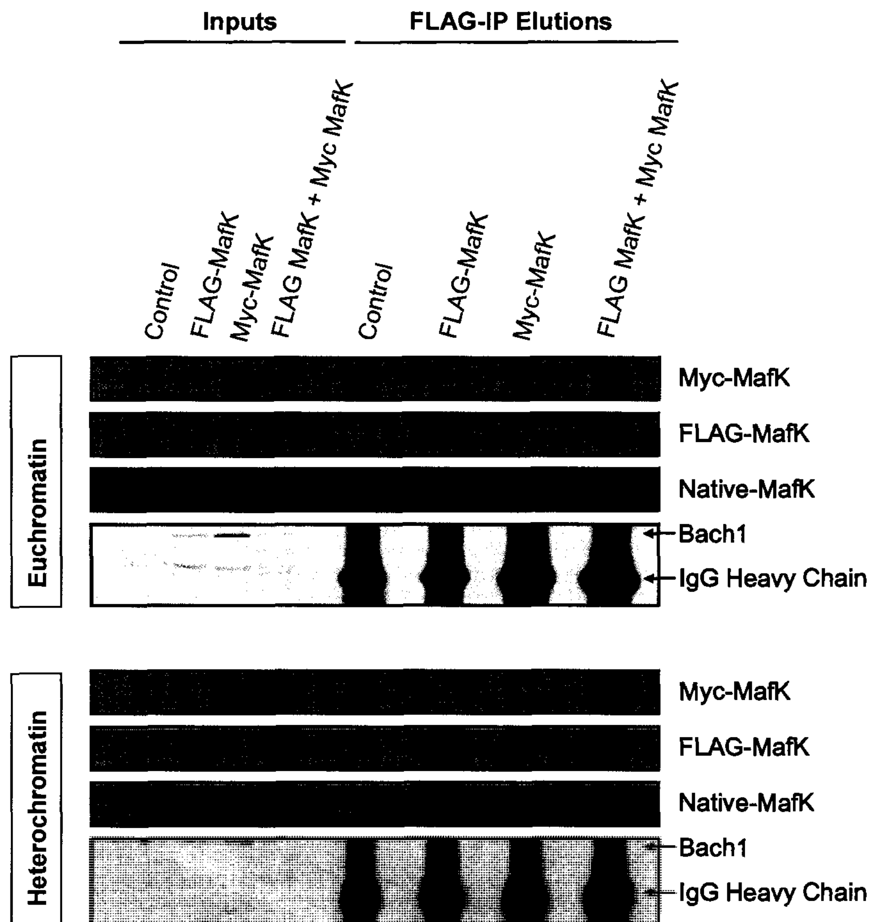


Figure 5. MafK homodimers are preferentially associated with the heterochromatic fraction. Undifferentiated MEL cells were transiently transfected to express FLAG-tagged MafK, Myc-tagged MafK, or both; untransfected MEL cells were used as a negative control for the transfections. Heterochromatic and euchromatic extracts prepared from these MEL cells were incubated with FLAG M2-specific antibody conjugated to Protein-A sepharose beads. Proteins were eluted by boiling in SDS buffer and resolved with 12% SDS-PAGE. Western blot is shown here with proteins indicated on the right revealed by specific antibodies. Cells expressing only Myc-MafK served as a negative control for the IP; cells expressing only FLAG-MafK served as a positive control. The presence of Myc-MafK as well as endogenous MafK in the FLAG-IP elution from the FLAG-MafK + Myc-MafK expressing cells indicates homodimerization, which appears to occur preferentially in the heterochromatin. Bach1 co-precipitates with FLAG-MafK only in the euchromatic FLAG-IP elutions.

3.4 A functional DNA-binding domain is critical to MafK heterochromatic localization.

Francastel et al. (2001) demonstrated by immunostaining that deletion of the leucine zipper motif of MafK resulted in its delocalization away from pericentromeric heterochromatin, indicating that protein-protein interactions were involved in the phenomenon. However, since dimerization is a prerequisite for the DNA binding of MafK, this experiment failed to rule out protein-DNA interactions as a mechanism for heterochromatic localization. Furthermore, given our observations that MafK forms homodimers in heterochromatin but does not appear to be associated with other proteins, it seemed unlikely that protein-protein interactions alone could account for its heterochromatic localization. Therefore, we chose to proceed by investigating whether protein-DNA interactions of MafK are involved in its localization to pericentromeric heterochromatin of MEL cells before terminal erythroid differentiation.

The first step was to identify whether the DNA binding domain of MafK is necessary for heterochromatic localization. Here we made use of MafK mutants MafK-L2PM4P and MafK-R22E (**Figure 6a**). MafK-L2PM4P contains two point mutations within the leucine zipper motif, thereby disrupting the interaction between two leucine zippers and inhibiting dimerization. MafK-R22E carries a mutation in the basic region of MafK and cannot bind to DNA, although they can form dimers through the intact leucine zipper domains (Kataoka et al. 1995). We cloned MafK, MafK-L2PM4P and MafK-R22E separately into the p3XFLAG-CMV-10 mammalian expression vector which were then used to transiently

transfect MEL cells. The cells were then immunostained with FLAG-M2 specific antibody and the results visualized by immunofluorescence microscopy (**Figure 6b**). In cells expressing FLAG-MafK, there was a clear overlap in fluorescence of regions stained with FLAG-M2 antibody and regions of dense DAPI staining, which identifies highly compact (peri)-centromeric heterochromatin. This verified previous immunostaining results and served as a positive control confirming the accuracy of our immunostaining protocol. The results for cells expressing FLAG-MafK-L2PM4P demonstrated that inhibition of MafK dimerization (and thus DNA binding as well) results in the delocalization of MafK from heterochromatin. This too confirmed previous results with leucine zipper-truncated MafK and served as a negative control for our experiment. Finally, the lack of overlap of DAPI and FLAG-M2 immunostaining in cells expressing FLAG-MafK-R22E revealed for the first time that an intact DNA binding domain is critical to MafK heterochromatic localization.

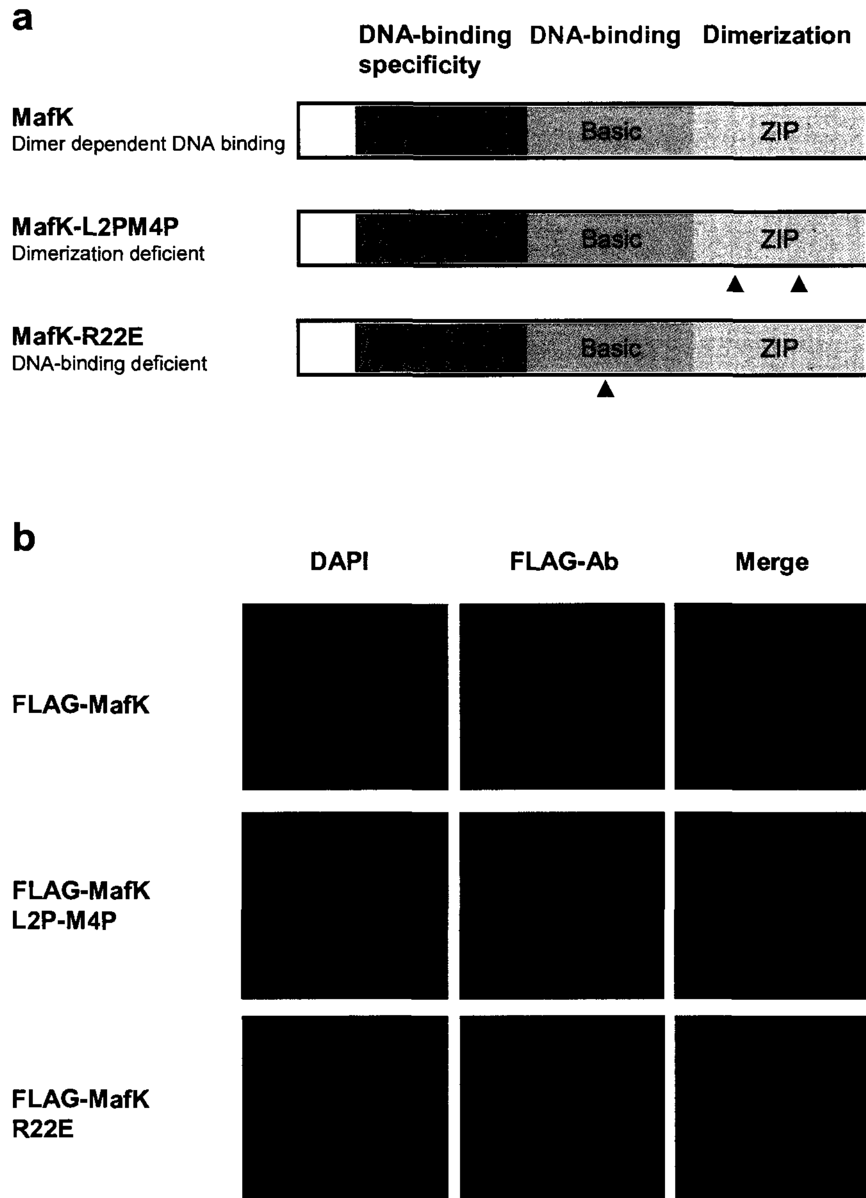


Figure 6. MafK DNA-binding domain is required for heterochromatic localization. (a) A schematic representation of MafK indicating the basic and leucine zipper (ZIP) domains, as well as the extended homology (Ext. Hom.) region. Mutations L2PM4P and R22E disrupt MafK dimerization and DNA binding, respectively. (b) Undifferentiated MEL cells were transiently transfected with FLAG-tagged MafK, MafK-L2PM4P and MafK-R22E, affixed to poly-L-lysine coated slides and immunostained with FLAG-M2 specific antibody. Immunofluorescence microscopy shows a clear overlap in fluorescence of regions stained with FLAG-M2 antibody (red) and regions of dense DAPI staining (blue) in cells expressing FLAG-MafK, while cells expressing FLAG-MafK-L2PM4P and FLAG-MafK-R22E show no such overlap.

3.5 MafK interacts with pericentromeric heterochromatin major satellite repeats.

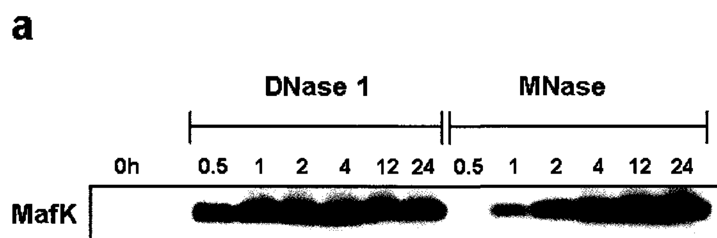
Having demonstrated that the DNA binding domain of MafK is involved in its heterochromatic localization by overexpression of a DNA-binding deficient MafK mutant, we investigated the interaction of native MafK with heterochromatin in MEL cells. To this end, we chose to conduct native chromatin immunoprecipitation (ChIP) experiments, the first step in which was to prepare extracts from MEL cell nuclei that would permit us to precipitate MafK along with any associated DNA. In the process of creating an effective procedure for extracting heterochromatic proteins from the insoluble nuclear pellet at the outset of this study, we had experimented with the use of micrococcal nuclease (MNase) instead of DNase 1 and had noticed that the release of MafK from the insoluble nuclear pellet was much lower with MNase than with DNase 1 (data not shown). Based on those preliminary observations, we decided to compare heterochromatic extracts prepared using equivalent quantities of DNase 1 and MNase for varying digestion times. Western blot analysis of the HEs revealed that the release of MafK into solution was related to the degree of MNase digestion, while the potent and indiscriminate nuclease action of DNase1 resulted in an almost complete release of MafK within 30 minutes of digestion (**Figure 7a**). Agarose electrophoresis analysis of the chromatin from the MNase treated samples confirmed that increased MNase digestion resulted in increased concentration of chromatin in solution, as well as the accumulation of mono- and

di-nucleosomal DNA (**Figure 7b**). These results underscored the strong interaction between MafK and heterochromatin.

Heterochromatic extracts prepared with MNase digestion for 0, 0.5, 1, 2, 4 and 12 hours were then used as inputs in the immunoprecipitation using MafK-specific antibody. MafK and associated chromatin were then eluted and the DNA was purified by phenol chloroform extraction and ethanol precipitation. This DNA was then used as the template in qPCR amplification using primers designed specifically for the *Mus musculus* major satellite DNA of pericentromeric heterochromatin. A mock-ChIP conducted in parallel using normal rabbit IgG was subtracted from the MafK-ChIP qPCR values to reveal specific binding, and the specific binding was expressed as a function of input (**Figure 8**). The ChIP results for the 0h digest, where little MafK was released in solution, showed the non-specific binding of major satellite DNA. As the amount of MafK increased (with longer MNase digestion), there was a concomitant increase in the amount of major satellite DNA in the ChIP elution relative to the input HE, indicating an increase in MafK binding to pericentromeric heterochromatin. The apparent MafK binding decrease noted at 12 hours was likely due to degradation of the 234-bp major satellite repeat unit by continued MNase digestion as evidenced by the increase in mono-nucleosomal DNA (**Figure 7b**), resulting in less major satellite sequence amplification by qPCR. The observed interaction between MafK and pericentromeric heterochromatin is likely due to direct DNA binding since native ChIPs, conducted without chemically cross-linking proteins to DNA before the ChIP, do not typically identify indirect protein-DNA interactions that

occur through weaker interactions such as cofactor binding (Brand et al. 2008). Overall, these results strongly suggest that MafK binds directly to pericentromeric heterochromatin major satellites in undifferentiated MEL cells.

a



b

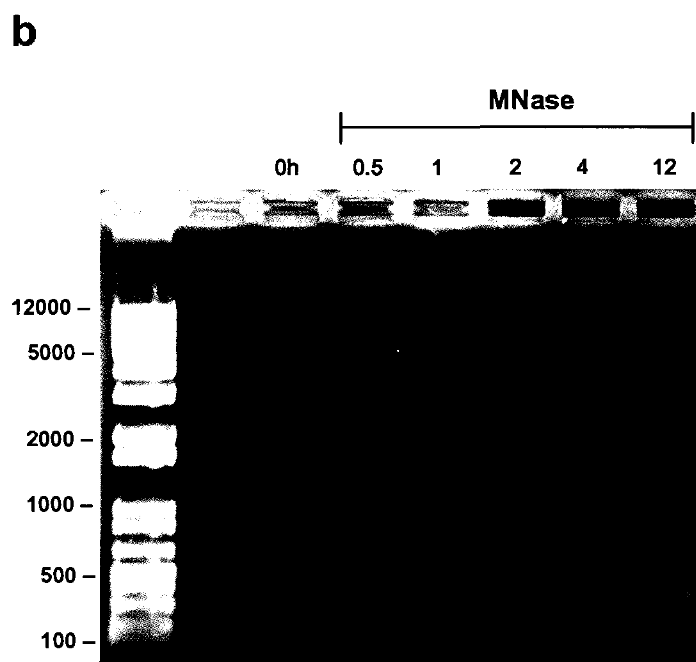


Figure 7. MafK interacts strongly with heterochromatin. Heterochromatic extracts (HEs) were prepared from undifferentiated MEL cells using either DNase 1 or MNase digestion of the insoluble nuclear pellet (NP); HE prepared without digestion (0h) served as a control. (a) Western blot analysis reveals that the amount of MafK released from heterochromatin is directly related to the length of MNase digestion time. (b) DNA was purified from the MNase digested HE samples 0-12h by phenol chloroform and resolved by agarose gel electrophoresis. Increased MNase digestion results in increased concentration of chromatin in solution, as well as the accumulation of mono- and di-nucleosomal DNA (bands at approx. 150 and 300 bp, respectively).

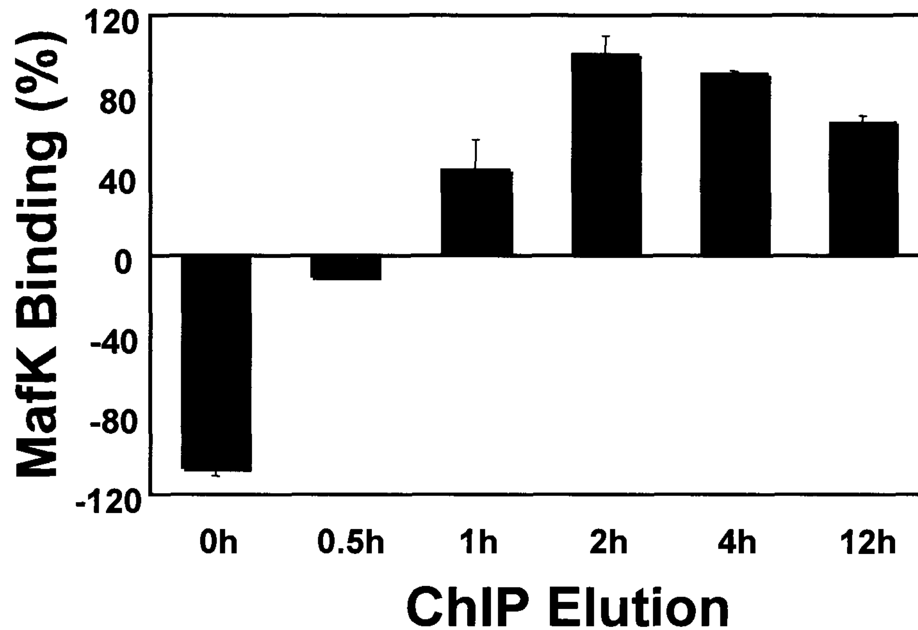


Figure 8. MafK interacts directly with pericentromeric chromatin. Native-ChIP was conducted on undifferentiated MEL cells by preparing heterochromatic extracts with MNase digestion from 0-12h and incubating with MafK-specific antibody conjugated to Protein-A Dynabeads. MafK and associated chromatin were eluted and the DNA was then extracted with phenol chloroform and used as input in qPCR amplification using primers for the major satellite repeat. A mock-ChIP was conducted in parallel using normal rabbit IgG. The bar graph shows MafK binding to major satellite sequence as calculated by subtracting the amount of the amplified major satellite sequence in the mock-ChIP elution from values for the MafK-ChIP elution and then dividing by the input fraction values. The enrichments are expressed as a function of the highest enrichment obtained on the locus (i.e. 2h sample, set to 100%). Error bars represent standard error of the mean (SEM).

3.6 Heterochromatic satellite DNA appears to contain a novel MafK binding site.

Given our findings which indicate that MafK homodimers interact directly with pericentromeric heterochromatin in undifferentiated MEL cell nuclei, we questioned whether known MafK binding sites, termed Maf recognition elements (MAREs), were found in satellite DNA sequences. Classically, MAREs are palindromic sequences classified as either T-MAREs (TGATGACTCAGCA) or C-MAREs (TGATGACGTCAGCA). More recent studies have shown that the Maf proteins are able to recognize a much larger variety of binding sites than originally thought. In an elegant study using surface plasmon resonance (SPR) imaging in microarray format, Yamamoto et al. (2006) were able to not only elucidate a number of single base pair substitutions in the T-MARE sequence that can bind small Maf proteins, but also categorized those sequences according to binding preference for Maf homodimers or heterodimers, as well as ambivalent or silent MAREs (**Table 4**). It is important to note that the Maf binding site found in the β-globin LCR that binds MafK:Bach1 and MafK:NF-E2p45 dimers was correctly classified in this system as a heterodimer-oriented MARE.

In addition to a wide variety of binding sites, Yoshida et al. (2005) demonstrated that Maf dimers do not even require the complete MARE sequence; rather, they are able to bind a half-T-MARE site (TGATGAC) when it is preceded by a 5′-AT rich sequence. As heterochromatin is known to contain multiple stretches of AT-rich DNA, this appeared to be a reasonable possibility for MafK heterochromatin binding. Given the large number of potential MafK

binding sites, we decided on a bioinformatics approach to examine whether heterochromatic satellite DNA contains any of the known MAREs.

An NCBI BLAST search of the *Mus musculus* genome revealed that neither the full T-MARE or C-MARE nor the core TRE or CRE sequences are found in murine heterochromatin (data not shown). We then chose to search for the more recently discovered Maf binding sites. To this end, with assistance from C. Porter and M. Huska of the Ottawa Health Research Institute Bioinformatics Group, we devised a simple program to search for the half-MARE sequence with every combination of single nucleotide substitutions identified by Yamamoto et al. (2006). The program, named FindMatches.pl (see Appendix A), was also designed to indicate the identity of any successfully matched sequences, identify the location of the match according to nucleotide numbering, and categorize the MARE variant according to Maf binding preferences. As the MARE variants were originally divided into symmetric and asymmetric substitutions based on whether the corresponding nucleotide 5′ of the center (0) position was also mutated, we chose to keep this notation even though symmetric did not apply in our case as we were searching for only half of the MARE sequence. Even though the mutations were given in the 3′ half of the MARE, we took advantage of the palindromic nature of MAREs to ensure that matches in the 5′ half of the MARE would be indentified by searching both forward and reverse strands of heterochromatic repeats in the 5′ to 3′ direction.

Using FindMatches.pl to search murine heterochromatic major and minor satellite DNA sequences (see Appendix A) resulted in 12 matches; eight in major

satellite and four in minor satellite DNA (**Table 5**). Having narrowed down the possibilities, we then manually inspected the matches and found that none of the 12 sequences were MARE variants because each contained greater than one nucleotide substitution and Yamamoto et al. (2006) demonstrated Maf binding with only single nucleotide substitutions at a time and not combinations thereof. These results indicate that murine centromeric and pericentromeric heterochromatin do not contain any known Maf binding sites that would account for the interaction observed by chromatin immunoprecipitation. This then suggests that heterochromatin contains novel binding site(s) with an affinity for MafK homodimers, the identification of which remains to be determined.

Table 4 – Maf recognition element (MARE) single nucleotide substitution variants[a]

Consensus of MARE		C / G	T	C	A	G	C	A
Nucleotide position		0	+1	+2	+3	+4	+5	+6
Group I	Symmetric		g / a	t	t / g	a	t / g	
	Asymmetric							
Group II	Symmetric			a				t / c
	Asymmetric		g / c/ a	t / g	t / g / c			
Group III	Symmetric	a / t						g
	Asymmetric			a		a	a / t / g	t / g / c
Group IV	Symmetric					t		
	Asymmetric					c / t		
Group V	Symmetric		c	g	c	c	a	
	Asymmetric							

[a]Adapted from Yamamoto et al. (2006)
MARE binding preference legend:
Red – homodimer oriented
Blue– heterodimer oriented
Black– prohibitive
Green – ambivalent

Table 5 – Results of search for known Maf recognition element (MARE) variants in murine heterochromatic satellite DNA using Findmatches.pl

MARE Category		Major Satellite		Minor Satellite	
		Forward Strand	Reverse Strand	Forward Strand	Reverse Strand
Group I	Symmetric	cactgta (44) ggcaaga (64) ggcgaga (180)		cactgta (19) cgttgga (65)	cattgta (9)
	Asymmetric				
Group II	Symmetric				
	Asymmetric		gtgtgca (73)		
Group III	Symmetric				
	Asymmetric	ctaaaaa (132)	ctaaagt (10) gtcaagt (126) ctaaatt (199)		ctaaaag (85)
Group IV	Symmetric				
	Asymmetric				
Group V	Symmetric	gcgagaa (181)			
	Asymmetric				

CHAPTER 4

DISCUSSION

The mechanisms involved in regulating the activation and/or repression of eukaryotic gene expression are critical components of the process of cellular differentiation whereby pluripotent cells become specialized in function. Transcription factors play a key role in the regulation of gene expression by binding to key *cis*-regulatory elements of genes and creating environments that either encourage or discourage the transcription of those genes. One such transcription factor is MafK, a small bZIP protein which switches from repressor to activator of β-globin gene expression during terminal erythroid differentiation. While MafK accomplishes this "switching" by the exchange of repressive cofactors with activating cofactors at the β-globin locus control region (LCR) (Brand et al. 2004), MafK is also observed to relocalize within the nucleus from heterochromatic to euchromatic regions upon differentiation (Francastel et al. 2001). As this intriguing phenomenon could represent an additional level of transcriptional control of β-globin expression and erythroid differentiation, we chose to investigate the characteristics of MafK localization to heterochromatin.

With the co-incidence of MafK subnuclear relocalization and its exchange of dimerization partners, we initially hypothesized that MafK was retained to heterochromatin through protein-protein interactions and that this localization contributed to β-globin silencing before differentiation. The observation that the MafK:Bach1 dimer interacts with a large complex of repressive proteins including chromatin remodelling proteins (Brand et al. 2004) indicated that perhaps the localization of MafK resulted from the heterochromatinization of the β-globin locus (**Figure 9a**). Alternatively, the repressor complex could include some

protein(s) that interacts with heterochromatin, thereby "tethering" the β-globin locus to heterochromatin via MafK protein-protein interactions (**Figure 9b**). Either one of these models would account for the overlap of both β-globin and MafK signals with DAPI stained regions in MEL cell nuclei noted by Francastel et al. (2001). It is to test this hypothesis that we started our investigation with the biochemical separation of euchromatic and heterochromatic protein complexes; by subsequently purifying MafK with its interacting partner(s) from the two fractions, we could compare euchromatic and heterochromatic MafK protein-protein interactions. We carried out separation on the basis of solubility in high-salt conditions, which separated nuclear proteins tighly bound to DNA, such as histones, from those less tightly bound in open (euchromatin) conformations, such as TBP. This definition, though perhaps somewhat simplistic, suited our needs as it reproduced MafK subnuclear relocalization (**Figure 3**), and was validated using markers such as heterochromatin protein 1α (HP1α), histones H2A/H2B, NF-E2p45, Bach1 and TBP (**Figures 3** and **4**). Future studies could potentially further characterize and compare the two extracts, for example by differentially labelled quantitative mass spectrometry, in order to further refine the methods of separating euchromatic and heterochromatic proteins.

Nevertheless, the immunoprecipitation of MafK and associated proteins from the two fractions revealed that MafK does not heterodimerize with Bach1 in the heterochromatic fraction (**Figure 4b**). This finding argued against our primary hypothesis and indicated that heterochromatic and euchromatic MafK are in fact distinct. This distinction was later clarified by our findings that

heterochromatic MafK exists as a homodimer, whereas euchromatic MafK heterodimerizes with Bach1 (**Figure 5**). Additionally, while Francastel et al. (2001) observed by fluorescence in-situ hybridization (FISH) that β-globin is associated with heterochromatin in undifferentiated MEL cells, recent studies using more sensitive techniques have demonstrated that the β-globin locus is in fact looping away from its chromosome territory in an open conformation and is "poised for activation" (Ragoczy et al. 2003). These observations are clearly inconsistent with the aforementioned models where heterochromatic MafK is not distinguished from MafK at the β-globin LCR.

Given that heterochromatic MafK appears to be independent of the β-globin locus, we considered another possibility regarding protein-protein interactions, that perhaps MafK forms novel interactions with proteins enriched in heterochromatin, such as HP1α, histones, or some other unknown protein(s). Mass spectrometry analysis, however, did not identify any other protein as specifically interacting with MafK. Co-immunoprecipitation of differentially tagged MafK demonstrated that heterochromatin is in fact enriched in MafK homodimers. The co-IP also addressed a concern regarding our heterochromatic extract protocol; as no protein (Bach1, p45, novel protein, etc.) had been identified as interacting with MafK, we questioned whether our novel extraction conditions disrupted protein-protein interactions. However, the elution of MafK homodimers under the same conditions (**Figure 5**) demonstrated that both the heterochromatic extraction and MafK immunoprecipitation protocols were indeed non-denaturing. Overall, our results indicated that protein-protein interactions

are not solely responsible for MafK heterochromatic localization. We were therefore revised our hypothesis and proceeded in a different direction: investigating protein-DNA interactions.

Considering the repetitive, gene-poor and highly-compact nature of pericentromeric heterochromatin, we initially considered it unlikely that MafK would interact directly with heterochromatin. An initial BLAST search revealed that the consensus Maf recoginition element (MARE) is not present in murine constitutive heterochromatin. Nevertheless, based on our observations regarding MafK protein-protein interactions, we proceeded to investigate the role of DNA binding in MafK heterochromatic localization. Immunostaining indicated that the DNA-binding basic region of MafK is necessary for its heterochromatic localization since the DNA-binding deficient MafK-R22E does not localize to heterochromatin (**Figure 6**). Interestingly, the dimerization-deficient MafK-L2PM4P displayed much more diffuse immunostaining than MafK-R22E, which had a few distinct foci of staining away from heterochromatin. Given that both mutants were unable to bind DNA, this difference in staining may be attributed to the intact leucine zipper motif of MafK-R22E, which may have facilitated its accumulation to particular foci through protein-protein interactions. Regardless of the nature of this accumulation, the fact remains that disruption of MafK's DNA binding domain resulted in its exclusion from pericentromeric regions, thus implicating it in MafK's heterochromatic localization.

We followed up these observations by chromatin immunoprecipitation experiments using MafK specific antibody on heterochromatic extracts prepared

with MNase digestion. Quantitative PCR amplification with primers specific for major satellite DNA revealed for the first time that MafK interacts with pericentromeric heterochromatin (**Figure 8**). ChIP analysis does not necessarily demonstrate direct DNA binding as would other experiments such as electrophoretic mobility shift assays (EMSA) because the interaction could theoretically occur indirectly through intermediaries such as cofactors. Such indirect interactions, however, are normally not identified by native ChIP, that is, ChIP without the customary formaldehyde crosslinking of proteins to DNA. Therefore, the fact that we observed an interaction using native ChIP strongly suggests that MafK homodimers bind directly to sequences within pericentromeric satellite repeats, though further studies are required to confirm the binding. Direct binding to satellite DNA has also been demonstrated for other transcription factors such as C/EBPα (Liu et al. 2007) and Ikaros (Cobb et al. 2000), though no distinguishing characteristics amongst the known heterochromatin-binding transcription factors have been elucidated thus far.

While ChIP analysis suggests binding to heterochromatin, subsequent bioinformatics search of satellite repeats using custom designed software revealed that constitutive heterochromatin does not contain any of the known Maf recognition elements. This implies that novel binding sequence(s) exist for MafK in pericentromeric satellite DNA, and future studies can be directed towards the identification of such site(s). Methods for binding site determination such as ChIP-on-chip or ChIP-seq could be utilized for this purpose, though the fact that satellite repeats themselves have a degree of internal repetition could present

some challenges to these approaches. Alternative methods include EMSA and/or SPR-microarray to narrow down the MafK binding site, followed by identification using a series of sequential nucleotide substitutions to demonstrate binding specificity. As these processes typically require relatively large quantities of purified protein, we optimized a non-denaturing protocol for the expression and purification of Histidine-tagged MafK from *E. coli* for use in future applications (data not shown). Experiments leading to the identification of apparently novel MafK binding site(s) could shed light not only on the phenomenon of heterochromatic localization but could also reveal novel locations within the genome where MafK is involved in transcriptional regulation.

Overall, our results indicate that in undifferentiated MEL cells, the majority of MafK is retained to the heterochromatic subnuclear compartment as homodimers interacting directly with novel binding sites within the heterochromatic satellite repeats. The finding that MafK interacts with heterochromatin as a homodimer is reasonable considering that heterochromatin is very compact and would not favour large protein complex formation. In addition, direct interaction with heterochromatin is supported by the fact that overexpression of MafK in (non-hematopoietic) NIH 3T3 cells also shows heterochromatic localization (Francastel et al. 2001) indicating that heterochromatin appears to be the default location for MafK binding. Overall, these results favour a model wherein heterochromatic localization serves as a mechanism for regulating nuclear MafK levels by retaining excess MafK away from transcriptionally sensitive sites (**Figure 9c**). In this model, Bach1 serves as

the limiting factor in undifferentiated MEL cells and excess MafK is isolated; this appears plausible as MafK homodimers are capable of passive repression and could potentially interfere with the MafK:Bach1-mediated active repression of β-globin expression. Once the cell is committed to terminal erythroid differentiation and Bach1 is sequestered by Heme (Sun et al. 2004; Tahara et al. 2004), NF-E2p45 then binds MafK in order to activate high level β-globin expression. Thus, heterochromatic localization can be considered a mechanism of "regulating the regulator," carefully controlling MafK availability which in turn modulates β-globin expression during differentiation. Such a model supports assertions that small Maf proteins function as "quantity-dependant bidirectional transcriptional effectors" based on observations in small Maf null mice (Nagai et al. 1998; Motohashi et al. 2000).

While the focus of this study was the characterization of MafK heterochromatic localization prior to differenatiation, future studies can be directed towards characterizing the molecular mechanism involved in the relocalization of MafK away from heterochromatin upon commitment to differentiation. One possibility is that differentiation triggers an increase in NF-E2p45 concentration such that MafK concentration becomes the limiting factor in NF-E2 formation; this could account for the lack of heterochromatic MafK after differentiation. Another possibility is that MafK is actually "released" from heterochromatin upon differentiation, for example by post-translational modifications that affect the dimerization and/or binding site preference of MafK. In fact, studies with the closely related MafG demonstrated that acetylation of

lysine residues in the DNA-binding domain increases the DNA binding affinity of the MafG:NF-E2p45 dimer (Hung et al. 2001). Small Maf proteins are also amenable to sumoylation, though it appears that this post-translational modification does not affect dimer formation and DNA recognition, but rather influences the ability of small Maf homodimers to recruit repressor complexes (Motohashi et al. 2006). Indeed, very recent results demonstrate that phosphorylation regulates the recruitment of the transcription factor Ikaros to pericentromeric regions in lymphocytes (Gurel et al. 2008; Popescu et al. 2009). Differing affinities of MafK homo- and hetero-dimers for MARE variants are also likely to play a role in the relocalization; increasing the affinity for heterodimer formation could influence relocalization of MafK away from the apparently homodimer oriented heterochromatic MARE to the heterodimer oriented MARE located at the β-globin LCR. As MafK dimerization is a prerequisite for DNA-binding, future studies will have to take both of these factors into account when characterizing the molecular basis of MafK subnuclear relocalization and clarifying its role in the transcriptional regulation of β-globin expression during terminal erythroid differentiation.

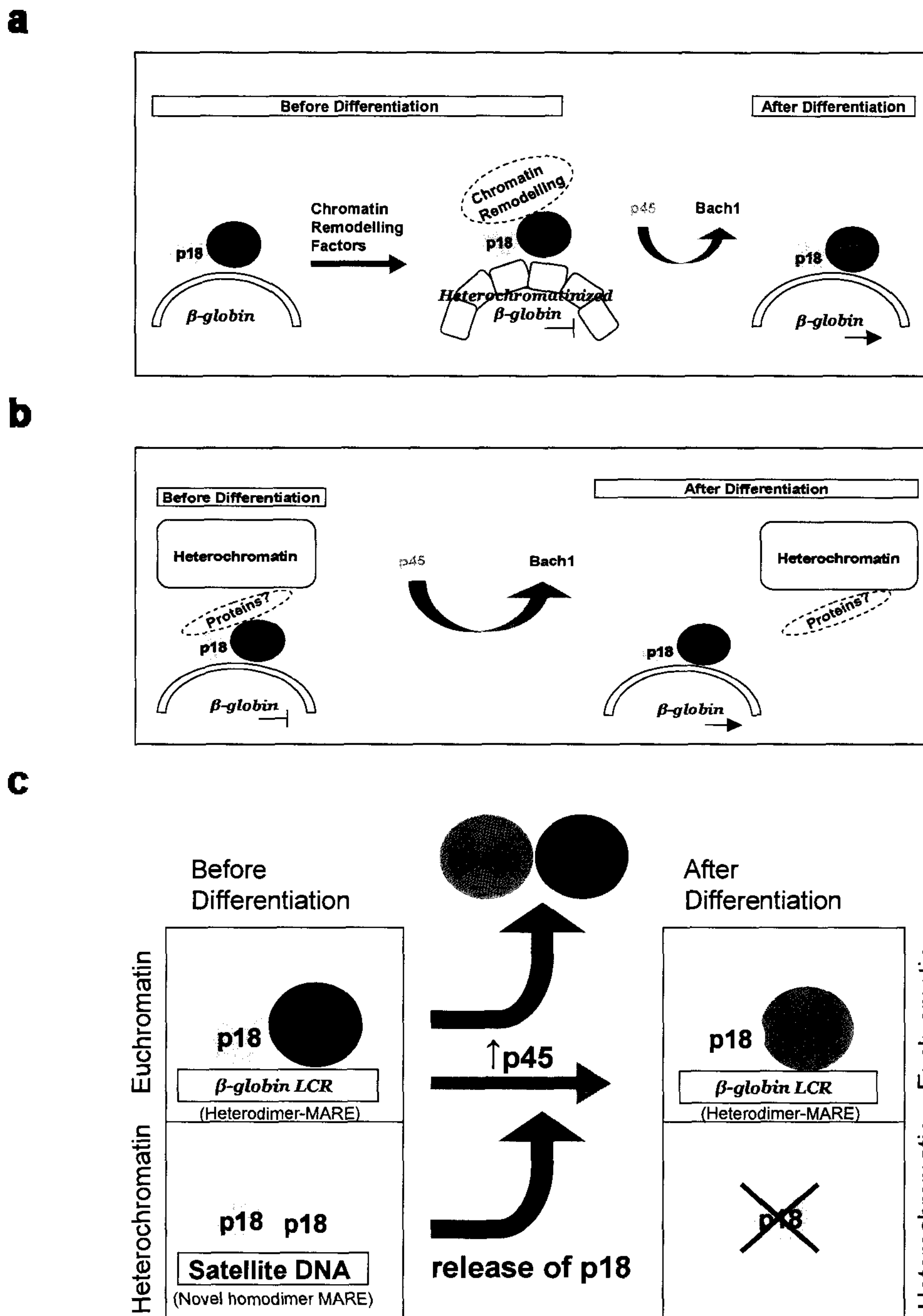


Figure 9. Possible roles of MafK subnuclear relocalization in β-globin silencing before erythroid differentiation. While we initially considered that MafK (p18) bound to the β-globin LCR would contribute to the transcriptional repression by either heterochromatinization of β-globin (a) or localization of the locus to transcriptionally repressive nuclear domains (b), our results support a role for subnuclear relocalization as a mechanism of precisely regulating the spatial distribution and local concentration of MafK during erythropoiesis (c).

References

Alter, B.P. 1994. Biology of erythropoiesis. *Ann N Y Acad Sci* **731**: 36-47.

Andrews, N.C. 1998. The NF-E2 transcription factor. *Int J Biochem Cell Biol* **30**(4): 429-432.

Andrews, N.C., Erdjument-Bromage, H., Davidson, M.B., Tempst, P., and Orkin, S.H. 1993a. Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature* **362**(6422): 722-728.

Andrews, N.C., Kotkow, K.J., Ney, P.A., Erdjument-Bromage, H., Tempst, P., and Orkin, S.H. 1993b. The ubiquitous subunit of erythroid transcription factor NF-E2 is a small basic-leucine zipper protein related to the v-maf oncogene. *Proc Natl Acad Sci U S A* **90**(24): 11488-11492.

Bank, A., Mears, J.G., and Ramirez, F. 1980. Disorders of human hemoglobin. *Science* **207**(4430): 486-493.

Barnabe-Heider, F. and Frisen, J. 2008. Stem cells for spinal cord repair. *Cell Stem Cell* **3**(1): 16-24.

Becker, W.M., Reece, J.B., and Poenie, M.F. 1996. *The World of the Cell*. The Benjamin/Cummings Publishing Company, Inc., Menlo Park, California.

Bender, M.A., Bulger, M., Close, J., and Groudine, M. 2000. Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region. *Mol Cell* **5**(2): 387-393.

Berg, J.M., Tymoczko, J.L., and Stryer, L. 2002. *Biochemistry*. W.H. Freeman, New York.

Blank, V., Kim, M.J., and Andrews, N.C. 1997. Human MafG is a functional partner for p45 NF-E2 in activating globin gene expression. *Blood* **89**(11): 3925-3935.

Boldrin, L. and Morgan, J.E. 2007. Activating muscle stem cells: therapeutic potential in muscle diseases. *Curr Opin Neurol* **20**(5): 577-582.

Boyer, S.H., Wuu, K.D., Noyes, A.N., Young, R., Scher, W., Friend, C., Preisler, H.D., and Bank, A. 1972. Hemoglobin biosynthesis in murine virus-induced leukemic cells in vitro: structure and amounts of globin chains produced. *Blood* **40**(6): 823-835.

Brand, M., Rampalli, S., Chaturvedi, C.P., and Dilworth, F.J. 2008. Analysis of epigenetic modifications of chromatin at specific gene loci by native chromatin immunoprecipitation of nucleosomes isolated using hydroxyapatite chromatography. *Nat Protoc* **3**(3): 398-409.

Brand, M., Ranish, J.A., Kummer, N.T., Hamilton, J., Igarashi, K., Francastel, C., Chi, T.H., Crabtree, G.R., Aebersold, R., and Groudine, M. 2004. Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nat Struct Mol Biol* **11**(1): 73-80.

Branden, C. and Tooze, J. 1999. *Introduction to Protein Structure*. Garland Publishing Inc., New York.

Brown, K.E., Guest, S.S., Smale, S.T., Hahm, K., Merkenschlager, M., and Fisher, A.G. 1997. Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* **91**(6): 845-854.

Cammas, F., Oulad-Abdelghani, M., Vonesch, J.L., Huss-Garcia, Y., Chambon, P., and Losson, R. 2002. Cell differentiation induces TIF1beta association with centromeric heterochromatin via an HP1 interaction. *J Cell Sci* **115**(Pt 17): 3439-3448.

Cantor, A.B. and Orkin, S.H. 2001. Hematopoietic development: a balancing act. *Curr Opin Genet Dev* **11**(5): 513-519.

Chakravarthy, B.R., Tremblay, R., Macdonald, P., Krsmanovic, V., Whitfield, J.F., and Durkin, J.P. 1992. The activation of inactive membrane-associated protein kinase C is associated with DMSO-induced erythroleukemia cell differentiation. *Biochim Biophys Acta* **1136**(1): 83-90.

Cobb, B.S., Morales-Alcelay, S., Kleiger, G., Brown, K.E., Fisher, A.G., and Smale, S.T. 2000. Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding. *Genes Dev* **14**(17): 2146-2160.

Collins, F.S. and Weissman, S.M. 1984. The molecular genetics of human hemoglobin. *Prog Nucleic Acid Res Mol Biol* **31**: 315-462.

de Laat, W. and Grosveld, F. 2003. Spatial organization of gene expression: the active chromatin hub. *Chromosome Res* **11**(5): 447-459.

Deierborg, T., Soulet, D., Roybon, L., Hall, V., and Brundin, P. 2008. Emerging restorative treatments for Parkinson's disease. *Prog Neurobiol* **85**(4): 407-432.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **11**(5): 1475-1489.

Durkin, J.P., Chakravarthy, B., Tremblay, R., Jouishomme, H., Whitfield, J.F., Biquard, J.M., and Krsmanovic, V. 1992. Evidence that a novel human differentiation-inhibiting protein blocks the dimethyl sulfoxide-induced differentiation of erythroleukemia cells by inhibiting the activation of membrane protein kinase C. *Cancer Res* **52**(22): 6329-6334.

Eissenberg, J.C., Morris, G.D., Reuter, G., and Hartnett, T. 1992. The heterochromatin-associated protein HP-1 is an essential protein in Drosophila with dosage-dependent effects on position-effect variegation. *Genetics* **131**(2): 345-352.

Ellis, J., Tan-Un, K.C., Harper, A., Michalovich, D., Yannoutsos, N., Philipsen, S., and Grosveld, F. 1996. A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region. *EMBO J* **15**(3): 562-568.

Eng, J. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **5**(11): 976-989.

Epner, E., Reik, A., Cimbora, D., Telling, A., Bender, M.A., Fiering, S., Enver, T., Martin, D.I., Kennedy, M., Keller, G., and Groudine, M. 1998. The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol Cell* **2**(4): 447-455.

Forrester, W.C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G., and Groudine, M. 1987. Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res* **15**(24): 10159-10177.

Forrester, W.C., Thompson, C., Elder, J.T., and Groudine, M. 1986. A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proc Natl Acad Sci U S A* **83**(5): 1359-1363.

Francastel, C., Magis, W., and Groudine, M. 2001. Nuclear relocation of a transactivator subunit precedes target gene activation. *Proc Natl Acad Sci U S A* **98**(21): 12120-12125.

Francastel, C., Poindessous-Jazat, V., Augery-Bourget, Y., and Robert-Lezenes, J. 1997. NF-E2p18/mafK is required in DMSO-induced differentiation of Friend erythroleukemia cells by enhancing NF-E2 activity. *Leukemia* **11**(2): 273-280.

Fraser, P. and Grosveld, F. 1998. Locus control regions, chromatin activation and transcription. *Curr Opin Cell Biol* **10**(3): 361-365.

Fraser, P., Hurst, J., Collis, P., and Grosveld, F. 1990. DNaseI hypersensitive sites 1, 2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucleic Acids Res* **18**(12): 3503-3508.

Frenette, P.S. and Atweh, G.F. 2007. Sickle cell disease: old discoveries, new concepts, and future promise. *J Clin Invest* **117**(4): 850-858.

Friend, C. 1957. Cell-free transmission in adult Swiss mice of a disease having the character of a leukemia. *J Exp Med* **105**(4): 307-318.

Friend, C., Scher, W., Holland, J.G., and Sato, T. 1971. Hemoglobin synthesis in murine virus-induced leukemic cells in vitro: stimulation of erythroid differentiation by dimethyl sulfoxide. *Proc Natl Acad Sci U S A* **68**(2): 378-382.

Fujiwara, K.T., Kataoka, K., and Nishizawa, M. 1993. Two new members of the maf oncogene family, mafK and mafF, encode nuclear b-Zip proteins lacking putative trans-activator domain. *Oncogene* **8**(9): 2371-2380.

Girton, J.R. and Johansen, K.M. 2008. Chromatin structure and the regulation of gene expression: the lessons of PEV in Drosophila. *Adv Genet* **61**: 1-43.

Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kollias, G. 1987. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51**(6): 975-985.

Groudine, M., Kohwi-Shigematsu, T., Gelinas, R., Stamatoyannopoulos, G., and Papayannopoulou, T. 1983. Human fetal to adult hemoglobin switching: changes in chromatin structure of the beta-globin gene locus. *Proc Natl Acad Sci U S A* **80**(24): 7551-7555.

Guenatri, M., Bailly, D., Maison, C., and Almouzni, G. 2004. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J Cell Biol* **166**(4): 493-505.

Gurel, Z., Ronni, T., Ho, S., Kuchar, J., Payne, K.J., Turk, C.W., and Dovat, S. 2008. Recruitment of ikaros to pericentromeric heterochromatin is regulated by phosphorylation. *J Biol Chem* **283**(13): 8291-8300.

Hahm, K., Cobb, B.S., McCarty, A.S., Brown, K.E., Klug, C.A., Lee, R., Akashi, K., Weissman, I.L., Fisher, A.G., and Smale, S.T. 1998. Helios, a T cell-restricted Ikaros family member that quantitatively associates with Ikaros at centromeric heterochromatin. *Genes Dev* **12**(6): 782-796.

Hardison, R., Slightom, J.L., Gumucio, D.L., Goodman, M., Stojanovic, N., and Miller, W. 1997. Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* **205**(1-2): 73-94.

Heidari, M., Rice, K.L., Phillips, J.K., Kees, U.R., and Greene, W.K. 2006. The nuclear oncoprotein TLX1/HOX11 associates with pericentromeric satellite 2 DNA in leukemic T-cells. *Leukemia* **20**(2): 304-312.

Hivin, P., Frederic, M., Arpin-Andre, C., Basbous, J., Gay, B., Thebault, S., and Mesnard, J.M. 2005. Nuclear localization of HTLV-I bZIP factor (HBZ) is mediated by three distinct motifs. *J Cell Sci* **118**(Pt 7): 1355-1362.

Hoffman, Furie B, Benz EJ, McGlave P, Silberstein LE, and SJ, S. 2008. *Hematology - Basic Principles and Practice.* Churchill Livingstone, New York.

Hu, M., Krause, D., Greaves, M., Sharkis, S., Dexter, M., Heyworth, C., and Enver, T. 1997. Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev* **11**(6): 774-785.

Hung, H.L., Kim, A.Y., Hong, W., Rakowski, C., and Blobel, G.A. 2001. Stimulation of NF-E2 DNA binding by CREB-binding protein (CBP)-mediated acetylation. *J Biol Chem* **276**(14): 10715-10721.

Igarashi, K., Itoh, K., Hayashi, N., Nishizawa, M., and Yamamoto, M. 1995a. Conditional expression of the ubiquitous transcription factor MafK induces erythroleukemia cell differentiation. *Proc Natl Acad Sci U S A* **92**(16): 7445-7449.

Igarashi, K., Itoh, K., Motohashi, H., Hayashi, N., Matuzaki, Y., Nakauchi, H., Nishizawa, M., and Yamamoto, M. 1995b. Activity and expression of murine small Maf family protein MafK. *J Biol Chem* **270**(13): 7615-7624.

Igarashi, K., Kataoka, K., Itoh, K., Hayashi, N., Nishizawa, M., and Yamamoto, M. 1994. Regulation of transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins. *Nature* **367**(6463): 568-572.

Ingram, V.M. 1956. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature* **178**(4537): 792-794.

Johnson, G.R. and Metcalf, D. 1977. Pure and mixed erythroid colony formation in vitro stimulated by spleen conditioned medium with no detectable erythropoietin. *Proc Natl Acad Sci U S A* **74**(9): 3879-3882.

Kataoka, K., Igarashi, K., Itoh, K., Fujiwara, K.T., Noda, M., Yamamoto, M., and Nishizawa, M. 1995. Small Maf proteins heterodimerize with Fos and may act as competitive repressors of the NF-E2 transcription factor. *Mol Cell Biol* **15**(4): 2180-2190.

Kataoka, K., Nishizawa, M., and Kawai, S. 1993. Structure-function analysis of the maf oncogene product, a member of the b-Zip protein family. *J Virol* **67**(4): 2133-2141.

Kataoka, K., Noda, M., and Nishizawa, M. 1994. Maf nuclear oncoprotein recognizes sequences related to an AP-1 site and forms heterodimers with both Fos and Jun. *Mol Cell Biol* **14**(1): 700-712.

Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74**(20): 5383-5392.

Kioussis, D., Vanin, E., deLange, T., Flavell, R.A., and Grosveld, F.G. 1983. Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature* **306**(5944): 662-666.

Klug, C.A., Morrison, S.J., Masek, M., Hahm, K., Smale, S.T., and Weissman, I.L. 1998. Hematopoietic stem cells and lymphoid progenitors express different Ikaros isoforms, and Ikaros is localized to heterochromatin in immature lymphocytes. *Proc Natl Acad Sci U S A* **95**(2): 657-662.

Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Le Beau, M.M., Fisher, A.G., and Singh, H. 2002. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296**(5565): 158-162.

Kotkow, K.J. and Orkin, S.H. 1995. Dependence of globin gene expression in mouse erythroleukemia cells on the NF-E2 heterodimer. *Mol Cell Biol* **15**(8): 4640-4647.

Krantz, S.B., Sawada, K., Sawyer, S.T., and Civin, C.I. 1987. Specific binding of erythropoietin to human erythroid colony-forming cells. *Trans Assoc Am Physicians* **100**: 164-172.

Kusunoki, H., Motohashi, H., Katsuoka, F., Morohashi, A., Yamamoto, M., and Tanaka, T. 2002. Solution structure of the DNA-binding domain of MafG. *Nat Struct Biol* **9**(4): 252-256.

Levings, P.P. and Bungert, J. 2002. The human beta-globin locus control region. *Eur J Biochem* **269**(6): 1589-1599.

Lu, S.J., Rowan, S., Bani, M.R., and Ben-David, Y. 1994. Retroviral integration within the Fli-2 locus results in inactivation of the erythroid transcription factor NF-E2 in Friend erythroleukemias: evidence that NF-E2 is essential for globin expression. *Proc Natl Acad Sci U S A* **91**(18): 8398-8402.

Martelli, A.M., Riccio, M., Bareggi, R., Manfioletti, G., Tabellini, G., Baldini, G., Narducci, P., and Giancotti, V. 1998. Intranuclear distribution of HMGI/Y proteins. An immunocytochemical study. *J Histochem Cytochem* **46**(7): 863-864.

Mikkola, H.K. and Orkin, S.H. 2006. The journey of developing hematopoietic stem cells. *Development* **133**(19): 3733-3744.

Mizushima, S. and Nagata, S. 1990. pEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Res* **18**(17): 5322.

Moreau-Gachelin, F., Wendling, F., Molina, T., Denis, N., Titeux, M., Grimber, G., Briand, P., Vainchenker, W., and Tavitian, A. 1996. Spi-1/PU.1 transgenic mice develop multistep erythroleukemias. *Mol Cell Biol* **16**(5): 2453-2463.

Morley, B.J., Abbott, C.A., Sharpe, J.A., Lida, J., Chan-Thomas, P.S., and Wood, W.G. 1992. A single beta-globin locus control region element (5' hypersensitive site 2) is sufficient for developmental regulation of human globin genes in transgenic mice. *Mol Cell Biol* **12**(5): 2057-2066.

Morrison, S.J., Uchida, N., and Weissman, I.L. 1995. The biology of hematopoietic stem cells. *Annu Rev Cell Dev Biol* **11**: 35-71.

Motohashi, H., Katsuoka, F., Miyoshi, C., Uchimura, Y., Saitoh, H., Francastel, C., Engel, J.D., and Yamamoto, M. 2006. MafG sumoylation is required for active transcriptional repression. *Mol Cell Biol* **26**(12): 4652-4663.

Motohashi, H., Katsuoka, F., Shavit, J.A., Engel, J.D., and Yamamoto, M. 2000. Positive or negative MARE-dependent transcriptional regulation is determined by the abundance of small Maf proteins. *Cell* **103**(6): 865-875.

Motohashi, H., Shavit, J.A., Igarashi, K., Yamamoto, M., and Engel, J.D. 1997. The world according to Maf. *Nucleic Acids Res* **25**(15): 2953-2959.

Nagai, T., Igarashi, K., Akasaka, J., Furuyama, K., Fujita, H., Hayashi, N., Yamamoto, M., and Sassa, S. 1998. Regulation of NF-E2 activity in erythroleukemia cell differentiation. *J Biol Chem* **273**(9): 5358-5365.

Nemeth, A. and Langst, G. 2004. Chromatin higher order structure: opening up chromatin for transcription. *Brief Funct Genomic Proteomic* **2**(4): 334-343.

Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. 2003. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**(17): 4646-4658.

Ney, P.A., Sorrentino, B.P., Lowrey, C.H., and Nienhuis, A.W. 1990a. Inducibility of the HS II enhancer depends on binding of an erythroid specific nuclear protein. *Nucleic Acids Res* **18**(20): 6011-6017.

Ney, P.A., Sorrentino, B.P., McDonagh, K.T., and Nienhuis, A.W. 1990b. Tandem AP-1-binding sites within the human beta-globin dominant control region function as an inducible enhancer in erythroid cells. *Genes Dev* **4**(6): 993-1006.

Nielsen, A.L., Oulad-Abdelghani, M., Ortiz, J.A., Remboutsika, E., Chambon, P., and Losson, R. 2001. Heterochromatin formation in mammalian cells: interaction between histones and HP1 proteins. *Mol Cell* **7**(4): 729-739.

Nishizawa, M., Kataoka, K., Goto, N., Fujiwara, K.T., and Kawai, S. 1989. v-maf, a viral oncogene that encodes a "leucine zipper" motif. *Proc Natl Acad Sci U S A* **86**(20): 7711-7715.

Orkin, S.H. 1995. Transcription factors and hematopoietic development. *J Biol Chem* **270**(10): 4955-4958.

Palstra, R.J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. 2003. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* **35**(2): 190-194.

Patrinos, G.P., de Krom, M., de Boer, E., Langeveld, A., Imam, A.M., Strouboulis, J., de Laat, W., and Grosveld, F.G. 2004. Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev* **18**(12): 1495-1509.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**(18): 3551-3567.

Persons, D.A. and Tisdale, J.F. 2004. Gene therapy for the hemoglobin disorders. *Semin Hematol* **41**(4): 279-286.

Peterson, K.R., Clegg, C.H., Navas, P.A., Norton, E.J., Kimbrough, T.G., and Stamatoyannopoulos, G. 1996. Effect of deletion of 5'HS3 or 5'HS2 of the human beta-globin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. *Proc Natl Acad Sci U S A* **93**(13): 6605-6609.

Philipsen, S., Talbot, D., Fraser, P., and Grosveld, F. 1990. The beta-globin dominant control region: hypersensitive site 2. *EMBO J* **9**(7): 2159-2167.

Piwien Pilipuk, G., Galigniana, M.D., and Schwartz, J. 2003. Subnuclear localization of C/EBP beta is regulated by growth hormone and dependent on MAPK. *J Biol Chem* **278**(37): 35668-35677.

Popescu, M., Gurel, Z., Ronni, T., Song, C., Hung, K.Y., Payne, K.J., and Dovat, S. 2009. Ikaros stability and pericentromeric localization are regulated by protein phosphatase 1. *J Biol Chem*.

Purdie, A., Wells, R.M., and Brittain, T. 1983. Molecular aspects of embryonic mouse haemoglobin ontogeny. *Biochem J* **215**(2): 377-383.

Ragoczy, T., Telling, A., Sawado, T., Groudine, M., and Kosak, S.T. 2003. A genetic analysis of chromosome territory looping: diverse roles for distal regulatory elements. *Chromosome Res* **11**(5): 513-525.

Ramirez, F., Mears, J.G., and Bank, A. 1980. The molecular basis of disorders of human hemoglobin synthesis. *Mol Cell Biochem* **31**(3): 133-145.

Reem, G.H. and Friend, C. 1975. Purine metabolism in murine virus-induced erythroleukemic cells during differentiation in vitro. *Proc Natl Acad Sci U S A* **72**(4): 1630-1634.

Reik, A., Telling, A., Zitnik, G., Cimbora, D., Epner, E., and Groudine, M. 1998. The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol Cell Biol* **18**(10): 5992-6000.

Roberts, D.J. and Williams, T.N. 2003. Haemoglobinopathies and resistance to malaria. *Redox Rep* **8**(5): 304-310.

Rosenbauer, F., Wagner, K., Kutok, J.L., Iwasaki, H., Le Beau, M.M., Okuno, Y., Akashi, K., Fiering, S., and Tenen, D.G. 2004. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat Genet* **36**(6): 624-630.

Ross, J., Ikawa, Y., and Leder, P. 1972. Globin messenger-RNA induction during erythroid differentiation of cultured leukemia cells. *Proc Natl Acad Sci U S A* **69**(12): 3620-3623.

Sawada, K., Krantz, S.B., Dai, C.H., Koury, S.T., Horn, S.T., Glick, A.D., and Civin, C.I. 1990. Purification of human blood burst-forming units-erythroid and demonstration of the evolution of erythropoietin receptors. *J Cell Physiol* **142**(2): 219-230.

Sawada, K., Krantz, S.B., Dai, C.H., Sato, N., Ieko, M., Sakurama, S., Yasukouchi, T., and Nakagawa, S. 1991. Transitional change of colony stimulating factor requirements for erythroid progenitors. *J Cell Physiol* **149**(1): 1-8.

Sawada, K., Krantz, S.B., Kans, J.S., Dessypris, E.N., Sawyer, S., Glick, A.D., and Civin, C.I. 1987. Purification of human erythroid colony-forming units and demonstration of specific binding of erythropoietin. *J Clin Invest* **80**(2): 357-366.

Shestakova, E.A., Mansuroglu, Z., Mokrani, H., Ghinea, N., and Bonnefoy, E. 2004. Transcription factor YY1 associates with pericentromeric gamma-satellite DNA in cycling but not in quiescent (G0) cells. *Nucleic Acids Res* **32**(14): 4390-4399.

Sun, J., Brand, M., Zenke, Y., Tashiro, S., Groudine, M., and Igarashi, K. 2004. Heme regulates the dynamic exchange of Bach1 and NF-E2-related factors in the Maf transcription factor network. *Proc Natl Acad Sci U S A* **101**(6): 1461-1466.

Tahara, T., Sun, J., Nakanishi, K., Yamamoto, M., Mori, H., Saito, T., Fujita, H., Igarashi, K., and Taketani, S. 2004. Heme positively regulates the expression of beta-globin at the locus control region via the transcriptional factor Bach1 in erythroid cells. *J Biol Chem* **279**(7): 5480-5487.

Talbot, D. and Grosveld, F. 1991. The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. *EMBO J* **10**(6): 1391-1398.

Tenen, D.G. 2003. Disruption of differentiation in human cancer: AML shows the way. *Nat Rev Cancer* **3**(2): 89-101.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**(6): 1453-1465.

Tuan, D., Solomon, W., Li, Q., and London, I.M. 1985. The "beta-like-globin" gene domain in human erythroid cells. *Proc Natl Acad Sci U S A* **82**(19): 6384-6388.

Voltarelli, J.C., Couri, C.E., Stracieri, A.B., Oliveira, M.C., Moraes, D.A., Pieroni, F., Barros, G.M., Madeira, M.I., Malmegrim, K.C., Foss-Freitas, M.C., Simoes, B.P., Foss, M.C., Squiers, E., and Burt, R.K. 2008. Autologous hematopoietic stem cell transplantation for type 1 diabetes. *Ann N Y Acad Sci* **1150**: 220-229.

Wade, N. 2006. Some scientists see shift in stem cell hopes. *NY Times (Print)*: A18.

World Health Organization. 2006. Thalassaemia and other haemoglobinopathies. In.

Wu, C.J., Krishnamurti, L., Kutok, J.L., Biernacki, M., Rogers, S., Zhang, W., Antin, J.H., and Ritz, J. 2005. Evidence for ineffective erythropoiesis in severe sickle cell disease. *Blood* **106**(10): 3639-3645.

Yamamoto, T., Kyo, M., Kamiya, T., Tanaka, T., Engel, J.D., Motohashi, H., and Yamamoto, M. 2006. Predictive base substitution rules that determine the binding and transcriptional specificity of Maf recognition elements. *Genes Cells* **11**(6): 575-591.

Yoshida, T., Ohkumo, T., Ishibashi, S., and Yasuda, K. 2005. The 5'-AT-rich half-site of Maf recognition element: a functional target for bZIP transcription factor Maf. *Nucleic Acids Res* **33**(11): 3465-3478.

Zhang, P., Behre, G., Pan, J., Iwama, A., Wara-Aswapati, N., Radomska, H.S., Auron, P.E., Tenen, D.G., and Sun, Z. 1999. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. *Proc Natl Acad Sci U S A* **96**(15): 8705-8710.

APPENDIX

```
$seq =
"tatggcgaggaaaactgaaaaaggtggaaaatttagaaatgtccactgtag
gacgtggaatatggcaagaaaactgaaaatcatggaaaatgagaaacatcc
acttgacgacttgaaaaatgacgaaatcactaaaaaacgtgaaaaatagaa
atgcacactgaaggacctggaatatggcgagaaaactgaaatcacggaaa
atgagaaatacacactttaggacgtgaaa";
print "\n";
print "Group I: Symmetric";
print "\n";
#!/usr/bin/perl
while ($seq =~ /([cg][tga][ct][atg][ga][ctg]a)/gi){
  $pos = pos($seq)-6;
  print ("$1\t$pos\n");
}
print "\n";
print "Group II: Symmetric";
print "\n";
while ($seq =~ /([cg]t[ca]agc[tca])/gi){
  $pos = pos($seq)-6;
  print ("$1\t$pos\n");
}
print "\n";
print "Group II: Asymmetric";
print "\n";
while ($seq =~ /([cg][gcat][gct][atgc]gca)/gi){
  $pos = pos($seq)-6;
  print ("$1\t$pos\n");
}
print "\n";
print "Group III: Symmetric";
print "\n";
while ($seq =~ /([tacg]tcagc[ag])/gi){
  $pos = pos($seq)-6;
  print ("$1\t$pos\n");
}
print "\n";
print "Group III: Asymmetric";
print "\n";
while ($seq =~ /([cg]t[ca]a[ga][cgat][atgc])/gi){
  $pos = pos($seq)-6;
  print ("$1\t$pos\n");
}
print "\n";
print "Group IV: Symmetric";
print "\n";
while ($seq =~ /([cg]tca[gt]ca)/gi){
  $pos = pos($seq)-6;
  print ("$1\t$pos\n");
}
print "\n";
print "Group IV: Asymmetric";
print "\n";
while ($seq =~ /([cg]tca[gct]ca)/gi){
  $pos = pos($seq)-6;
  print ("$1\t$pos\n");
}
print "\n";
print "Group V: Symmetric";
print "\n";
while ($seq =~ /([cg][ct][gc][ac][gc][ca]a)/gi){
  $pos = pos($seq)-6;
  print ("$1\t$pos\n");
}
```

Figure S1. Findmatches.pl script. The program written in Perl programming language searches the sequence of interest (entered between the quotation marks after $seq =) for the classic Maf recognition element (MARE) and MARE variants identified by Yamamoto et al. (2006) (summarized in Table 4).

Major Satellite DNA

Forward Strand (5′ - 3′)

```
  1 tatggcgagg aaaactgaaa aaggtggaaa atttagaaat gtccactgta
 51 ggacgtggaa tatggcaaga aaactgaaaa tcatggaaaa tgagaaacat
101 ccacttgacg acttgaaaaa tgacgaaatc actaaaaaac gtgaaaaatg
151 agaaatgcac actgaaggac ctggaatatg gcgagaaaac tgaaaatcac
201 ggaaaatgag aaatacacac tttaggacgt gaaa
```

Reverse Strand (5′ - 3′)

```
  1 tttcacgtcc taaagtgtgt atttctcatt ttccgtgatt ttcagttttc
 51 tcgccatatt ccaggtcctt cagtgtgcat ttctcatttt tcacgttttt
101 tagtgatttc gtcatttttc aagtcgtcaa gtggatgttt ctcattttcc
151 atgattttca gttttcttgc catattccac gtcctacagt ggacatttct
201 aaattttcca ccttttttcag ttttcctcgc cata
```

Minor Satellite DNA

Forward Strand (5′ - 3′)

```
  1 ggaaaatgat aaaaaccaca ctgtagaacc ttttagagga gtgagttaca
 51 ctgaaaaaca cattcgttgg aaacgggatt tgtagaacag tgtatatcaa
101 tgagttacaa tgagaaacat
```

Reverse Strand (5′ - 3′)

```
  1 atgtttctca ttgtaactca ttgatataca ctgttctaca aatcccgttt
 51 ccaacgaatg tgtttttcag tgtaactcac tcctctaaaa ggttctacag
101 tgtggttttt atcattttcc
```

Figure S2. Murine Satellite DNA sequences. Major and minor satellite DNA repeats constitute the majority of (peri)-centromeric heterochromatin. Shown here are the consensus sequences for each, both forward and reverse strands in the 5′ to 3′ direction, which were searched using Findmatches.pl for MARE and MARE variants.