



VCU

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2014

Structural and Functional Characterization of the MBD2-NuRD Co-Repressor Complex

Megha Desai

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>



Part of the [Genetics Commons](#), [Genomics Commons](#), [Molecular Biology Commons](#), and the [Molecular Genetics Commons](#)

© The Author

Downloaded from

<https://scholarscompass.vcu.edu/etd/3617>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

© Megha A. Desai, 2014

All Rights Reserved

**Structural and Functional Characterization of the MBD2-NuRD
Co-Repressor Complex**

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy at Virginia Commonwealth University.

by
Megha A. Desai
Bachelor of Engineering, NMAM Institute of Technology, 2009

Director: Gordon D. Ginder, M.D.
Director of Massey Cancer Center, Professor of Internal Medicine, Human and
Molecular Genetics, Microbiology and Immunology

Virginia Commonwealth University
Richmond, Virginia
December 2nd, 2014

Acknowledgements

I would first like to thank my advisor, Dr. Ginder, for giving me the wonderful opportunity to pursue my PhD thesis under his guidance. I am very grateful to Dr. Ginder for sharing his knowledge, wisdom and enthusiasm for science. His optimism has always motivated me to push through the many upsets of research. Dr. Ginder's humility is an example to follow and I couldn't have asked for a better mentor.

I must thank Shou Zhen and Sheng Zhu for their constant help and guidance in our lab. They hold our lab together. My research over the 5 years would not have been possible without both of them.

Justin has been a great friend and colleague in Ginder lab. Thank you, Justin, for especially being there when I needed someone to bounce off ideas of my research.

In our time together in the lab, Maria and I became very dear friends and I truly admire her as a scientist and even more for the wonderful person she is. Maria has always shared her knowledge and experiences generously and extended her support whenever I needed it.

Merlin, a past graduate student in the lab, was a wonderful mentor to me when I first joined the lab as a rotation student. I thank her for the projects she initiated during her graduate training, which later became the focus of my thesis research.

It was also a pleasure to have known Omar and Q, who were both past MD/PhD students in the lab. There was a lot learnt from the knowledge and experiences they shared.

I would like to thank all my committee members for their guidance and support throughout the course of my graduate research. Their help and ideas were invaluable in shaping the course of my research projects.

A special thanks to Dr. David Williams for being an amazing collaborator; the work presented in Chapter 2 would not have been possible without his expertise. His passion for science is highly contagious.

Julie Farnsworth, from the VCU flow cytometry core, has been the heart and soul of my FACS studies and my experiments would not have been possible without her dedication and enthusiasm to see my research succeed.

I thank Amy and Freida for all the help with scheduling meetings, reserving conference rooms and for all the thoughtful things they do that make working in Massey Cancer Center such a wonderful experience.

I would like to extend a big thank you to the team of Massey operations: Neil, Mike(s) and George. They are an amazing team and make the work experience here appear seamless. Thank you for being our 24x7 helpline.

I must thank Rita Shiang for being such a great program director for Dept. of Human and Molecular Genetics. She does an amazing job on keeping us on track with department requirements and for recognizing student achievements in various ways. We are very fortunate to have her as a strong advocate of the students.

I cannot thank enough my parents and grandmother for their selfless love and support through the years. I wouldn't be the person I am today without their encouragement and many sacrifices. I thank them for teaching me to believe in myself and to pursue my dreams.

I consider myself very lucky to have the most amazing parents-in-law. I thank Mom and Dad for the love and encouragement and for being there when I needed them the most. I look up to my Dad as my guide and mentor.

I run out of words to thank my fiancé, Hernal for everything he does for me. He is my pillar of support. His unconditional love and support have been and continue to be my biggest strength. I am grateful to have him in my life. I love him very much.

And last but not the least, I thank my wonderful friends, Apoorva, Gopichand, Divya, Anting, Maria who made the roller coaster of graduate life so memorable.

Dedication

I dedicate my thesis to my parents, who dedicated their life to the betterment of their patients and inspired many students, including me, to contribute to the advancement of medicine.

Table of Contents

Acknowledgements	ii
Dedication	iv
List of Figures.....	viii
List of Tables.....	x
List of Abbreviations.....	xi
Abstract	xv
CHAPTER 1: Introduction	1
I. Hemoglobin and the hemoglobin switch:	1
A. Hemoglobin.....	1
B. The β -globin gene locus and the hemoglobin switch:.....	2
II. Erythropoiesis:.....	5
III. β -hemoglobinopathies:	7
A. Sickle Cell Disease (SCD)	8
B. β -thalassemias	11
C. Management of β -hemoglobinopathies:	15
i. Therapeutic options for β -hemoglobinopathies:	15
ii. Fetal Hemoglobin Stimulants:.....	16
iii. Futuristic therapeutic options:.....	19
IV. Human β -globin gene regulation	21

A.	Cis-acting sequences.....	21
B.	Trans-acting factors:	23
V.	Epigenetics:.....	33
A.	DNA Methylation:	35
i.	Methyl Cytosine Binding Proteins:.....	35
B.	Histone Modifications.....	39
C.	MicroRNAs	40
D.	NuRD complex:.....	42
i.	MBD2-NuRD Complex in normal development.....	45
ii.	MBD2-NuRD Complex in Tumorigenesis:	46
iii.	MBD2-NuRD complex in hemoglobin switching:.....	47
VI.	Primary adult human erythroid cells as a model to study hemoglobin switching :	49
VII.	Scope of the thesis.....	50
VIII.	Significance.....	51
CHAPTER 2: An Intrinsically Disordered Region of MBD2 recruits the Histone Deacetylase Core of the NuRD Complex		53
I.	Introduction:.....	53
II.	Methods:.....	56
III.	Results:.....	63
A.	The MBD2 _{IDR} region is intrinsically disordered.....	63
B.	A polypeptide linkage stabilizes the MBD2-p66 α coiled-coil complex.....	63
C.	NMR analysis of intrinsically disordered MBD2 _{IDR}	65
D.	The MBD2 _{IDR} modifies DNA binding kinetics and overall binding affinity:	67
E.	The MBD2 _{IDR} binds the histone deacetylase core complex of NuRD:.....	69
F.	Key residues within minimal MBD2 _{IDR} are Evolutionarily Conserved:.....	71
G.	Functional biological assay of double mutant MBD2:	77

IV. Discussion:	81
CHAPTER 3: Role of the MBD2-NuRD complex in developmental regulation of the fetal γ -globin gene .	87
I. Introduction:.....	87
II. Methods:.....	97
III. Results:.....	106
A. Over-expression of ZBTB32 in CD34+ adult human erythroid cells activates the fetal γ -globin gene:	106
B. MicroRNA expression profile of MBD2 knockout mice:	109
C. Validation of miR-210 expression in adult erythrocytes from β -YAC transgenic mice:	109
D. Functional analyses of miR-210 expression in adult mouse bone marrow cells:	111
E. Validation of miR-210 expression in MBD2 knockdown CD34 ⁺ hematopoietic progenitor cells:.....	114
F. Functional analysis of miR-210 in CD34 ⁺ hematopoietic progenitor cells:.....	117
IV. Discussion:	120
CHAPTER 4: Summary, Future Directions and Perspectives	126
Literature Cited.....	140
Vita	154

List of Figures

Figure 1. The hemoglobin switch.....	4
Figure 2. Human erythropoiesis.....	6
Figure 3. Pathophysiology of sickle cell disease.	10
Figure 4. Pathophysiology of β -thalassemia.....	13
Figure 5. Trans-acting factors regulating hemoglobin switch..	34
Figure 6. The role of DNA methylation in gene silencing.....	38
Figure 7. MicroRNA Biogenesis.....	41
Figure 8. The MBD2/NuRD co-repressor complex..	44
Figure 9. The MBD2 _{IDR} of MBD2 is intrinsically disordered in isolation..	64
Figure 10. A polypeptide linkage stabilizes the MBD2-p66 α coiled-coil complex.....	66
Figure 11 NMR analysis of intrinsically disordered MBD2IDR.....	68
Figure 12 The MBD2IDR modifies DNA binding kinetics and affinity..	70
Figure 13. The MBD2IDR binds the histone deacetylase core complex of NuRD..	72
Figure 14. MBD2IDR does not interact with p66 α and Mi2 β subunits of NuRD.	74
Figure 15. Evolutionary conservation of key residues within minimal MBD2IDR.	75
Figure 16. Two conserved residues from minimal MBD2IDR are necessary to bind the histone deacetylase core of NuRD.	78
Figure 17. Double mutant MBD2 shows reduced transcriptional repression of its methylated target gene PRSS8.....	80
Figure 18. A working model of the architecture of the MBD2-NuRD co-repressor complex..	86
Figure 19. Gene expression microarray in erythroid cells of MBD2 ^{-/-} mice	89
Figure 20. Validation of candidate genes by Q-RT PCR.	92
Figure 21. Chromatin Immunoprecipitation of MBD2 in ZBTB32 gene..	94
Figure 22. Validation of candidate genes in CID β -YAC cells.	96

Figure 23 Two step miRNA qPCR.	103
Figure 24. Validation and functional analysis of ZBTB32 in CD34+ adult erythroid cells.....	107
Figure 25. ZBTB32 over-expressing CD34+ cells show normal erythroid differentiation profile..	108
Figure 26. Validation of miR-210 expression in mouse splenic erythroblasts.	110
Figure 27. mir-210 regulates human β -type globin gene expression in β -YAC mouse bone marrow cells.	113
Figure 28. miR-210 expression in MBD2 knockdown CD34+ erythroid cells.....	115
Figure 29 Functional analysis of miR-210 over-expression in CD34+ cells.....	118
Figure 30. Over-expression of miR-210 in erythroid cells does not affect its normal differentiation profile..	119

List of Tables

Table 1. List of primers used for mutagenesis	61
Table 2. List of primers used for qPCR in Chapter 2	62
Table 3. List of primers used for qPCR in Chapter 3	105
Table 4. miR-210 expression in a background of varying MBD2 knockdown and γ -globin gene expression in erythroid cells.	116

List of Abbreviations

5-aza.....	5-azacytidine
5-HMF.....	5-hydroxymethyl-2-furfural
α	Alpha
β	Beta
β maj.....	Beta Major
β min.....	Beta Minor
β YAC.....	β -globin Locus Yeast Artificial Chromosome
γ	Gamma
δ	Delta
ϵ	Epsilon
μ	Micro
ADP.....	Adenosine Diphosphate
APC.....	Adenomatous polyposis coli
bp.....	Base Pair
BCL11A.....	B-cell lymphoma/leukemia 11A
BFU-E.....	Burst Forming Unit-Erythroid
C.....	Celsius

CD.....	Cluster of Differentiation
CFU-E.....	Colony Forming Unit-Erythroid
CHD4.....	Chromodomain-helicase-DNA-binding protein 4
ChIP.....	Chromatin Immunoprecipitation
CID.....	Chemical Inducer of Dimerization
CpG.....	Cytosine-Guanine Dinucleotide
CRISPR.....	Clustered Regulatory Interspersed Short Palindromic Repeat
DEPC.....	Diethylpyrocarbonate
DM.....	Differentiation Medium
DNA.....	Deoxyribonucleic Acid
DNMT.....	DNA Methyltransferase
DR.....	Direct Repeat
EPO.....	Erythropoietin
FACS.....	Fluorescence-Activated Cell Sorting
FBS.....	Fetal Bovine Serum
FOG-1.....	Friend of GATA-1
GM.....	Growth Medium
GFP.....	Green Fluorescent Protein
GWAS.....	Genome-Wide Association Study
H3.....	Histone H3
HbA.....	Hemoglobin A
HbA2.....	Hemoglobin A2
HbF.....	Fetal Hemoglobin

HbS.....	Sickle Cell Hemoglobin
HDAC.....	Histone Deacetylase
HLA.....	Human Leukocyte Antigen
HMGA2.....	High-Mobility Group AT-hook 2
HPFH.....	Hereditary Persistence of Fetal Hemoglobin
HPLC.....	High Performance Liquid Chromatography
HS.....	Hypersensitive site
IgG.....	Immunoglobulin G
IMDM.....	Iscove's Modified Dulbecco's Medium
IP.....	Immunoprecipitation
iPSC.....	Induced Pluripotent Stem Cell
KD.....	Knockdown
KO.....	Knockout
LCR.....	Locus Control Region
LSD1.....	Lysine-Specific Demethylase 1
KLF1.....	Krüppel-like Factor 1
M.....	Molar
MBD.....	Methyl Binding Domain
MBD2.....	Methyl-CpG-Binding Domain Protein 2
MBD3.....	Methyl-CpG-Binding Domain Protein 3
miR.....	microRNA
MTA.....	Metastasis-Associated
MYB.....	Myeloblastosis
NF-E.....	Nuclear Factor, Erythroid Derived 2

NuRD.....	Nucleosome Remodeling and Histone Deacetylase
PBS.....	Phosphate Buffered Saline
PCR.....	Polymerase Chain Reaction
PYR.....	SWI/SNF Related Protein
qPCR.....	Quantitative Polymerase Chain Reaction
RNA.....	Ribonucleic acid
RbAp.....	Retinoblastoma Associated Protein
SCD.....	Sickle Cell Disease
SCR.....	Scrambled Control
SDS.....	Sodium Dodecyl Sulfate
SFEM.....	Serum-Free Expansion Medium
shRNA.....	short hairpin RNA
SIN-LV.....	Self-Inactivating Lentivirus
siRNA.....	small interfering RNA
SNP.....	Single Nucleotide Polymorphism
Sox6.....	Sry-related HMG box 6
SSE.....	Stage Specific Element
SUMOylation.....	Small Ubiquitin-like Modifier
TALEN.....	Transcription-Activator like Nucleases
Th.....	T Helper Cell
TR2.....	Testicular Receptor 2
TR4.....	Testicular Receptor 4
Ugt8.....	UDP glycosyltransferase 8
UTR.....	Untranslated Region
WT.....	Wild Type
ZBTB32.....	Zinc-finger and BTB-domain containing 32
ZFN.....	Zinc-Finger Nuclease

Abstract

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF THE MBD2-NURD CO-REPRESSOR COMPLEX

By
Megha A. Desai, B.E.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2014

Thesis Advisor: Gordon D. Ginder, M.D.
Director, Massey Cancer Center
Professor, Departments of Internal Medicine, Human and Molecular Genetics and
Microbiology and Immunology

The MBD2-NuRD co-repressor complex is an epigenetic regulator of the developmental silencing of embryonic and fetal β -type globin genes in adult erythroid cells as well as aberrant methylation-dependent silencing of tumor suppressor genes in neoplastic diseases. Biochemical characterization of the MBD2-NuRD complex in chicken

erythroid cells identified RbAp46/48, HDAC1/2, MTA1/2/3, p66 α/β , Mi2 α/β and MBD2 to comprise this multi-protein complex.

In the work presented in Chapter 2, we have pursued biophysical and molecular studies to describe a previously uncharacterized domain of human MBD2 (MBD2_{IDR}). Biophysical analyses show that MBD2_{IDR} is an intrinsically disordered region (IDR). Despite this inherent disorder, MBD2_{IDR} increases the overall binding affinity of MBD2 for methylated DNA. MBD2_{IDR} also recruits the histone deacetylase core components (RbAp48, HDAC2 and MTA2) of NuRD through a critical area of contact requiring two contiguous amino acid residues, Arg²⁸⁶ and Leu²⁸⁷. Mutation of these critical residues abrogates interaction of MBD2 with the histone deacetylase core and impairs the ability of MBD2 to repress the methylated tumor suppressor gene *Prostasin* in MDA-MB-435 breast cancer cells. These findings expand our knowledge of the multi-dimensional interactions of the MBD2-NuRD complex that govern its function.

In Chapter 3, we have discussed a novel mechanism for MBD2-mediated silencing of the fetal γ -globin gene. Through microarray expression analyses in adult erythroid cells of MBD2^{-/-} mice, we identified ZBTB32 and miR-210 as downstream targets of MBD2. Over-expression of ZBTB32 and miR-210 in adult erythroid cells causes increased expression of the silenced fetal γ -globin gene. Thus, our results indicate that MBD2 may regulate γ -globin gene expression indirectly through ZBTB32 and miR-210 in adult erythroid cells.

CHAPTER 1: Introduction

I. Hemoglobin and the hemoglobin switch:

A. Hemoglobin

Hemoglobin is the oxygen carrying metalloprotein in the red blood cells in our body.

Hemoglobin carries oxygen from the lungs to other tissues in the body. Although there are many variants of hemoglobin in normal embryonic, fetal and adult life as well as different pathologic conditions, each hemoglobin molecule is a tetramer consisting of four subunits including two α -like chains and two β -like globin chains. In humans, the genes coding the α -chains are located on the short arm of chromosome 16 (16p13.3). The α -globin locus contains one functional embryonic ζ -globin gene and two adult α -globin genes (α_1 and α_2). The genes coding for the β -globin chains are located on the short arm of chromosome 11 (11p15.5) and it determines different types of hemoglobin chains that are expressed in different stages of development (David G. Nathan, David Ginsburg, Stuart H. Orkin, A. Thomas Look, 2003).

In adult humans, the most common form of hemoglobin is HbA ($\alpha_2\beta_2$) which comprises ~95% of the total hemoglobin. HbA₂ ($\alpha_2\delta_2$) comprises about 1.5-3.5% of the total

hemoglobin whereas the fetal hemoglobin HbF ($\alpha_2\gamma_2$) comprises <2% of the total hemoglobin in adults (David G. Nathan, David Ginsburg, Stuart H. Orkin, A. Thomas Look 2003).

B. The β -globin gene locus and the hemoglobin switch:

The human β -globin gene cluster lies on the short arm of chromosome 11 (11p15.5) and consists of different genes that code for the β -globin subunit of hemoglobin. This cluster includes a pseudogene ($\psi\beta$) and 5 functional genes: 5' - epsilon(ϵ) - gamma G($^G\gamma$) - gamma A($^A\gamma$) - delta(δ) - beta(β) which are expressed in the order of their arrangement within the locus (Stamatoyannopoulos et al. 2001). All of the expressed β -type globin gene products have 146 amino acids and consist of two introns and 3 exons.

The earliest hemoglobin to be expressed is the embryonic hemoglobin ($\zeta_2\epsilon_2$) which consists of the ϵ -globin subunit from the β -globin gene cluster and the ζ -globin subunit from the α -globin gene cluster. The ϵ -globin gene is expressed in the first wave of primitive erythropoiesis from the yolk sac during the first 5-7 weeks of gestation.

At 5-7 weeks post gestation, the primitive erythroid lineage is replaced by the definitive erythroid lineage and undergoes maturational globin switching at which time the ϵ -globin gene expression ceases and the fetal γ -globin gene expression begins. This first major “hemoglobin switching” event coincides with the changes in the sites of erythropoiesis from the embryonic yolk sac to the developing fetal liver. Fetal hemoglobin ($\alpha_2\gamma_2$) consists of the fetal β -type globin subunits, $^G\gamma$ and $^A\gamma$, and the α -globin subunit from the α -globin gene cluster (Figure 1). The γ -chains arose due to gene

duplications in primate evolution and differ only in the presence of glycine or alanine, respectively, at position 136. The $\epsilon\gamma$ subunit is produced three times higher than the $\delta\gamma$ subunit in these fetal erythrocytes.

The second major “hemoglobin switch” occurs perinatally in humans and Old World primates around 32 weeks post-gestation when the fetal γ -globin subunits are gradually replaced by two adult β -globin subunits forming the adult hemoglobin: $\alpha_2\beta_2$. By six months of age, adult hemoglobin HbA₁ ($\alpha_2\beta_2$) is the primary hemoglobin that is expressed. The adult bone marrow is now the primary site of erythropoiesis for the adult erythrocytes (Figure 1) (Boyer et al. 1975; Ley et al. 1989; Peschle et al. 1985).

The first “hemoglobin switch” occurring in the primitive embryonic erythroid cells to give way to the definitive erythroid cells is found among all vertebrates. However the second “hemoglobin switch” leading from fetal-to-adult definitive erythroid cells occurs only in the primates and humans and is believed to be due to the unique expression of the γ genes during the fetal period - an event which took place around 35 to 55 million years ago during primate evolution. Although the γ - and β -globin chains are highly homologous, there are significant structural differences between them which confer different functional properties to HbF and HbA. HbF has higher average affinity for oxygen compared to HbA; this property is physiologically relevant since the maternal O₂

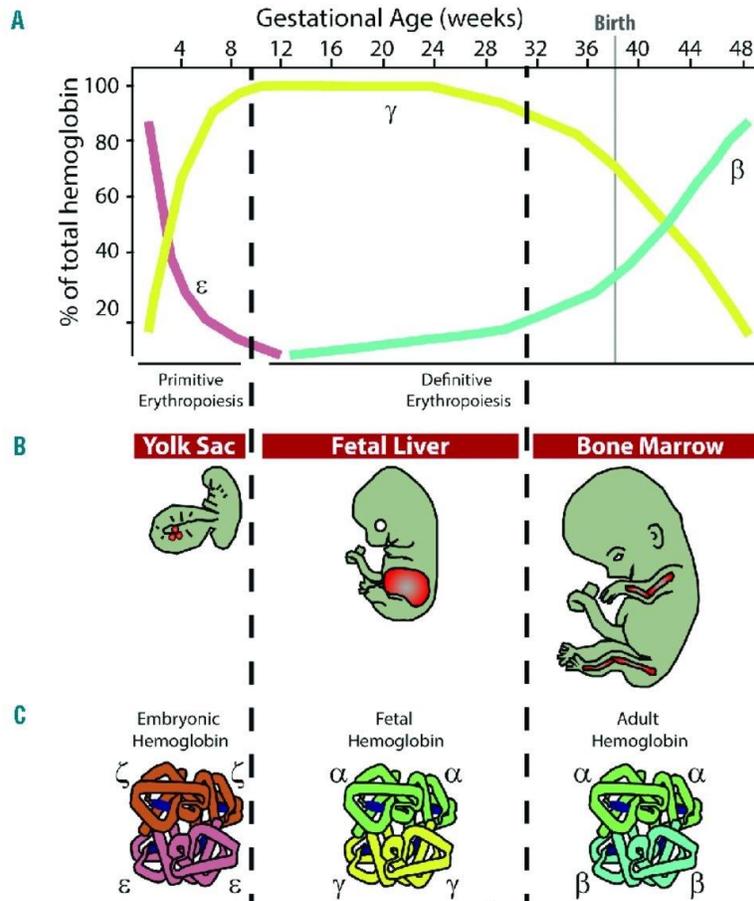


Figure 1. The hemoglobin switch. (A) Hemoglobin switching in man. Around week 6 of gestation, embryonic globin (ϵ) is silenced and fetal globin (γ) starts to be expressed. Perinatally the switch to adult globin (β) occurs. For the α -like globins, a single switch from the embryonic (ζ) to adult (α) globin occurs (not shown). (B) Major anatomical sites of hematopoiesis during development. Erythropoiesis occurs in the blood islands of the yolk sac in the first 8 weeks of gestation, then in the fetal liver between 8 and 32 weeks, and finally in the bone marrow from 32 weeks on. Around birth the spleen serves as a transient erythropoietic organ (not shown). (C) Structure of the main human hemoglobins expressed during development. Embryonic globin ($\zeta_2\epsilon_2$; HbE Gower-1); fetal hemoglobin ($\alpha_2\gamma_2$; HbF) and adult hemoglobin ($\alpha_2\beta_2$, HbA). Adapted from Cantu and Philipsen, 2014.

needs to diffuse to the fetal circulation and the higher O₂ affinity of HbF ensures that some of the maternal O₂ will be trapped by the fetal circulatory system. Also, the lower O₂ affinity of adult hemoglobin (maternal hemoglobin) readily allows transfer of its O₂ to fetal hemoglobin (Bell 1999).

II. Erythropoiesis:

Erythropoiesis is the process by which red blood cells are formed from erythroid precursor cells. There are two types of erythropoiesis in humans, primitive erythropoiesis and definitive erythropoiesis. Primitive erythropoiesis occurs from the “blood islands” in the yolk sac early on during embryonic development and gives rise to large, nucleated erythroid cells which mature and enucleate in circulation. Definitive erythropoiesis occurs during late fetal life and throughout adulthood when the sites of erythropoiesis are the fetal liver and postnatal bone marrow respectively, giving rise to mature, enucleated erythroid cells. Unlike primitive erythropoiesis, definitive erythropoiesis occurs within erythroblastic islands, composed of erythroblasts physically attached to central macrophage cells. (McGrath and Palis 2008; Palis 2014).

Hematopoietic stem cells differentiate into a common myeloid progenitor which gives rise to the lineage-committed, definitive erythroid progenitors called burst forming unit-erythroid (BFU-E) and colony forming unit-erythroid (CFU-E). These give rise to the morphologically identifiable, nucleated precursors that progress from pro erythroblasts (ProE) to basophilic erythroblasts (BasoE), polychromatophilic erythroblasts (PolyE) and orthochromatophilic erythroblasts (OrthoE) to reticulocyte. Reticulocytes are released

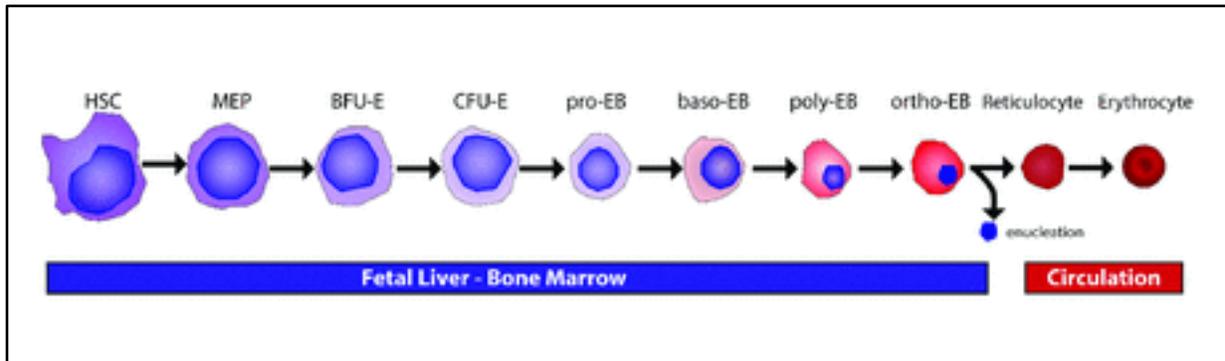


Figure 2. Human erythropoiesis. Hematopoiesis begins in the bone marrow with the multi-potent hematopoietic stem cell. This cell gives rise to a common myeloid progenitor (MEP), followed by a megakaryotic/erythroid pluripotent progenitor. In erythropoiesis, the MEP goes through a series of morphological changes culminating in the formation of a Hb-filled and organelle-lacking erythrocyte. These stages include the Burst forming unit-erythroid (BFU-E), Colony forming unit-erythroid (CFU-E), proerythroblast, basophilic erythroblast, polychromatophilic and orthochromatophilic erythroblasts which loses the nucleus and differentiates into the reticulocyte. The reticulocyte is released into bloodstream and matures in RBC. Adapted from Cantu and Philipsen, 2014

into the bloodstream and comprise about 1% of circulating red blood cells (Figure 2). The reticulocytes mature into erythrocytes after one to two days in circulation. These different stages of erythropoiesis can be visualized by staining with Wright staining and examination by light microscopy. Progressive erythroid maturation is characterized by erythroblast expansion through a limited set of symmetric cell divisions, accumulation of hemoglobin, decrease in cell size, nuclear pyknosis and decrease in RNA content. As erythrocytes age, surface area and volume, but not the hemoglobin content, are progressively lost which leads to increased erythrocyte cell density. Senescent erythrocytes are ultimately cleared by splenic macrophages (Palis 2014).

III. β -hemoglobinopathies:

Disorders of the β -globin gene (β -hemoglobinopathies) are single gene disorders which are inherited as autosomal recessive traits. According to a recent estimate by the World Health Organization, about 0.7% of the world's population is carrier of sickle cell disease, affecting over 275,000 babies and of β -thalassemia, affecting around 56,000 babies every year (Weatherall and Clegg 2001). About 60% of total and 70% of all pathological cases are found in Africa. β -hemoglobinopathies are most common in the ethnic populations from Africa, the Mediterranean basin and South-east Asia.

Hemoglobinopathies provide a selective advantage to heterozygous populations for survival against malaria, hence these variant alleles have been evolutionarily selected in geographical areas where malaria is endemic (Weatherall and Clegg 2001). The β -hemoglobinopathies can be characterized into two groups. The first group is

characterized by structurally abnormal β -globin chains most commonly caused by missense mutations. Over 900 qualitative structural variants of hemoglobin have been reported, although most of them are rare and benign. Most of these are caused due to amino acid substitutions, although some are also brought about by insertions and deletions of amino acids, elongation of globin chains and poly-peptide fusions as a result of recombination between globin chains (Examples, Hb S, E, D^{punjab}, C, Lepore etc.) (Weatherall and Clegg 2001). Sickle cell disease is the most common type of β -hemoglobinopathy in this group. The second group is characterized by the production of abnormally low amounts of the β -globin chains, through mutations as well as deletions that affect regulation of the gene. β -thalassemias comprise this second group of disorders.

A. Sickle Cell Disease (SCD)

Sickle cell disease is one of the most common severe monogenic disorders worldwide. It is a multi-system disorder that is caused by inheritance of the mutated β -globin chain and involves periods of acute illness and progressive organ damage arising from hemoglobin polymerization. Sickle cell anemia is the most common form of sickle cell disease and results from a single base-pair mutation which causes substitution of a valine in place of glutamic acid in the 6th position in the adult β -globin gene (β^S). This mutation produces a hydrophobic motif in the deoxygenated HbS tetramer such that it results in binding between the β_1 and β_2 chains of two hemoglobin molecules. The crystallization produces a polymer nucleus which fills the erythrocyte, disrupting its architecture and flexibility and promoting cellular dehydration with physical and oxidative

stress (Figure 3). A normal red cell typically functions for 90-120 days, whereas sickle cells only last 10-20 days.

Sickle cell anemia manifests as two major pathophysiological processes: vaso-occlusion with ischemia-reperfusion injury, and hemolytic anemia. Acute vaso-occlusion pain is caused by entrapment of erythrocytes and leucocytes in the microcirculation, causing vascular obstruction and tissue ischemia. Restoration of blood flow further promotes tissue damage by reperfusion. These cycles of ischemia and reperfusion lead to oxidative stress. Hemolysis causes anemia and fatigue; cutaneous leg ulcerations, priapism and pulmonary hypertension are also associated with hemolysis. The release of hemoglobin in the plasma during hemolysis also produces reactive oxygen species which inhibit endothelial nitric oxide signaling, leading to endothelial cell dysfunction and nitric oxide resistance (David G. Nathan, David Ginsburg, Stuart H. Orkin, A. Thomas Look 2003).

The prevalence of sickle cell disease is the highest in the sub-Saharan population; a recent estimate suggests about 230,000 affected children are born every year in this region. In comparison, the number of affected children in North America is 2600, with 1300 in Europe every year. Due to the adaptive advantage to heterozygotes, the occurrence of SCD is highest in people with ancestry from regions where malaria has been endemic, including Africa, the Mediterranean, Indian and the middle-east (Weatherall and Clegg 2001). In USA, 1 in 500 African-American children and 1 in 36,000 Hispanic children are born with SCD every year. It is estimate that 2.5 million Americans are heterozygous carriers of the sickle disease trait, occurring among 1:12 African-Americans and 1:100

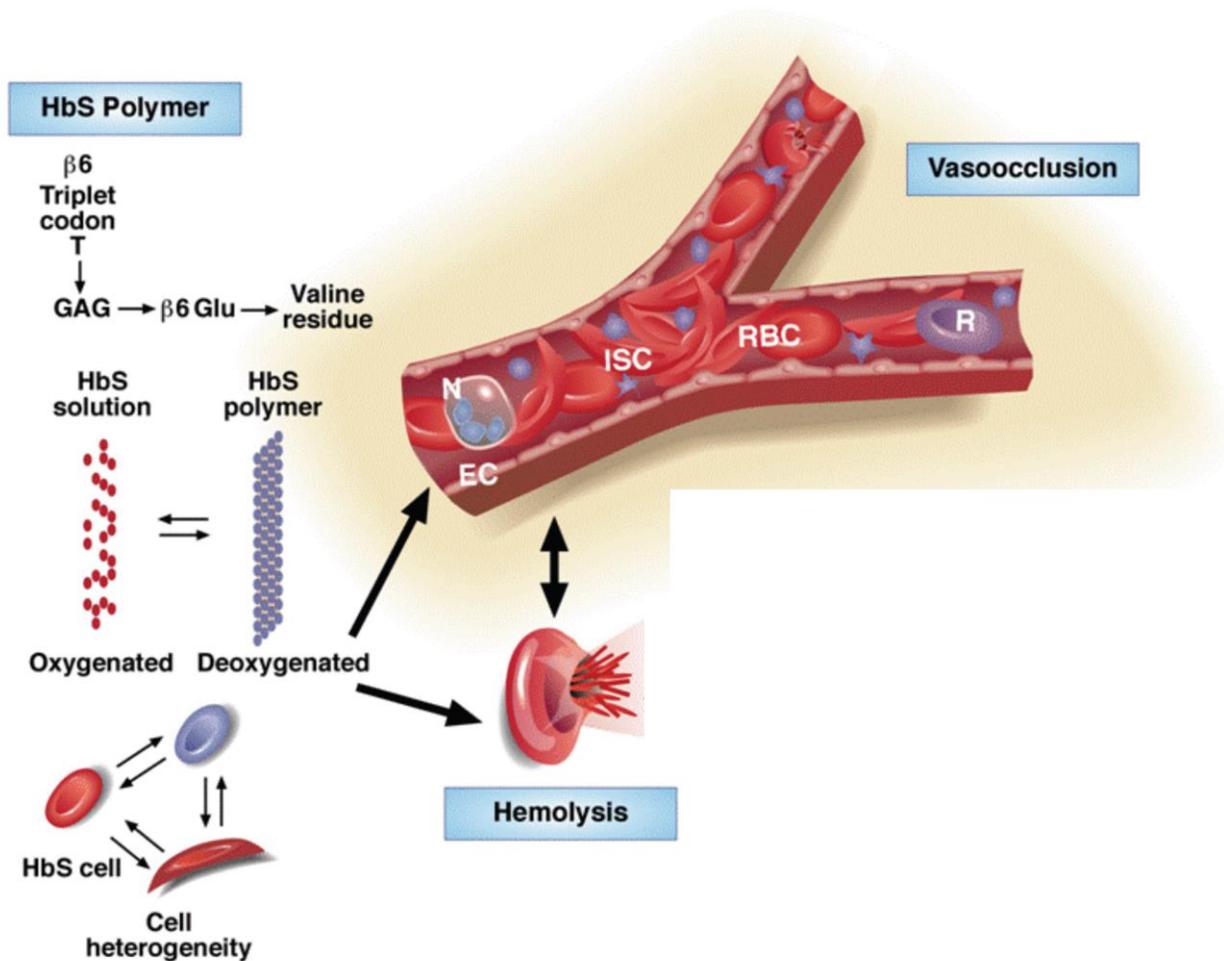


Figure 3. Pathophysiology of sickle cell disease. The HbS mutation, HBB glu6val, leads to β -globin chains that, when incorporated into hemoglobin tetramers with normal α -globin chains, produce a hemoglobin, HbS, which can undergo reversible polymerization when deoxygenated. The sickle polymer injures the erythrocyte and eventually produces irreversible membrane damage. The cells have a shortened life-span due to hemolysis. Adapted from Steinberg MH, 2008

Hispanic-Americans. About 90% patients with SCD survive to the age of 20 years and about 50% survive beyond their fifth decade. In the USA, the mean survival age for male patients with SCD is 53 years and that for female patients is 58 years.

B. β -thalassemias

β -thalassemias are one of the most common autosomal recessive monogenic disorders worldwide. β -thalassemias are characterized by reduced or absent production of the β -globin chains of hemoglobin that leads to microcytic anemia.

β -thalassemias are highly heterogenous at the molecular level; more than 200 different disease-causing mutations have been identified so far. The majority of mutations are single base-pair substitutions, deletions or insertions of few base-pairs leading to a frameshift. Mutations leading to deficient production of β -globin chains are categorized as: 1. Mutations causing defective β -globin gene transcription which includes mutations in the promoter and 5'-UTRs; 2. Mutations that cause affect its mRNA processing which includes splice-junction and consensus sequence mutations, polyadenylation and 3'-UTR mutations; 3. Mutations resulting in abnormal mRNA translation including non-sense, frameshift and initiation codon mutations (Stamatoyannopoulos et al. 2001).

β^0 -thalassemias are characterized by mutations that cause complete absence of production of the β -globin chains whereas β^+ -thalassemias results from reduced production of β -globin chains. The three clinically recognized forms of β -thalassemias in the order of increasing severity include: β -thalassemia trait (carrier status), β -thalassemia intermedia and β -thalassemia major. β -thalassemia minor (also β -

thalassemia trait) results from heterozygosity for the β -thalassemia mutations (β^0 or β^+) and is clinically asymptomatic. β -thalassemia intermedia consists of a genetically and clinically heterogeneous group of thalassemia-like disorders with severity ranging from asymptomatic to severe transfusion-dependent anemia. These individuals could be heterozygous or homozygous for any of the β -thalassemia genotypes. β -thalassemia major (also called Cooley's anemia for Thomas Cooley who first described it in 1925) is the most severe form of the disease leading to severe transfusion-dependent anemia and is caused by homozygosity for the β^0 alleles or compound heterozygosity for mild β -thalassemia alleles (Stamatoyannopoulos et al. 2001).

The clinical severity of β -thalassemias depend on the extent of imbalance between the alpha globin and non-alpha globin chains. In red blood cell precursors, the absent or reduced amounts of β -globin chains causes the unassembled alpha-globin chains to precipitate, which in turn leads to oxidative damage to the red cell membranes and apoptosis due to ineffective erythropoiesis. Anemia stimulates the production of erythropoietin leading to intensive but ineffective expansion of the bone marrow, which results in bone deformities. Prolonged anemia and increased erythropoietin stimulation in turn results in hepatosplenomegaly and extra-medullary hematopoiesis (Figure 4).

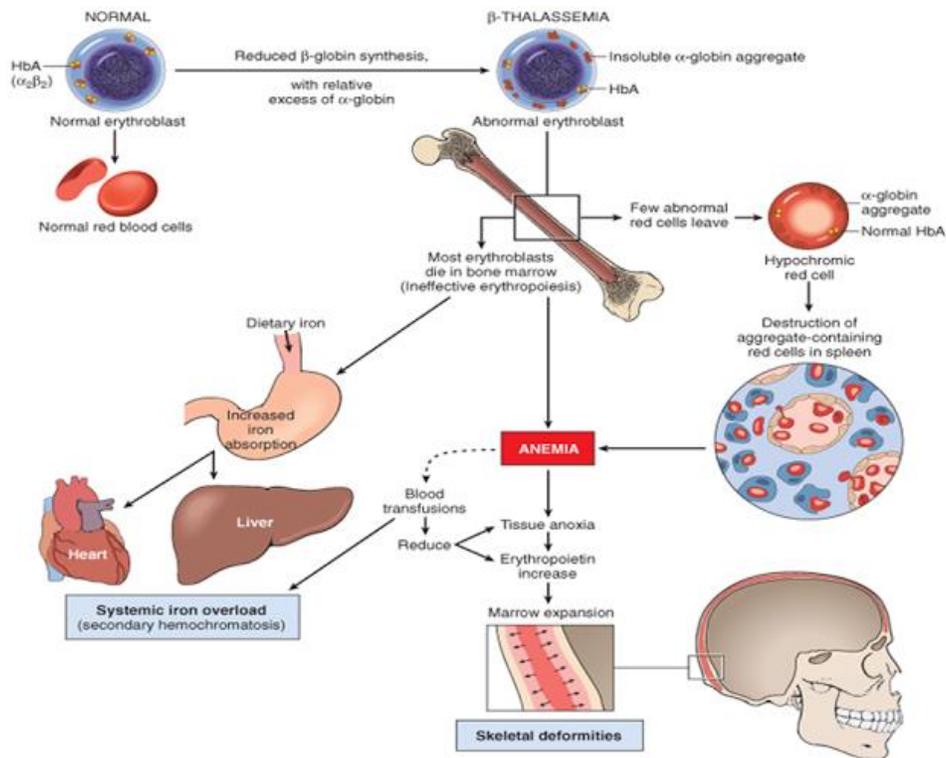


Figure 4. Pathophysiology of β -thalassemia. Reduced production of β -globin causes aggregation of the excess α -globin which leads to an abnormal erythroblast. This causes ineffective erythropoiesis leading to anemia, bone marrow expansion and skeletal deformities. Organ damage causes liver cirrhosis, cardiomyopathy and organ damage due to iron deposits. Iron overload due to blood transfusions leads to secondary hemochromatosis. Adapted from Pathologic basis of blood diseases, Aster JC 2003.

Individuals with β -thalassemia major usually come to medical attention within 1-2 years of age and require regular blood transfusions for survival. If a regular transfusion program is maintained, the growth and development of infants is normal till 10-11 years of age. After the age of 10-11 years, there is an increased risk for developing complications due to iron overdose, however it can be managed with adherence to iron chelation therapy. Complications of iron overload affect the heart (myocardiopathy and pericarditis), liver (chronic hepatitis, fibrosis and cirrhosis) and endocrine glands (resulting in diabetes and insufficiency of the thyroid and pituitary glands) and causes hypersplenism, osteoporosis and lung hypertension (Figure 4) (Weatherall and Clegg 2001).

Patients who undergo regular blood transfusions and iron chelation therapy survive well beyond 40 years of age. Myocardial disease is the major life-limiting complication of iron toxicity in β -thalassemias. An estimate suggests that 71% of deaths in patients with β -thalassemia major are caused by cardiac complications.

β -thalassemias are prevalent in Africa, the Mediterranean, the Middle-east, central Asia, India, Southern China and the Far East, with incidences as high as 10% in these regions presumably due to the adaptive association with malaria. Estimates indicate that about 1.5% of the global population are carriers of β -thalassemias and about 1 in every 100,000 people is symptomatic for the beta-thalassemia disease, with majority of the cases being recorded in developing countries (Weatherall and Clegg 2001).

C. Management of β -hemoglobinopathies:

i. Therapeutic options for β -hemoglobinopathies:

Management of sickle cell disease involves symptomatic treatment of acute events like vaso-occlusive crises and acute chest crises; prevention of bacterial infections; iron chelation to reduce the iron overload; folic acid supplementation; management of different forms of organ damage and stroke and controlling complications, if and when they occur. Newer therapeutic options include: 5-hydroxymethyl-2-furfural (5-HMF) which is a natural compound that binds to red blood cells and increases their oxygen , thus preventing the red blood cells from sickling (Abdulmalik et al. 2005). Adenosine A2A receptor antagonists have potent anti-inflammatory activity and are used for relieving pain-related symptoms due to vaso-occlusive crises in sickle cell disease (Field et al. 2013).

Treatment options for β -thalassemia disorders include regular blood transfusions, iron chelation therapy, splenectomy and folic acid supplementation. The primary focus of these therapies is to alleviate the symptoms of the illness.

Allogenic bone marrow or cord blood transplantation is the only curative therapy available for small number of patients with severe β -hemoglobinopathies. The overall survival following matched sibling donor hematopoietic stem cell transplants is excellent in children, with disease-free survival rates greater than 80%, and a graft failure rate of about 10% (Angelucci 2010; Shenoy 2013). However its therapeutic application is limited by the inability to find an HLA-identical sibling donor for the patients and also the concerns of morbidity and mortality related to the high risk procedure.

ii. Fetal Hemoglobin Stimulants:

Hereditary persistence of fetal hemoglobin (HPFH) is a naturally occurring condition in which the fetal γ -globin gene continues to be expressed well into adulthood without any pathological manifestations. Many deletion mutations resulting from large deletions at the β -globin gene cluster and non-deletion mutations resulting from single base mutations in the promoters of the $G\gamma$ and $A\gamma$ genes have been characterized in HPFH, however how some of these mutations affect γ -globin gene expression remains unknown. Rare individuals with β -hemoglobinopathies and HPFH do not manifest the disease because the fetal γ -globin gene can substitute for the functionally impaired adult β -globin gene. This phenomenon was first observed in infants with sickle cell disease who had fewer symptoms and their deoxygenated erythrocytes took longer to sickle; Watson attributed this effect to the presence of fetal hemoglobin in these cells (WATSON 1948). However, this effect lasted only for a few months after birth until adult β -globin gene expression became pre-dominant. The ameliorating effect of HbF was later noted in individuals who were co-inherited sickle cell disease and hereditary persistence of fetal hemoglobin. These individuals were clinically asymptomatic despite the presence of HbS (Edington and Lehmann 1955).

The protective effect of HbF was understood through studies which postulated that varying amounts of HbF would be beneficial in different complications of sickle cell disease. It was accepted that higher amounts of HbF had ameliorating effects in those symptoms of the disease that were associated with sickle vaso-occlusion and blood viscosity (Kato, Gladwin, & Steinberg, 2007). However, any increase in HbF had a beneficial effect on mortality (Platt et al. 1994; Powars et al. 1984). The pathophysiology

of sickle cell disease depends on the polymerization of the deoxygenated HbS, but HbF interferes with this process. Importantly, the Hb tetramer composed of both HbF and HbS ($\alpha_2\beta^S\gamma$) cannot enter the polymerized sickle state on deoxygenation. Also, the hybrid tetramer containing β^S and β^A globin chains have only half the probability of entering the polymer to form HbS. The anti-polymerization effect of HbF resides in its two critical residues: glycine $\gamma 87$ and aspartic acid $\gamma 80$. HbF prevents the tendency of deoxy sickle hemoglobin to polymerize, which in turn retards the many downstream pathophysiological effects caused by cellular damage due to HbS (Steinberg 2008).

Cell based therapies for induction of fetal hemoglobin as a therapeutic strategy for β -hemoglobinopathies began in the 1970s after cloning of the human globin genes and in the 1980s once the mechanisms regulating globin gene expression were elucidated. Epigenetic mechanisms like DNA methylation and histone deacetylation gained increased importance in developmental silencing of the fetal globin gene in adult erythrocytes. 5-azacytidine was the first agent to be used to induce HbF in adults patients. The rationale for using a DNA methylation inhibitor came from the fact that actively transcribed adult β -globin gene is hypomethylated whereas the non-transcribed fetal γ -globin gene is hypermethylated in adult life. DNA methylation of genes was shown to be inversely proportional to gene expression. 5-azacytidine was shown to induce very high levels of HbF in anemic baboons, which led to treatment of patients with sickle cell disease and β -thalassemias with 5-azacytidine (Charache et al. 1983; Ley et al. 1982; Lowrey and Nienhuis 1993). Although the adult patients with β -hemoglobinopathies showed promising induction of fetal hemoglobin, azacytidine was never tested in large-clinical trials due to concerns about potential carcinogenicity and

toxicity. Recently, decitabine (5-aza-2-deoxycytidine) - a newer and safer analog of 5-azacytidine, has been introduced in small scale clinical trials for patients with SCD and has shown promising results in inducing HbF and ameliorating the pathophysiology of SCD. However, larger long-term trials are required to confirm its safety and efficacy (Sauntharajah et al. 2003).

Hydroxyurea, an S-phase specific ribonucleotide reductase inhibitor and an approved chemotherapeutic agent has demonstrated great success in induction of HbF in adult patients with SCD (Charache et al. 1995; Dover and Charache 1992; Steinberg et al. 1997). Induction of HbF by hydroxyurea has been proposed to occur through various different mechanisms: perturbation of erythroid differentiation through stress erythropoiesis pathways, stimulation of cyclic GMP signaling pathway, increase in cAMP levels, induction of small GTP-binding protein and RAS-related (SAR) protein and induction of soluble guanylate cyclase through its nitrosylation (Atweh and Fathallah 2010). Currently, hydroxyurea is the only FDA approved treatment for adults with moderate or severe sickle cell disease, however its effects are highly variable (Ma et al. 2007).

Butyrate, a short chain fatty acid that inhibits histone deacetylases, was also investigated as a fetal hemoglobin stimulating agent in chicken, mice and baboons. Treatment of SCD patients with arginine butyrate also showed sustained stimulation of HbF. However, due to difficulty in the mode of administering large volumes of this drug its therapeutic use has been limited (Atweh and Fathallah 2010; Perrine et al. 1993). Recently, an oral butyrate derivative sodium 2,2-dimethyl butyrate (HQH-1001) was shown to induce γ -globin expression and was tolerated in phase I/II clinical trials in

patients with β -thalassemia (Fucharoen et al. 2013). Although butyrates and its derivatives have promising results, the effect is often variable among patients.

iii. Futuristic therapeutic options:

The lack of effective treatments for β -hemoglobinopathies has encouraged efforts to investigate newer therapeutic options. Gene therapy has been promising in correcting many immunodeficiency diseases like X-linked severe combined immunodeficiency (X-SCID), Wiskott-Aldrich syndrome (WAS) and chronic granulomatous disease (CGD). Gene therapy for β -hemoglobinopathies is in the early phases of development and preliminary results have been encouraging. The first successful gene-therapy for a transfusion-dependent patient with hemoglobin-E- β -thalassemia was done using a self-inactivating lentivirus (SIN-LV)-based β^{T87Q} vector and reported in a French trial (Cavazzana-Calvo et al. 2010). Synthetic β -globin (β^{T87Q}) has anti-sickling properties. An initial safety concern was raised due to activation of the HMGA2 locus in the erythroid cells, however it gradually reduced to only 2% to 3% of the circulating nucleated cells. At 5 years after gene therapy, this patient is transfusion-independent. One more patient has been enrolled in the program and results are awaited. In the United States, a phase 1 to 2 β -thalassemia gene therapy trial was started by Sadelain and colleagues at Memorial Sloan Kettering Institute in 2011; 2 patients received gene-corrected CD34+ cells and results are awaited. Following successes in β -thalassemia gene-therapy, similar studies were begun for sickle cell disease. Townes et al. developed anti-sickling β -globin by introducing three substitutions to the normal β -globin (T87Q, E22A, and G16D); the recombinant β -globin designated β^{AS3} inhibits sickling of

deoxygenated hemoglobin and also provides a competitive advantage compared with β -globin for its interaction with alpha-globin (Levasseur et al. 2004). Correction of the SCD phenotype in a murine model by expressing β^{T87Q} with an LV vector was reported for the first time by Pawlik and colleagues (Levasseur et al. 2003). Recently, Romero and group used a LV carrying the recombinant β^{AS3} to show consistent expression of the modified β -globin and reversal of RBC properties in red cells differentiated from CD34+ cell derived from the bone marrow of a sickle cell disease patient (Romero et al. 2013). Malik and group will be initiating clinical trials for patients with SCD using a LV vector carrying the γ -globin gene driven by the β -globin LCR and a 130-bp β -globin promoter (Perumbeti et al. 2009; Pestina et al. 2009).

Induced pluripotent stem cells (iPSCs) combined with the powerful tool of genome editing offers many exciting avenues for correcting genetic disorders using autologous somatic cells. iPSCs offer the advantage of an endless supply of stem cells for genome editing followed by the possibility to select the ideal clone with safe integration and high-efficiency expression (Papapetrou et al. 2011). Zinc-finger nucleases (ZFNs), transcription-activator like nucleases (TALENs) and clustered regulatory interspersed short palindromic repeat (CRISPR/Cas) endonuclease systems have been used for genome editing. Proof-of-concept studies have showed ZFN-mediated correction of β -thalassemia in human iPSCs (Chang and Bouhassira 2012), *in situ* ZFN and TALEN enabled correction of SCD mutation in iPSCs derived from an SCD patient (Sebastiano et al. 2011) and correction of the sickle cell mutation in HBB gene in human iPSCs using TALENs and *piggyBac* transposon (Sun and Zhao 2014). A very recent study published correction of β -thalassemia causing HBB mutations in patient-derived iPSCs

using CRISPR/Cas9 and *piggyBac* technologies (Xie et al. 2014). Although the genome editing approach is very promising, more work is needed to establish the safety and efficacy of these methods in gene therapy.

IV. Human β -globin gene regulation

Molecular therapies which target developmental silencing of the fetal γ -globin gene are also promising because they pose less toxicity and better effectiveness, which necessitates a detailed understanding of the process of hemoglobin switching. Various regulators of the human β -globin gene locus are discussed in detail.

A. Cis-acting sequences

The locus control region was first identified as DNase-I hypersensitive sites located approximately 6 to 20 kb upstream of the β -globin genes and necessary for expression of the β -globin genes. The LCR consists of 5 DNase-I hypersensitive sites and each HS has a core sequence approximately 250 bp long which contains motifs for binding of various transcriptional factors. It was initially proposed through studies in transgenic mice that the LCR was required for opening the β -globin gene locus and for its active transcription (Grosveld et al. 1987), however it was later shown that the LCR is primarily required for high level expression of the β -globin genes (Bender et al. 2000). The LCR confers lineage-specific expression of the β -globin genes while also acting as an enhancer of the β -globin locus promoting its high expression and serving as an insulator for the locus to protect it from surrounding heterochromatin (Grosveld et al. 1987). Of the 5 HSs, HS2 is important for the position independent activity of the LCR (Collis,

Antoniou, Grosveld 1990), whereas HS3 is essential for its enhancer activity (Ellis et al. 1996).

Studies in transgenic mice have identified two potential mechanisms for the function of LCR in regulating expression of the β -globin gene locus: Gene competition (Behringer et al. 1990; Enver et al. 1990) and Autonomous silencing. The competitive mechanism (also called chromosome looping) suggests that specific genes from the β -globin gene locus preferentially interact with the LCR at a given developmental stage depending on its proximity to the LCR, leading to its active transcription. The γ - and β -globin genes compete with the LCR during fetal and adult erythropoiesis such that during the fetal stage there is preferential interaction between the γ -globin gene and the LCR whereas in the adult stage the LCR interacts with the β -globin gene, silencing the γ -globin gene. The autonomous silencing mechanism suggests that all factors necessary for turning off gene expression are available within the canonical gene or nearby sequences (Raich et al. 1990). The ϵ -globin gene is highly regulated by autonomous gene silencing by various stage-specific factors (Raich et al. 1990). Autonomous silencing is the primary mechanism that turns off γ -globin gene expression in adult erythropoiesis; competition by the β -globin gene also has a contributory effect (Dillon and Grosveld 1991). The β -globin gene is mainly regulated by chromatin looping leading to its robust expression in adult erythropoiesis (Dillon and Grosveld 1991).

B. Trans-acting factors:

GATA1 and FOG1:

GATA1 is a central regulator of essential genes involved in erythropoiesis. GATA1 knockout mice die during gestation at E12.5 due to ineffective hematopoiesis in the yolk sac (Takahashi et al. 1997); GATA1 null-cells of primitive or definitive lineage fail to develop beyond the pro erythroblast stage (Fujiwara et al. 1996). GATA1 was among the first factors to be identified to have a role in hemoglobin switching through its binding at the β -globin gene locus (Evans and Felsenfeld 1989; Martin, Tsai, Orkin 1989; Tsai et al. 1989). GATA1 may be involved in hemoglobin switching through facilitating a chromatin loop formation between *Kit* and β -globin gene (Jing et al. 2008; Vakoc et al. 2005).

A SNP located 567 base-pairs 5' to the $\text{G}\gamma$ globin gene was identified in an Iranian-American family that showed heredity persistence of fetal hemoglobin (Chen et al. 2008). This SNP disrupted the GATA-binding motif in the far upstream promoter of the $\text{G}\gamma$ -globin gene, and a similar SNP is also present in the $\text{A}\gamma$ -globin gene located at 566 base-pairs upstream (5') to $\text{A}\gamma$ -globin start site (Harju-Baker et al. 2008). Disruption of the -566 $\text{A}\gamma$ GATA-binding site in β -YAC transgenic mice caused elevated expression of fetal hemoglobin in adult mice. Later, Mi2 β was identified to be an essential component for gene repression at this site (Costa et al. 2012). GATA1 has been shown to function as a transcriptional activator as well repressor in erythroid cells through interaction with its cofactor Friend of GATA1 (FOG1) and the MeCP1 complex (Miccio et al. 2010; Miccio and Blobel 2010; Rodriguez-Paredes and Esteller 2011).

FOG1 is a zinc finger protein which is critical to the function of GATA1 and important for the normal differentiation and maturation of megakaryocytic and erythroid precursor cells (Tsang et al. 1998). FOG1 knockout mice die due to fatal anemia at E10.5-11.5, similar to GATA1 knockout mice (Tsang et al. 1998). The functional interaction between GATA1 and FOG1 is highlighted by a mutation in the N-terminal zinc-finger of GATA1 that impairs FOG1 binding; patients and mice with this mutation suffer from severe anemia (Hong et al. 2005). The N-terminal region of FOG1 recruits the NuRD co-repressor complex to mediate the transcriptional repression by GATA1; mutation of the residues critical for interaction between FOG1 and NuRD impairs its gene repression activity as well repression by GATA1 (Hong et al. 2005).

BCL11A , HBS1L-MYB and miRNAs-15a and 16-1:

Because hereditary persistence of fetal hemoglobin is a quantitative trait, genome-wide association studies were applied in subjects with and without β -hemoglobinopathies, to identify novel regulators of the γ -globin gene. These studies identified common variation at two loci beyond the β -globin gene cluster, *BCL11A* on chromosome 2 (Lettre et al. 2008; Menzel et al. 2007; Uda et al. 2008) and the *HBS1L-MYB* intergenic region on chromosome 6 (Thein et al. 2007; Thein et al. 2009). Variants at these two loci are also associated with severity of the β -globin disorders, with higher levels HbF associated with milder phenotypes. *BCL11A* is a zinc finger protein and functions as a transcriptional repressor. *Bcl11a* knockout mice show impaired silencing of endogenous murine embryonic globin genes and the transgenic human fetal globin genes (Sankaran et al. 2009). Knockdown of *BCL11A* in primary human erythroid cells show elevated

expression of the γ -globin gene, thus establishing a role for BCL11A in maintaining repression of the γ -globin gene in adult erythroid cells (Sankaran et al. 2008). BCL11A interacts with many transcriptional regulators like GATA1, FOG1, SOX6, other erythroid transcription factors and the NuRD complex to bring about silencing of the murine embryonic and human fetal globin genes (Xu et al. 2010). BCL11A exerts its repressive effects in part by occupying critical sites in the β -globin gene locus (intergenic δ - β sequences which are deleted in HPFH), which promote long-range interactions between the LCR and the β -globin gene, at the expense of the γ -globin genes (Xu et al. 2010). Recent studies have identified a developmental stage-specific erythroid-lineage restricted enhancer of BCL11A through analysis of SNPs in regulatory DNA elements such that loss of this enhancer results in absence of BCL11A expression in erythroid cells. Common genetic variants of this enhancer display modestly reduced BCL11A expression, transcription factors GATA1 and TAL1 binding and increased HbF (Bauer et al. 2013).

The HBS1L-MYB locus regulates the γ -globin gene primarily through the effect of MYB (Jiang et al. 2006; Wahlberg et al. 2009). c-MYB is a hematopoietic transcription factor essential for definitive hematopoiesis, is highly expressed in hematopoietic stem cells and progenitors and its expression reduces as the erythroid cells mature (Vegiopoulos et al. 2006). Knockdown of MYB in primary adult human erythroid cells has been shown to increase HbF expression, possibly through its activation of KLF1/BCL11A and TR2/TR4 pathways (Suzuki et al. 2013). A recent study analyzed several HBS1L-MYB intergenic variants and showed that these variants reduce critical transcription factor

binding, affecting long-range interactions by MYB and also MYB expression levels (Stadhouders et al. 2014).

Trisomy of chromosome 13 is one of the few viable trisomies and is associated with a delayed fetal-to-adult hemoglobin switch and persistent elevated expression of HbF. Analysis of cases of partial trisomy 13 using an integrated genomic approach identified a precursor RNA (*DLEU2*) for two microRNAs, 15a and 16-1, with identical seed targeting sequences as the top candidate in the region of 13q14. Increased expression of miR-15a/16-1 in hematopoietic progenitor cells caused a modest increase in γ -globin gene expression. This effect was shown to be mediated via the *MYB* gene, wherein *MYB* was identified as a direct target of the miR-15a/16-1 (Sankaran et al. 2011).

SOX6:

SOX6 is a member of the Sry-related high-mobility-group (HMG) box transcription factors which serve as important regulators of cell fate and differentiation in various lineages. The role of SOX6 in globin gene regulation was first identified in Sox6 knockout mice; these mice show increased expression of embryonic $\epsilon\gamma$ -globin gene in the fetal livers, while the β h1 gene was rapidly down-regulated. Sox6 represses the $\epsilon\gamma$ -globin gene by directly binding at its promoter region (Yi et al. 2006). Sox6 has been suggested to enhance erythroid differentiation by stimulating erythroid survival, proliferation and terminal maturation (Dumitriu et al. 2006). A role for SOX6 in human globin gene regulation has not been elucidated yet, however recent studies in human erythroid progenitors has suggested a correlation between varying levels of SOX6 and the expression of the γ -globin gene (Sripichai et al. 2009). SOX6 has also been shown

to co-occupy the β -globin gene locus with BCL11A and silence the expression of γ -globin gene in adult erythroid cells (Xu et al. 2010). Knockdown of SOX6 in these adult erythroid cells demonstrated that SOX6 has a modest contribution in maintaining silencing of the γ -globin gene in the adult erythroid cells (Xu et al. 2010). The exact mechanism of transcriptional repression of the γ -globin gene by SOX6 is not known, however the Sox family of transcription factors interact with the minor groove of DNA and may effect conformational changes in DNA. Thus SOX6 may function as an architectural protein by enabling long range DNA interactions along the β -globin locus and assembling local chromatin structures with multi-protein complexes.

KLF1:

KLF1 (Krüppel-like factor 1), also known as EKLF (Erythroid Krüppel-like factor), belongs to the Krüppel-like zinc-finger containing family of transcription factors. KLF1 was first identified in a subtractive hybridization screen for genes highly expressed in erythroid vs. myeloid lineages (Miller and Bieker 1993). KLF1 has been shown to play critical roles in erythropoiesis and β -globin gene regulation (Miller and Bieker 1993; Nuez et al. 1995; Perkins, Gaensler, Orkin 1996; Wijgerde et al. 1996). KLF1^{-/-} mice show severe anemia and death *in utero* at E14.5 during development (Nuez et al., 1995). KLF1 has been shown to bind the CACCC binding motif in the proximal promoter of the β -globin gene and the HS sites, HS1, HS2 and HS4, in the β -globin LCR *in vivo* (Miller and Bieker 1993). Studies by Borg et al. linked a mutation in the first C2H2 zinc-finger of KLF1 to HPFH in a Maltese family; individuals heterozygous for this mutation showed up to 19.5% expression of HbF into adulthood (Borg et al. 2010). Other

mutations in the KLF1 gene promoter and another mutation affecting its splicing have been identified, both of which cause haploinsufficiency of KLF1 and elevated expression of HbF, suggesting a critical role of KLF1 in hemoglobin switching (Gallienne et al. 2012; Radmilovic et al. 2013; Satta et al. 2012; Singleton, Frayne, Anstee 2012). Haploinsufficient *Klf1* transgenic mice carrying the human β -globin gene locus show elevated HbF without a significant change in β -globin expression and erythropoiesis, an effect that is mediated by Klf1 binding the Bcl11a promoter and acting as a positive regulator of Bcl11a (Zhou et al. 2010). Interestingly, *Klf1*(wt/ko)::*Bcl11a*(cko/cko) mice also show increased expression of murine embryonic genes and human γ -globin gene during fetal development. However, after birth these fetal and embryonic globin genes show a significant reduction in expression, being maintained at modestly higher levels than control mice during adult life, possibly because of a role of other transcriptional repressors of the embryonic and fetal β -type globin genes in adulthood (Esteghamat et al. 2013). Thus, two possible mechanisms have been suggested for regulation the fetal-to-adult hemoglobin switch by KLF1: a) Preferential activation of the adult β -globin genes by establishing an adult stage-specific three-dimensional locus configuration through chromatin looping; and b) Indirect repression of the fetal γ -globin gene by activation of transcription factors like BCL11A which serve as repressors of γ -globin gene expression.

NF-E4:

NF-E4 was identified as a regulator of the chicken β -globin gene through its interaction with the stage specific element (SSE) in the chicken β -globin gene promoter (Zhou et al., 2004). It was later characterized as a member of the stage selector protein (SSP) heterodimeric complex which also contains the ubiquitous transcription factor CP2 (Zhou et al. 2004). Several SSP binding sites were identified in the HS2, HS3, ϵ -globin promoter and in an HPFH point mutation, which implicated a role for p22-NF-E4 (human homologue of chicken NF-E4) in human γ -globin gene regulation. Enforced expression of p22-NF-E4 in β -YAC transgenic mice caused a delay in the switch from fetal to adult globin genes, however the switch was eventually completed in the adult bone-marrow. Enforced expression of a short isoform of NF-E4 (p14-NF-E4) in human erythroid progenitor cord blood cells was reported to reduce the expression of the fetal γ -globin gene (Zhao et al. 2006).

PRMT1/5:

Protein arginine methyltransferase 5 (PRMT5) induces a repressive histone mark H4R3me2s, which serves as a template for binding of DNMT3A and subsequent DNA methylation (Zhao et al. 2009a). Reduction of PRMT5 or its enzymatic activity leads to increased expression of the γ -globin gene (Rank et al. 2010a). PRMT5 has also been reported to occupy the γ -globin gene promoter and its enzymatic activity is critical in recruiting a multi-protein repressive complex to the γ -globin promoter, an event which induces various repressive epigenetic marks. Disruption of this multi-protein complex results in reactivation of the fetal γ -globin gene (Rank et al. 2010a).

FOP:

Friend of PRMT1 (FOP) is a chromatin factor important in transcriptional regulation, including estrogen-dependent gene induction in breast cancer cells (van Dijk et al. 2010). It is a target of methylation by PRMT1 and PRMT5 arginine methyltransferases. Knockdown of FOP in murine fetal liver cells and human erythroid progenitor cells shows increased expression of the γ -globin gene (van Dijk et al. 2010). The exact mechanism of action of FOP in γ -globin expression is not clear, although it is presumed to function as a target for PRMT mediated regulation of the γ -globin gene.

DRED, TR2/TR4 and LSD1:

The embryonic and fetal globin genes, but not the adult β -globin genes, contain direct repeat (DR) elements in their promoters which serve as binding sites for non-steroidal nuclear receptors. Numerous mutations in the G γ and A γ globin gene promoter DR elements have been associated with specific HPFH phenotypes. Introduction of these DR element promoter mutations in β -YAC transgenic mice leads to increased expression of the respective ϵ - and γ -globin genes. In an attempt to identify potential repressive proteins that may bind these DR *cis* elements, the DR-element binding protein was identified as a complex between the erythroid-specific transcription factor NF-E3 and a member of the COUP transcription family member COUP-TFII. Named the DR-element erythroid specific (DRED) complex hereafter, this complex was characterized to be composed of a heterodimer of the orphan nuclear receptors TR2/TR4 (Tanimoto et al. 2000). TR2 and TR4 bind the DR1 sites in the embryonic ϵ -globin gene promoter and the single DR1 site in the fetal γ -globin gene promoter

(Tanabe et al. 2002). Silencing of the embryonic and fetal globin genes was shown to be delayed in *Tr2* and *Tr4* null mice (Tanabe et al. 2007), however paradoxically, forced over expression of TR2 and TR4 in mice also lead to induction of the fetal globin gene and a modest alleviating effect in the phenotype of SCD mouse model (Campbell et al. 2011; Tanabe et al. 2007). The mechanism of action underlying this phenomena is yet unknown, but two possible explanations have been proposed: 1. The over-expression of TR2 and TR4 may inhibit formation of the DRED complex due to limited availability of one or more of the DRED complex co-factors 2. TR2 and TR4, in addition to being repressors, may function as activators of the fetal gamma-globin gene by association with other cellular activators like PGC-1.

TR2 and TR4 associate with co-repressors DNMT1 and LSD1 (lysine-specific demethylase 1) which form a core heterotetrameric complex (Cui et al. 2011). LSD1 removes methyl-groups from mono- and dimethyl histone H3 lysine 4 which is an activating epigenetic mark. Knockdown of LSD1 in primary adult human erythroid cells can modestly activate γ -globin gene expression, however LSD1 is important for normal erythroid differentiation and LSD1 knockdown erythroid cells show delayed erythroid differentiation (Xu et al. 2013). Pharmacologic inhibition of LSD1 by Tranylcypramine (TCP)- a monoamine oxidase inhibitor was also shown to induce γ -globin gene expression *in vitro* and *in vivo* (Shi et al. 2013), however it poses the risk of hematopoietic toxicity and significant side effects. Next generation TCP compounds such as could be tested for better efficiency in γ -globin gene induction with fewer side

effects. However, its application may be limited due to the importance of LSD1 in normal hematopoiesis.

Ikaros:

The transcription factor Ikaros (also known as Lyf-1) is a nuclear regulator that has been implicated in control of early hematopoiesis. Ikaros is associated with the chromatin remodeling PYR complex which binds the intergenic region between γ - and β -globin genes that is involved in regulating the fetal-to-adult hemoglobin switch (O'Neill et al. 2000). The PYR complex consists of the NuRD and SWI/SNF chromatin remodeling complexes (O'Neill et al. 2000). Ikaros null mice show a modest delay in silencing of the murine embryonic globin genes and the human fetal globin genes in transgenic mice carrying the human β -globin mini locus, possibly due the absence of the PYR complex (Lopez et al. 2002). Mice carrying a specific mutation in Ikaros in the third zinc finger of its N-terminal domain show severe anemia and die between E15.5 to E17.5 due to failure in normal erythroblast growth and differentiation (Dijon et al. 2008).

Mi2 β :

Mi2 β , the chromatin remodeling component of the NuRD complex, belongs to the SNF2 family of DNA helicases (Eisen, Sweder, Hanawalt 1995). In addition to the NuRD complex, it is also a component of the PYR complex which has been associated in regulating the hemoglobin switch (O'Neill et al. 2000). Recent studies by our lab established a role for Mi2 β in regulation of the fetal γ -globin gene, in part through the NuRD complex, but largely as a positive regulator of KLF1 and BCL11A proteins such

that even a partial knockdown of Mi2 β in primary adult erythroid cells results in ~25% to 30% $\gamma/(\gamma+\beta)$ globin gene expression and 13.2% HbF protein levels as compared to control cells expressing <1% $\gamma/(\gamma+\beta)$ globin gene (Amaya et al. 2013). A following study by another group also identified Mi2 β as one of the most potent regulators of HbF in adult erythroid cells (Xu et al. 2013). However, the therapeutic potential of Mi2 β may be limited by its important roles in normal development.

As can be noted, a complex interplay of several different factors is involved in the developmental silencing of the fetal γ -globin gene in adult erythroid cells. Epigenetics comprises another layer of regulation of γ -globin gene expression.

V. Epigenetics:

Epigenetics (Greek: epi - beyond; genetics) is defined as changes in heritable gene expression brought about by mechanisms other than the canonical changes in the DNA sequence. These changes can involve chemical modifications to DNA or the proteins that closely interact with the DNA and a newly emerging field of functional non-coding RNA has also gained recognition as an epigenetic mechanism. Epigenetic mechanisms have been implicated in various biological processes, including both normal developmental processes like genomic imprinting, X-chromosome inactivation and cellular differentiation and pathological phenomena like imprinting disorders, cancer and developmental abnormalities. DNA methylation and histone modifications are the most extensively studied mechanisms; many therapeutics like 5-azacytidine and HDAC inhibitors targeted at these mechanisms have been clinically explored.

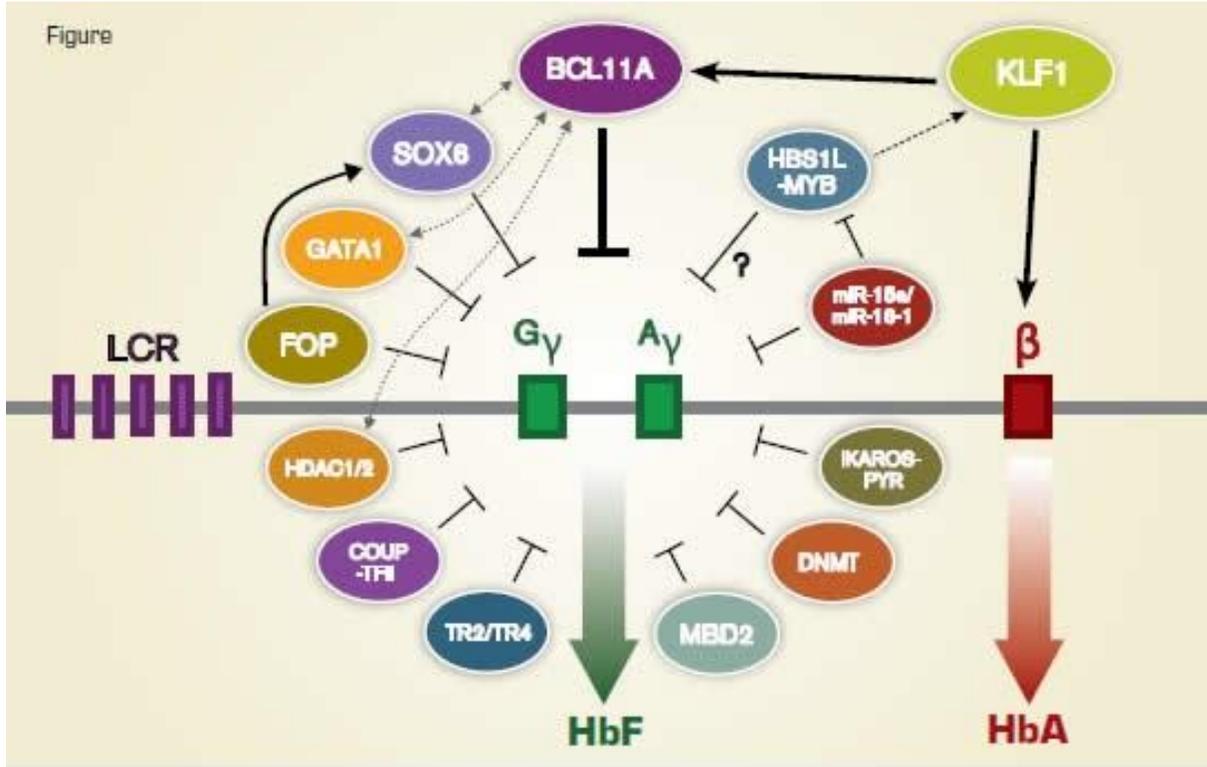


Figure 5. Trans-acting factors regulating hemoglobin switch. The switch from fetal (HbF) to adult (HbA) hemoglobin is regulated by various transcription factors and cofactors. Regulators of this process constitute potential therapeutic targets for patients with major hemoglobin disorders. Positive and negative interactions are denoted by printed and blunt arrows, respectively. Adapted from Xu, Bauer and Orkin 2011.

A. DNA Methylation:

DNA methylation involves addition of a methyl moiety to the 5th carbon of a cytosine residue in the DNA backbone. DNA methyl transferase enzymes are the epigenetic writers that establish the DNA methylation signature of genes and include the well characterized DNMT1 and DNMT3A and DNMT3B. DNMT1 is the maintenance methylase which adds a methyl moiety to the symmetric CpG on hemimethylated DNA during DNA replication. DNMT3A and DNMT3B are *de novo* methylases which symmetrically add a methyl group to unmethylated DNA (Kim, Samaranyake, Pradhan 2009). CpG dinucleotides are methylated throughout the genome at ~70-80% of CpG cytosines (Jabbari and Bernardi 2004), however there are clusters of non-methylated “CpG-islands” which are present in the promoter regions of 60-70% of all human genes (Saxonov, Berg, Brutlag 2006). Differential methylation of these CpG islands plays a regulatory role in developmental as well as tissue-specific genes (Deaton and Bird 2011). Aberrant methylation of CpG islands of tumor suppressor genes also has been shown to lead to carcinogenesis (Esteller 2008).

i. Methyl Cytosine Binding Proteins:

The DNA methylation signature of chromatin is read by the methyl-CpG binding domain (MBD) containing family of proteins. These proteins bring about repression of their target genes by recruiting associated co-repressor complexes containing histone deacetylase and nucleosome remodeling activities or by directly inhibiting interaction with the transcriptional machinery (Figure 6). MeCP2 was the founding member of the MBD family of proteins; later MBD1, MBD2, MBD3 and MBD4 were also identified as

members of this family of proteins (Hendrich and Bird 1998; Lewis et al. 1992). The structures of these proteins bear little resemblance to each other outside of their ~75 amino acid MBD motif. Another group of proteins known as the Kaiso and Kaiso-like proteins (ZBTB4 and ZBTB38) recognize and bind methyl cytosine residues through their conserved zinc finger motifs in a sequence specific manner (Filion et al. 2006; Prokhortchouk et al. 2001).

MeCP2: MeCP2 has been characterized to be an intrinsically disordered protein and mutations in all of its 6 domains are known to cause the X-linked condition Rett syndrome (Miltenberger-Miltenyi and Laccone 2003). The MBD of MeCP2 can recognize and bind even a single methylated CpG and through its transcriptional repression domain (TRD) it recruits the SIN3A co-repressor complex containing histone deacetylases (Jones et al. 1998; Nan et al. 1998). Although MeCP2 plays a vital role in neurological development, *Mecp2*^{-/-} female mice are viable and born normal but progressively develop symptoms of Rett syndrome (Guy et al. 2001).

MBD1: MBD1 has been shown to bind mCpG residues as well as unmethylated DNA based on its alternatively spliced isoform and has been shown to interact with co-repressor proteins SETDB1, MCAF, HDAC3 and MPG through its TRD (transcriptional repression domain) (Fujita et al. 2003; Nakao et al. 2001; Sarraf and Stancheva 2004). MBD1 null mice also have no developmental defects except minor neural defects and reduced genomic stability (Zhao et al. 2003). The C-terminus of MBD1 (MBD1-c) was recently shown to be intrinsically disordered but it governs selective interaction of MBD1 with its co-repressors proteins (Hameed et al. 2014).

MBD2 and MBD3: Early studies of the MBD2 and MBD3 proteins in transgenic mice revealed that *Mbd2*^{-/-} mice were viable and fertile with a defect in maternal nurturing whereas *Mbd3*^{-/-} mice were embryonic lethal, indicating non-overlapping functions of the MBD2 and MBD3 proteins in developmental gene regulation (Hendrich et al. 2001). MBD2 and MBD3 proteins share ~70% sequence homology and both proteins are highly expressed across various tissues types in adult mice as well as during embryogenesis (Hendrich and Bird 1998). MBD2 specifically binds to methylated DNA in vitro and in vivo (Hendrich and Bird 1998) and its binding orientation and affinity is influenced by the DNA sequence adjacent to the central mCpG (Scarsdale et al. 2011a). All MBD proteins except MBD3 have been shown to bind methyl-CpG residues, although with varying affinities; MBD3 has a sequence variation in its conserved MBD which impairs its ability to interact with methylated DNA (Hendrich and Bird 1998). MBD3 was shown to have high affinity for hydroxy-methylated DNA, however these studies may require further work to draw a stronger conclusion (Baubec et al. 2013; Yildirim et al. 2011). Recent work has shown that MBD3 preferentially localizes to methylated CpGs and to a lesser extent to non-methylated CpGs, while it does not distinguish between hydroxy-methylated and non-methylated DNA (Cramer et al. 2014).

MBD4: MBD4 binds to methylated DNA but is primarily involved in identifying T:G mismatches caused by deamination of the cytosine residue in mCpG dinucleotides and brings about DNA repair through its DNA glycosylase activities (Hendrich et al. 1999; Petronzelli et al. 2000). MBD4 deficient mice are viable and fertile, but display increased mutability in CG → TA transitions at CpG sites (Millar et al. 2002).

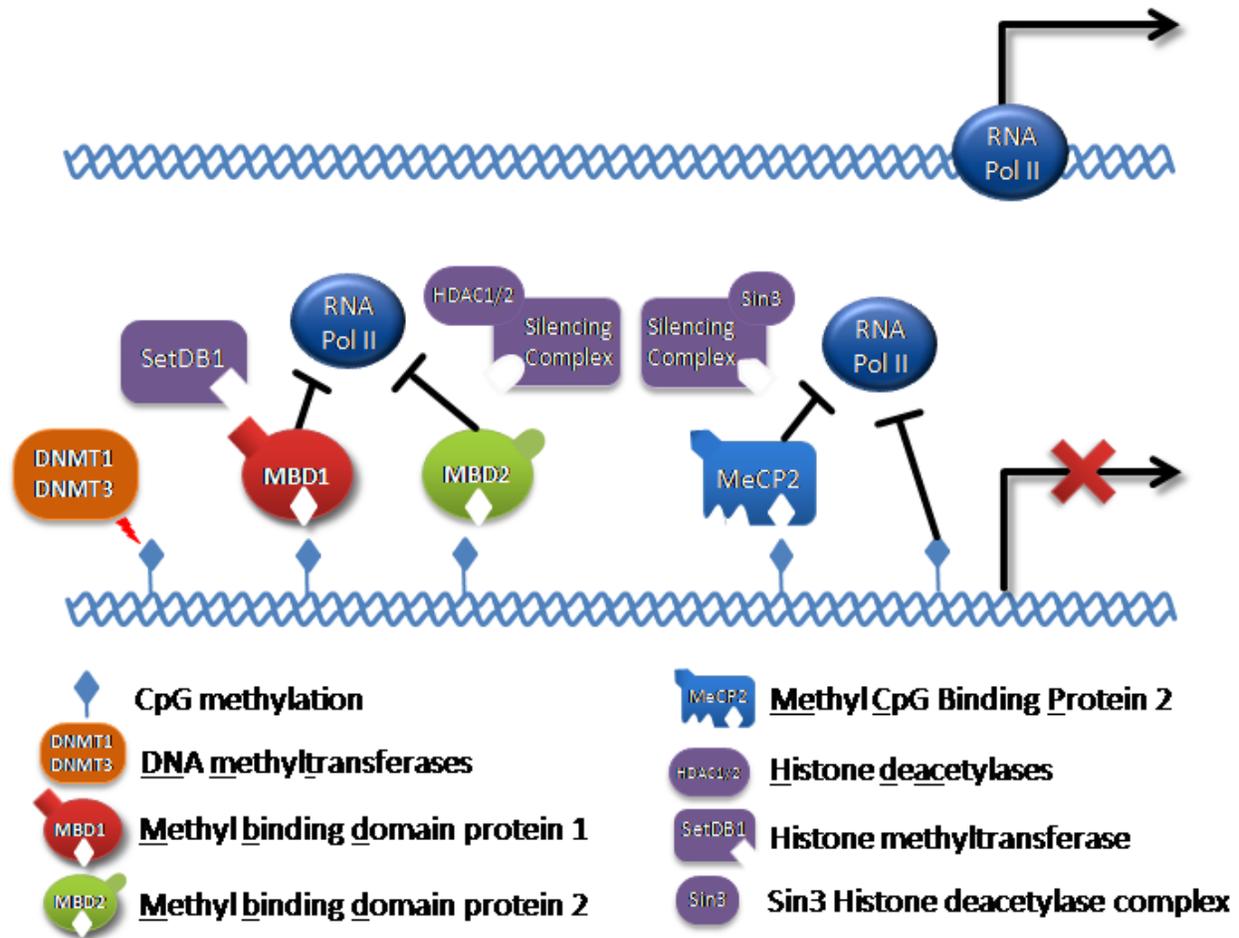


Figure 6. The role of DNA methylation in gene silencing. During development, a number of CpG rich promoter regions are methylated by DNMT enzymes. DNA methylation hampers RNA Pol II from advancing. DNA methylation also recruits methyl-binding domain proteins such as MBD2 and silencing complexes such as NuRD complex, which in turn alter histone modification patterns, further preventing gene expression. Adapted from Ginder, Gnanapragasam and Mian 2008.

Mutations in MBD4 have been reported in 26 to 43% of human colorectal tumors that show micro satellite instability (Bader, Walker, Harrison 2000; Riccio et al. 1999).

Likewise, MBD4 deficient mice heterozygous for *Min* allele of adenomatous polyposis coli gene (*Apc*^{Min}) show increased mutagenesis due to genomic instability (Millar et al. 2002).

B. Histone Modifications

Histones are alkaline proteins that package and order DNA into structural units called nucleosomes. There are five major categories of histones: H2A, H2B, H3 and H4 are core histones while H1 and H5 are linker histones. Histones allow regulation of gene expression by undergoing various histone modifications like acetylation, methylation, phosphorylation, SUMOylation, ubiquitination and ADP-ribosylation. These modifications are added to the N-terminal tails of histones through post translational modifications by specific enzymes. Histone acetylation and methylation are among the most well characterized modifications. Trimethylation of H3K4, H3K36 and H3K79 are associated with actively transcribed genes whereas trimethylation of H3K9, H3K27 and H4K20 have been associated with transcriptionally silenced genes. Similarly, histone acetylation is a mark of euchromatin where histone deacetylation is associated with heterochromatin. Combinations of such modifications are known to constitute the histone code. The histone code is recognized by large number of multi-protein complexes which are involved in functions like DNA repair, gene regulation, chromosome condensation and others.

In regulation of the β -globin gene locus, histone methyltransferases PRMT1 (van Dijk et al. 2010) and PRMT5 (Rank et al. 2010b; Zhao et al. 2009b) have been extensively studied. Lysine-specific demethylase 1 (LSD1) has also been identified as a silencer of the γ -globin gene in adult erythroid cells (Shi et al. 2013). Histone deacetylase inhibitors have a promising role in reactivation of HbF and recent studies have identified specific inhibitors of HDAC1 and HDAC2 as γ -globin gene inducers (Bradner et al. 2010).

C. MicroRNAs

MicroRNAs are endogenous ~22 nucleotide non-coding RNAs that have functional roles in gene regulation through mRNA cleavage or translational repression. The human genome codes for over 1000 miRNAs which may target about 60% of human protein-coding genes. miRNAs identify their target genes through a complementary 6-8 nucleotide sequence known as the “seed sequence”. The seed sequence may lie in the 5'-UTR, 3'-UTR or the coding region of the target gene. Several target prediction algorithms have been developed to identify putative targets of the currently known miRNAs in different species. Recently discovered functions of miRNAs include DNA repair, cell proliferation and cell death, modulation of hematopoietic lineage differentiation in mammals and others.

MiRNA biogenesis and processing are represented in Figure 7. miRNA genes are often intergenic; about 40% of miRNA genes may lie in the introns of protein and non-protein coding genes and are often transcribed together with their host gene. About 42-48% miRNAs are transcribed from a polycistronic unit encoding multiple discrete loops for

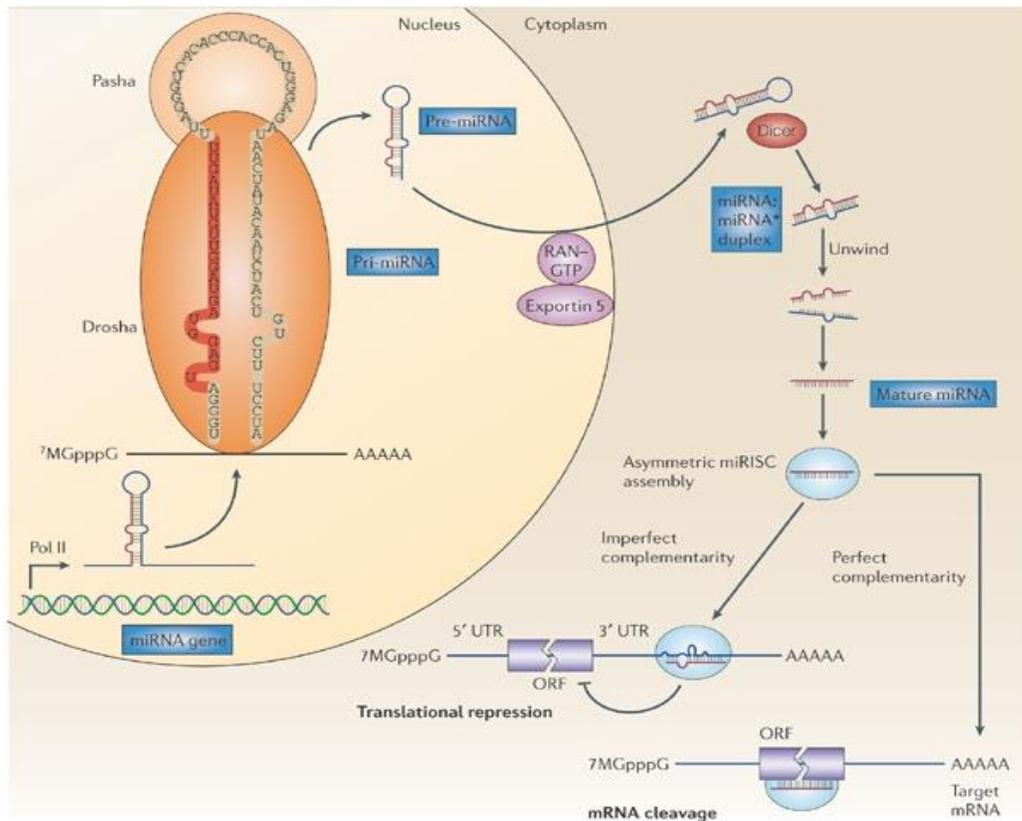


Figure 7. MicroRNA Biogenesis. MicroRNA (miRNA) genes are generally transcribed by RNA Polymerase II (Pol II) in the nucleus to form large pri-miRNA transcripts, which are capped (7MGpppG) and polyadenylated (AAAAA). These pri-miRNA transcripts are processed by the RNase III enzyme Drosha and its co-factor, Pasha, to release the ~70-nucleotide pre-miRNA precursor product. RAN-GTP and exportin 5 transport the pre-miRNA into the cytoplasm. Subsequently, another RNase III enzyme, Dicer, processes the pre-miRNA to generate a transient ~22-nucleotide miRNA:miRNA* duplex. This duplex is then loaded into the miRNA-associated multiprotein RNA-induced silencing complex (miRISC) (light blue), which includes the Argonaute proteins, and the mature single-stranded miRNA (red) is preferentially retained in this complex. The mature miRNA then binds to complementary sites in the mRNA target to negatively regulate gene expression in one of two ways: 1. miRNAs that bind to mRNA targets with imperfect complementarity block target gene expression at the level of protein translation (lower left). Complementary sites for miRNAs using this mechanism are generally found in the 3' untranslated regions (3' UTRs) of the target mRNA genes. 2. miRNAs that bind to their mRNA targets with perfect (or nearly perfect) complementarity induce target-mRNA cleavage (lower right). miRNAs using this mechanism bind to miRNA complementary sites that are generally found in the coding sequence or open reading frame (ORF) of the mRNA target. Adapted from Winter et al, 2009.

several mature miRNAs to be processed. Most miRNA genes are transcribed by RNA polymerase II and undergo processing stages from primary miRNA in the nucleus to precursor miRNA in the nucleus and precursor-miRNA to mature miRNA in the cytoplasm. This miRNA processing is mediated by the RNA-induced silencing complex (RISC) containing Drosha and many other proteins. MiRNA expression is also subject to various transcriptional regulatory mechanisms and its dysregulation has been reported to be causative for several pathophysiological conditions; cancer, heart disease, nervous disorders, obesity and also inherited disorders are some examples (Ardekani and Naeini 2010).

D. NuRD complex:

The nucleosome remodeling and histone deacetylase complex is an epigenetic regulator of gene expression. It is one of the four major types of ATP-dependent chromatin remodeling complexes. The NuRD complex was first purified over a decade ago in cells of different species and is the only multi-protein complex that couples two independent chromatin modifying enzymatic activities: 1) nucleosome remodeling brought by the chromodomain-helicase-DNA-binding protein 3 (CHD3; Mi2 α) and CHD4 (Mi2 β) subunits which have ATPase-dependent chromatin remodeling activity and which fall within the subclass of the SWI/SNF family and 2) protein deacetylation brought about by the histone deacetylases HDAC1 and HDAC2. Other non-enzymatic subunits of the complex include the retinoblastoma-binding protein 4 (RBBP4, also known as RbAp48) and RBBP7 (also known as RbAp46) which directly bind core histones; the GATAD2A (p66 α) and GATAD2B (p66 β) proteins which interacts with

MBD2/3 and also recruits Mi2 α / β proteins to the NuRD complex; both RbAp46/48 and p66 α / β proteins serve important roles as structural units of the complex too. The MTA family of proteins - MTA1/2/3 are also core components of the NuRD complex with potentially unique roles in transcriptional regulation, given cell- and tissue-type specificity (Figure 8). All of these components associate with a member of the MBD family of proteins, MBD2 or MBD3. The MBD2 and MBD3 subunits have been shown to associate with mutually exclusive NuRD complexes and have recently been implicated in fundamentally different transcriptional regulatory processes. Early studies of the MBD2 and MBD3 proteins in transgenic mice revealed that *Mbd2*^{-/-} mice were viable and fertile with a defect in maternal nurturing whereas *Mbd3*^{-/-} mice were embryonic lethal, indicating non-overlapping functions of the MBD2/3 proteins in developmental gene regulation (Hendrich et al. 2001).

In addition to these six core subunits, a large number of proteins have been shown to interact with the NuRD complex, such as GATA1/FOG1, BCL11A, LSD1 etc (Sankaran, Xu, Orkin 2010a; Xu et al. 2013). Such a combinatorial assembly of non-enzymatic subunits in the NuRD complex has been proposed to be the fundamental mechanism conferring functional specificity to the NuRD complex.

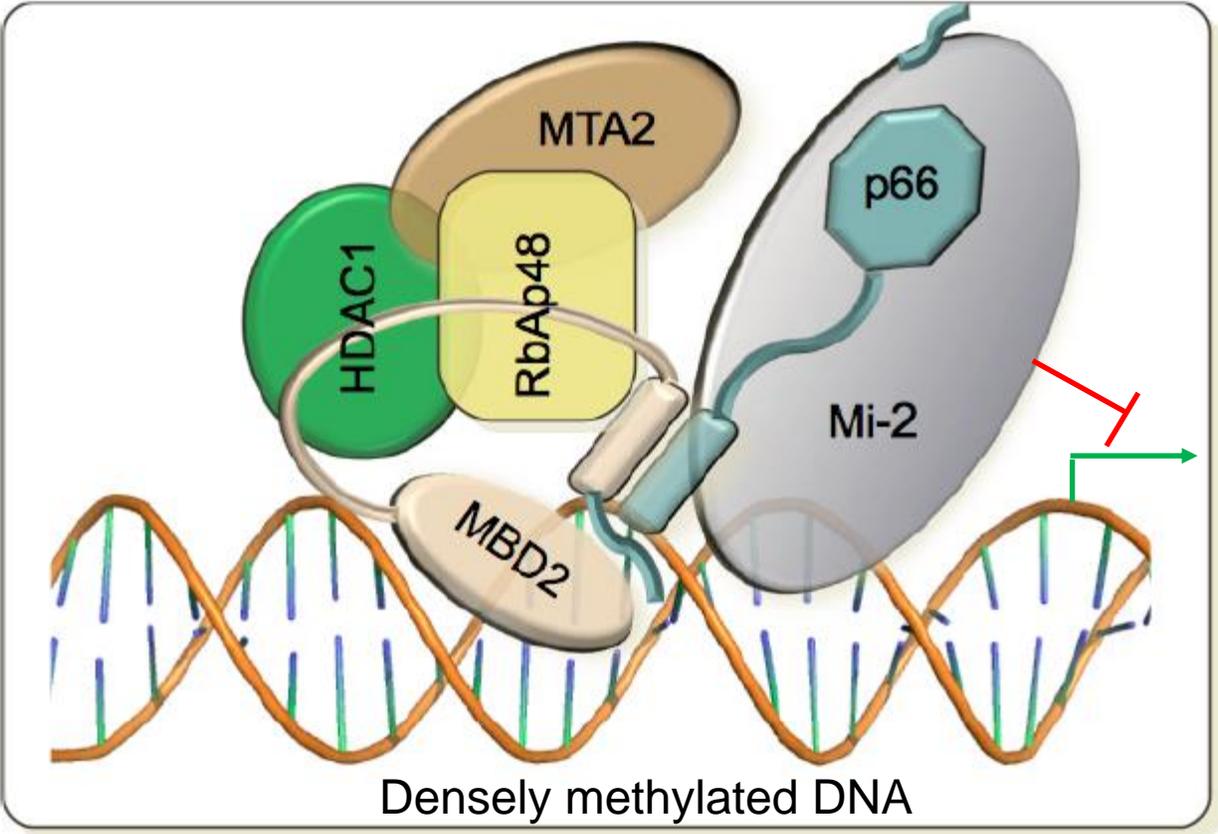


Figure 8. The MBD2/NuRD co-repressor complex. Densely methylated DNA is able to recruit methyl-binding domain protein 2 (MBD2) as well as the NuRD complex. Mi2 confers chromatin remodeling function to the complex. MTA2 and p66 act as transcriptional repressors. The histone deacetylase activity is mediated by HDAC1 and RbAp48 which is a histone-binding protein.

i. MBD2-NuRD Complex in normal development

MBD2 recruits the nucleosome remodeling and deacetylase (NuRD) complex which was first identified as a multi-protein complex known as MeCP1 (Ng et al. 1999). MeCP1 was shown to bind >12 methylated CpG dinucleotides, irrespective of the sequence context and its transcriptional repression was found to be correlated with methylation density. MBD2 was identified to be the methyl-DNA binding component of MeCP1 (Feng and Zhang 2001; Ng et al. 1999). The MBD2-NuRD complex has been isolated from cell lines and primary cells and consists of at least one component each of RbAp46/48, HDAC1/2, MTA1/2/3, p66 α/β , Mi-2 α/β and MBD2 (Feng and Zhang 2001; Kransdorf et al. 2006). MBD3 was also identified as a part of this complex initially, however later studies identified MBD2-NuRD and MBD3-NuRD as distinct complexes (Le Guezennec et al. 2006).

Studies of MBD2^{-/-} knockout mice suggested that MBD2 was not required for embryonic development since MBD2^{-/-} mice are viable, fertile and of normal appearance, although MBD2^{-/-} mothers display abnormal nurturing behavior (Hendrich et al. 2001). They do not show any defects in DNA methylation or imprinting of genes, however cell lines generated from MBD2^{-/-} knockout mice were unable to repress methylated reporter genes as effectively as wild-type and heterozygous mice.

Kersh et al. showed MBD2 plays a role in the differentiation of naive CD8⁺ T cells into effector and memory cells (Kersh 2006). MBD2^{-/-} knockout mice also display disoriented differentiation of helper T-cells which show mis-expression of interleukin 4 and interferon γ ; interleukin 4 and interferon γ are expressed mutually exclusively in Th1 and Th2 populations, however Mbd2^{-/-} mice show simultaneous expression of both cytokines in their T helper cells, which ultimately leads to altered immune responses in

the Mbd2^{-/-} mice. Gene expression analysis of differentially regulated genes in the gut of Mbd2^{-/-} mice identified abnormally high expression of exocrine pancreatic genes in the colon of these animals. The exocrine pancreatic genes code for digestive enzymes which are normally expressed in the duodenum and pancreas only (Berger et al. 2007). MBD2 deficiency also causes slightly elevated expression of Xist in SV-40 transformed fibroblast cell lines derived from the tail of knockout mice and add back of exogenous Mbd2 was able to recover the repression of the Xist gene (Barr et al. 2007). The Xist gene produces a non-translated RNA that localizes to the inactive X-chromosome from which it is transcribed and is responsible for cis-inactivation of the X-chromosome.

MBD2 has also been shown to affect the development of olfactory receptor neurons by controlling the expression of differentiation stage specific genes. Loss of Mbd2 results in the dysregulation of adult olfactory epithelium progenitor-driven neurogenesis, compromised olfactory receptor neuron survival, and prolonged deficits in the organization of olfactory receptor neurons synaptic fields (Macdonald et al. 2010).

ii. MBD2-NuRD Complex in Tumorigenesis:

Loss of MBD2 has been shown to strongly suppress adenoma formation in Apc^{min/+} mice (Sansom et al. 2003). Apc^{min/+} mice are heterozygous for a mutation in the tumor suppressor APC (adenomatous polyposis coli) which leads them to develop multiple intestinal neoplasias. MBD2 binds to and silences the expression of the Lect2 gene whose product is a negative regulator of the Wnt signaling pathway. Thus loss of MBD2 attenuates Wnt signaling and up regulates the Lect2 gene product, which in turn suppresses formation of intestinal neoplasms in Apc^{min/+} mice (Phesse et al. 2008).

MBD2 has been shown to bind to 5' upstream regulatory regions and repress the expression of tumor suppressor genes p16 and p14 in human colon cancer cell lines (Magdinier and Wolffe 2001). *p16* is a cyclin-dependent kinase inhibitor found to be hypermethylated in various malignant tissues and p14/ARF (ARF- alternative reading frame), a gene located 20kb upstream on chromosome 9p21, has also been found to be methylated in many carcinomas. Both of these gene products are critical in of major cell cycle regulatory pathways of p53 and retinoblastoma protein, Rb, respectively. MBD2 also silences the expression of *GSTP1* gene in MCF-7 breast cancer cells; GSTP1 encodes the pi-class glutathione S-transferase and is known to be hypermethylated in cancers of the breast, prostate and liver (Lin and Nelson 2003a). Studies in our lab identified a direct role of MBD2 in methylation-mediated silencing of tumor suppressor genes, DAPK1 and KLK10 in MCF-7 breast cancer cells as well as growth suppression of MBD2 deficient mammary epithelial cell lines in *in vivo* xenograft tumors in mice (Mian et al. 2011).

iii. MBD2-NuRD complex in hemoglobin switching:

Globin genes were among the first set of genes in which DNA methylation was shown to be inversely proportional to gene expression. Previous studies in our lab identified a critical role for MBD2 in the developmental regulation of the chicken embryonic ρ -globin gene and human embryonic ϵ -globin and fetal γ -globin genes. The chicken embryonic ρ -globin gene becomes highly methylated at the same time that it is transcriptionally silenced (Singal et al. 1997). Also, a protein complex containing MBD2 was purified from primary chicken erythrocytes and it was shown to bind the methylated DNA

sequences proximal to the ρ -globin gene (Singal et al. 2002). Further purification and characterization of this complex showed the presence of MBD2, but not MBD3 in the complex (Kransdorf et al. 2006). These studies were the first to identify a direct role of the MBD2/NuRD complex in developmental regulation of the β -type globin genes.

These results were followed by crossing transgenic mice bearing a yeast artificial chromosome with the entire human β -globin gene locus (β -YAC) with MBD2 knockout mice. Absence of MBD2 led to ~20 to 40% increase in the human fetal γ -globin gene expression (Rupon et al. 2006a) as well as the embryonic ϵ -globin gene (Rupon et al. 2011a) in adult mice. Given that the human β -globin gene locus is CpG sparse, we also showed that MBD2 does not directly interact with the γ -globin gene promoter, thereby suggesting that MBD2 regulates the expression of this locus via an indirect mechanism.

The role of MBD2 in developmental regulation of the γ -globin gene was also studied in primary adult human erythroid cells. Stable knockdown of MBD2 by ~75% reactivated expression of the fetal γ -globin gene to about 10% of the total β -type globin genes expressed in terminally differentiated erythroid cells (Amaya et al. 2013). These promising results reinforced MBD2/NuRD as a potential target for therapeutic reactivation of the γ -globin gene for treatment of β -hemoglobinopathies. A proof-of-concept study was conducted by our lab to examine the repressive function of MBD2 by disrupting its interaction with components of the NuRD complex. A small peptide that interfered with the coiled-coil interaction between MBD2 and p66 α dissociated the p66 α and Mi2 β proteins from the NuRD complex, resulting in increased expression of the γ -globin gene in adult mouse bone marrow erythroid cells bearing the human β -globin gene locus (Gnanapragasam et al. 2011a). These results encouraged studies to disrupt

other protein-protein interactions within the MBD2/NuRD complex with the potential of designing similar inhibitory small peptides.

VI. Primary adult human erythroid cells as a model to study hemoglobin

switching :

Cultured primary human erythroid cells derived from CD34+ progenitors serve as a powerful model to study developmental regulation of the globin genes because they recapitulate normal adult erythropoiesis at the molecular level (Fibach and Prus 2005; Pope et al. 2000). A two phase liquid culture differentiation system allows for extensive proliferation and terminal differentiation of the mobilized bone marrow cells harvested from the peripheral blood (Amaya et al. 2013). A major challenge in using hematopoietic stem cells for the study of regulators of the hemoglobin switch is their limited proliferative capacity, asynchronous differentiation manner and lack of ability to achieve terminal differentiation of these cells without significantly affecting cell viability in culture. Human induced pluripotent stem cells (hiPSCs) derived erythroid cells are also being developed which may overcome the limitations of limited life span and viability of the primary human erythroid cells and may be used for mass production of erythroid progenitor cells to study the hemoglobin switch in adult erythropoiesis. Although significant advances have been made in generating iPSCs from adult somatic cells, the system is challenged by the fact that iPSC-derived erythroid cells express mainly embryonic (ϵ) and fetal (γ) globins (Chang et al. 2006). Further studies are required to understand the transcriptional profiles of iPSC-derived hematopoietic cells and their application in studying human erythropoiesis.

VII. Scope of the thesis

The MBD2-NuRD complex was biochemically characterized in erythroid cells by our lab and found to contain RbAp46/48, HDAC1/2 and MTA1/2/3 components which form the histone deacetylase core of the complex and the p66 α / β and Mi2 α / β proteins which comprise the chromatin remodeling components of the NuRD complex, together with MBD2 which recruits the NuRD complex to sites of dense CpG methylation in the genome (Kransdorf et al. 2006). Little is known about the structural properties and functional interactions of various components within the MBD2/NuRD complex. The work presented in Chapter 2 characterizes the structure and function of an intrinsically disordered region of MBD2, MBD2_{IDR}, and identifies its important roles in modifying kinetics of DNA binding and interaction with the histone deacetylase core of the NuRD complex. A critical contact region for this interaction comprising two contiguous amino acid residues, Arg²⁸⁶ and Leu²⁸⁷ was identified and its functional significance was examined.

Studies in our lab have also demonstrated a critical role of MBD2 in developmental silencing of the fetal γ -globin gene in adult erythroid cells. However, it is not yet clear how MBD2, in concert with the NuRD complex, brings about this silencing of the fetal γ -globin gene. The MBD2/NuRD complex requires sites of dense CpG methylation for its interaction with methylated DNA and the human β -globin gene cluster lacks regions of high CpG density. In continuation of previous work in our laboratory that identified potential downstream regulators of MBD2 through a microarray screen, the studies detailed in Chapter 3 of this thesis functionally validate one of these candidate target

genes and a newly identified miRNA in the transcriptional activation of the fetal γ -globin gene in a human primary adult erythroid model system.

VIII. Significance

Currently, not many treatment options are available for patients with β -hemoglobinopathies. Blood transfusions and bone marrow transplants are recommended for a small percentage of patients because of the associated high risks and complications. Reactivation of fetal hemoglobin (HbF) in adult patients has been pursued as a therapeutic avenue because of its ameliorating effects in SCD and β -thalassemia. The DNA methylation inhibitor 5-azacytidine was tested in preclinical studies to induce HbF but was abandoned later due to concerns of potential carcinogenicity. Hydroxyurea - an S-phase specific agent without a primary hypomethylating effect and tolerable side effects, was later tested in clinical trials and presently it is the only therapeutic agent available for treatment of sickle cell anemia, but its effect varies highly in the patients.

Considering the limited efficacy of available treatment options, understanding of the molecular mechanism of hemoglobin switching will prove rewarding in the development of targeted therapies for stimulating HbF in adult patients with β -hemoglobinopathies. Previous work in the lab has identified MBD2 to be involved in developmental silencing of fetal hemoglobin. The therapeutic significance of MBD2 as a target is underscored by the observation that MBD2 has a minimal role in normal mammalian development as seen in MBD2 knockout mice which are viable, fertile and have a minimally abnormal phenotype.

Understanding the mechanistic role of MBD2 and delineating its functional interactions with the NuRD co-repressor complex will facilitate its research as an effective therapeutic target. Disrupting MBD2-mediated functions will have therapeutic potential not only in fetal hemoglobin induction in adult patients with β -hemoglobinopathies but also in the treatment of wide array cancers by de-repression of tumor suppressor genes targeted by MBD2.

CHAPTER 2: An Intrinsically Disordered Region of MBD2 recruits the Histone Deacetylase Core of the NuRD Complex

I. Introduction:

Epigenetic regulation comprises heritable changes in gene expression most commonly brought about by DNA methylation and histone modifications. The predominant form of DNA methylation in mammals involves addition of a methyl moiety to the C5 carbon of the cytosine residue in a cytosine-guanine dinucleotide (CpG) through the enzymatic activity of DNA methyl-transferases, DNMT1, DNMT3A and DNMT3B (Kim, Samaranayake, Pradhan 2009). Regions of high CpG density are often associated with gene promoters (Illingworth and Bird 2009), which remain unmethylated except for a subset of tissue-specific genes involved in normal differentiation and development (Berger et al. 2007; Kransdorf et al. 2006; Rupon et al. 2006a; Straussman et al. 2009). Aberrant hypermethylation of tumor suppressor gene promoters is associated with oncogenesis in a wide array of tissues (Jones and Baylin 2007). The methyl-CpG binding domain (MBD) family proteins recognize this methylated mark and repress the associated genes by recruiting different co-repressor complexes. The MBD family of proteins

include the first identified MeCP2 (Meehan et al. 1989) and MBD1, MBD2, MBD3 and MBD4 (Hendrich and Bird 1998). With the exception of mammalian MBD3, all MBD proteins bind to methylated DNA although with varying affinities (Cramer et al. 2014). MBD2 binds densely methylated CpG islands and represses transcription of the associated genes through recruitment of the NuRD co-repressor complex (Feng and Zhang 2001). The MBD2-NuRD co-repressor complex from both cell lines and primary cells has been characterized and is comprised of at least one copy each of the MTA1/2/3, HDAC1/2, RbAp46/48, p66 α / β , and Mi-2 α / β and MBD2 proteins (Feng and Zhang 2001; Kloet et al. 2014; Kransdorf et al. 2006; Smits et al. 2013). Recent studies have evaluated the stoichiometry of protein interactions in the NuRD complex (Kloet et al. 2014; Smits et al. 2013). However much remains to be explored about the nature and assembly of protein-protein interactions within this complex. Previous work in our laboratory identified MBD2 as a silencer of the chicken ρ -globin gene (Gnanapragasam et al. 2011b; Kransdorf et al. 2006; Singal et al. 2002) as well as murine and human embryonic and fetal β -type globin genes in adult erythroid cells (Gnanapragasam et al. 2011b; Rupon et al. 2006a). In addition, MBD2 has been implicated in aberrant silencing of methylated tumor suppressor genes in carcinogenesis (Esteller 2008; Lin and Nelson 2003b; Magdinier and Wolffe 2001; Mian et al. 2011; Sansom et al. 2003; Stefanska et al. 2013).

The MBD2 protein consists of an N-terminal glycine-arginine repeat region (GR), a methyl-binding domain (MBD) which binds *in vivo* to densely methylated DNA (Brackertz et al. 2006), an uncharacterized domain of MBD2 (MBD2_{IDR}) and a coiled-coil domain. In previous work we showed that the C-terminal coiled-coil of MBD2 binds to

the p66 α / β components of NuRD, which contributes to the recruitment of the Mi-2 α / β proteins and gene silencing. Consistent with these findings, the p66 α coiled-coil domain peptide can bind to native MBD2 in cells and relieve MBD2-mediated repression of target genes such as the embryonic and fetal β -type globin genes in adult erythroid cell culture systems (Gnanapragasam et al. 2011b). This proof-of-concept study underscored the biological significance of functional disruption of the MBD2-NuRD co-repressor complex and led us to pursue characterization of other MBD2 mediated interactions within the NuRD complex.

Intrinsically disordered proteins (IDPs) are a rapidly advancing area of research due to their importance in human biology. Although IDPs lack a stable three-dimensional structure under physiological conditions, they can serve as hubs of multi-protein interactions for diverse cellular functions like transcription regulation, chromatin remodeling and cell signaling because the intrinsic disorder permits transient, low affinity but high specificity protein-protein and nucleic acid-protein interactions (Tompa 2012). Among the MBD protein family members, the transcription repression domain of MBD1 (Hameed et al. 2014) and 60% of full-length MeCP2 protein have been identified to be intrinsically disordered (Adams et al. 2007), even in the presence of their binding partners.

Having previously determined the structures of the MBD and coiled-coil domains of MBD2, we present here the unique structural and functional features of the previously uncharacterized MBD2_{IDR} in collaboration with Dr. David Williams at UNC Chapel Hill (All biophysical studies were conducted by Dr. David Williams and his research group, whereas I was involved in biological assays presented in this work). We show that this

region is intrinsically disordered in isolation and in the context of the full-length protein bound to DNA; we identify its role in modifying kinetics and affinity of DNA-binding, and map the critical sites needed for MBD2_{IDR} to recruit the histone deacetylase core complex within the context of the intact MBD2 protein in cells. We anticipate these results will facilitate efforts for further biochemical and structural characterization of the MBD2-NuRD complex and open up avenues to target co-repressor activities of MBD2-NuRD via disruption of MBD2_{IDR}-mediated interactions with the NuRD complex.

II. Methods:

Protein expression and purification: Various MBD2 constructs were cloned into a modified pET32a vector (Cai et al. 2003) including: the MBD2_{IDR} of human (amino acids 238-356) and mouse (amino acids 241-359) MBD2; a single-chain coiled-coil domain construct (scMBD2-p66 α) comprised of the MBD2 coiled-coil domain (amino acids 361-393), a short GGSG linker, and the p66 α coiled-coil domain (amino acids 137-178); and a full-length MBD2 single-chain construct (MBD2FLsc) comprised of MBD2 (amino acids 150-393) and the scMBD2-p66 α . The resulting plasmids were transformed into the Rosetta II (DE3) (Invitrogen) *E. coli* strain, grown in either Luria Bertani medium (unlabeled) or M9 minimal media (¹³C, ¹⁵N labeled) and induced with 1mM isopropyl- β -d-thiogalactopyranoside at an A₆₀₀ ~ 0.8 for 2 or 4 hours, respectively. Bacterial pellets from 1L of growth media were lysed with 30 mL of the B-PER reagent (Thermo Scientific) and the expressed fusion protein purified by Nickel affinity chromatography. For analytical ultracentrifugation studies, the fusion protein was further purified by gel filtration chromatography (Superdex-75 26/60, GE Healthcare). For CD and NMR

studies, the MBD2_{IDR} was separated from the thioredoxin fusion domain by thrombin cleavage at room temperature overnight, and further purified by gel filtration (Superdex-75 26/60, GE Healthcare) and ion exchange (MonoS 10/100, GE Healthcare) chromatography. The final protein concentration was determined by UV absorbance at 280 nm.

Analytical ultracentrifugation: Protein was buffer exchanged in to 20mM Tris pH 8.0, 150 mM NaCl and sedimentation velocity analyzed at 40,000 rpm, 20 °C on a Beckman Optima XL-I analytical ultracentrifuge (Beckman Coulter Inc.) equipped with a four and eight-position AN-60Ti rotor. The sample partial specific volume, buffer density and viscosity were calculated with the SEDNTERP (Lebowitz, Lewis, Schuck 2002) software and the effective molecular weight determined by fitting the data to a continuous size distribution with the SEDFIT (Schuck 2000) software.

Circular dichroism: The MBD2_{IDR} was buffer exchanged into 10 mM NaPO₄, pH 6.5 at a final concentration of ~33 µg/mL protein. CD spectra were collected from 190-260 nm (0.5 nm interval, 50 nm/min, 1 cm path length, 20 °C) on a JASCO J-715 CD spectrometer (JASCO Corp). Helical content for each peptide was calculated as described previously (Kumita, Smart, Woolley 2000; Walavalkar, Gordon, Williams 2013).

Nuclear magnetic resonance: Uniform ¹³C, ¹⁵N labeled protein was buffered exchanged into 10 mM NaPO₄, pH 6.5, 0.02% sodium, 1mM dithiothreitol, and 10%

$^2\text{H}_2\text{O}$ and concentrated to 0.5-1 mM. NMR spectra were collected on a Bruker Avance III 700 MHz instrument, and data processed and analyzed with NMRPipe (Delaglio et al. 1995) and CcpNmr (Vranken et al. 2005), respectively. Standard double and triple resonance experiments (^{15}N -HSQC, HNC0, HNCACB, CBCA(CO)NH, HBHA(CO)NH, HN(CA)NNH, HNCACO, CCH-TOCSY, (H)CC(CO)NH, ^{15}N -NOESY-HSQC) were collected and the backbone and sidechain $\text{H}_{\alpha/\beta}$ resonances assigned for mouse MBD2_{IDR}. Given the sequence identity and similar ^{15}N -HSQC spectra, only HNC0, HNCACB, and HBHA(CO)NH triple resonance spectra were necessary to propagate backbone and sidechain $\text{H}_{\alpha/\beta}$ resonance assignments to human MBD2_{IDR}. Chemical shift index analyses were performed with the TALOS-N software to predict secondary structure formation and order parameters (Cornilescu, Delaglio, Bax 1999; Shen et al. 2009; Shen and Bax 2013).

Cell culture: Human embryonic kidney 293T cells were maintained in Dulbecco's modified Eagle's Medium (DMEM) containing 10% heat inactivated fetal bovine serum (Hyclone), 2mM L-glutamine and 100 U/mL penicillin and streptomycin. MDA-MB-435 breast cancer cells were maintained in DMEM supplemented with 10% heat inactivated fetal bovine serum and 100 U/mL penicillin and streptomycin. Cells were cultured at 37°C and 5% CO₂.

Co-immunoprecipitation: Various MBD2 constructs were cloned into the pCMVTag2B (Stratagene) vector in frame with an N-terminal flag-tag sequence. HEK 293T cells were transfected with the constructs (18 ug plasmid DNA) by calcium phosphate precipitation

method (Kingston, Chen, & Okayama, 2003) and harvested after 48 hours. Cells were lysed and immunoprecipitated with anti-flag M2 antibody (Sigma) and mouse IgG (Santa Cruz) controls according to the Sigma Flag-IPT kit protocol (Sigma-Aldrich, Inc., St. Louis, MO). The precipitated proteins were then analyzed for different components of the MBD2-NuRD complex by western blot using antibodies against RbAp48 (Abcam ab79416), HDAC2 (Millipore #05-814) and MTA2 (Santa Cruz sc-28731).

For full-length MBD2 pull downs, the cells were lysed in micrococcal nuclease (MNase) digestion buffer (25mM HEPES-KOH pH 7.6, 100mM NaCl, 5mM MgCl₂, 3mM CaCl₂, 10% glycerol, 0.2% NP40 and 1X EDTA-free protease inhibitor cocktail (Roche)) followed by MNase digestion using 1500 U/mL of MNase (Worthington Biochemical, Lakewood, NJ) for 2 hours on ice. Ethidium bromide was then added to the lysate at 300 ug/mL followed by a spin at 10,000xg for 15 mins at 4°C. The supernatant was then subjected to immunoprecipitation as previously described.

Site-directed mutagenesis: Mutant oligonucleotides targeting the conserved residues of the 1st and 2nd region of MBD2_{IDR} were designed using the QuikChange Primer Design Program (www.agilent.com/genomics/qcpd). Primers used for mutagenesis are listed in Table 1. Mutagenesis was carried out per manufacturer's protocol using the Quikchange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies, Inc., Santa Clara, CA). Clones were verified by sequencing and used for transfections.

Sequence Alignments: Protein-coding sequences of MBD2 for different species were obtained from the NCBI Protein database and aligned using the PRALINE software (<http://www.ibi.vu.nl/programs/pralinewww/>).

Local isoelectric point calculation: Each residue (i) in the protein was assigned a local isoelectric point (pI) calculated from the amino acid sequence (Bjellqvist et al. 1993) for residues spanning $i \pm 7$. The pI was calculated using the SeqUtils.IsoelectricPoint module in the Bio package from the Biopython Project (www.biopython.org). For those residues less than seven from the N- or C-termini, the isoelectric point was calculated from the amino acid sequence for residues spanning $i \pm n$ where n is the number of residues between i and the N- or C- terminus.

Lentiviral knockdown and expression of MBD2: The shMBD2 (5' - GGGTAAACCAGACTTGAA - 3') sequence was cloned into the pRRL.H1.shRNA vector. Full-length wild-type and mutant MBD2 gene sequences were cloned in the pLV203 vector with a flag-tag added to the C-terminus (Genecopoeia, Rockville, MD). Both full-length MBD2 genes contain a single silent mutation introduced into the MBD2 shRNA target sequence to confer shRNA resistance. The vectors were packaged into a lentivirus by calcium phosphate transfections of HEK 293T cells. MDA-MB-435 breast cancer cells were transduced with packaged virus and grown *in vitro* for 7 days post-transduction before quantitative PCR and western blot analyses. Primers used for qPCR analysis are listed in Table 2. Antibodies used for MBD2 and Flag epitope protein detection were from Santa Cruz (sc1244) and Abcam (ab1162) respectively.

Table 1. List of primers used for mutagenesis

	FORWARD	REVERSE
P244G	GAAAATTGATGCTGTTTGTCTAATTC CCAATGTTGTATTCAAGTCTGGTTTA	TAAACCAGACTTGAATACAACATTG GGAATTAGACAAACAGCATCAATTT TC
R246E	CCAGACTTGAATACAACATTGCCAAT TGAGCAAACAGCATCAATTTTCAA CAACC	GGTTGTTTGAAAATTGATGCTGTTT CTCAATTGGCAATGTTGTATTCAAGT CTGG
T248A	GACTTGAATACAACATTGCCAATTAG ACAAGCAGCATCAATTTTCAA	TTGAAAATTGATGCTGCTTGTCTAAT TGGCAATGTTGTATTCAAGTC
R246E_T2 48A	GGTTGTTTGAAAATTGATGCTGCTTG CTCAATTGGCAATGTTGTATTCAAGT CTGGTTTACC	GGTAAACCAGACTTGAATACAACAT TGCCAATTGAGCAAGCAGCATCAAT TTCAAACAACC
P255A	GACAAACAGCATCAATTTTCAAACA AGGGGTAACCAAAGTCACAAATCAT	ATGATTTGTGACTTTGGTTACCCCTT GTTTGAAAATTGATGCTGTTTGTGTC
V256A	CATCAATTTTCAAACAACCGGCAACC AAAGTCACAAATCATCC	GGATGATTTGTGACTTTGGTTGCCGG TTGTTTGAAAATTGATG
P278G	ACAACGAATGAATGAACAGGGACGT C	CTCCCAGAAAAGCTGACGTCCCTGTT
Q280A	CGAATGAATGAACAGCCACGTGCGC T	CTTCTCCCAGAAAAGCGCACGTGGC T
W283A	AGCCACGTCAGCTTTTCGCGGAGAA G	CTTGTAGCCTCTTCTCCGCGAAAAGC
R286E	TCAGCTTTTCTGGGAGAAGGAGCTAC	TGCACTAAGTCCTTGTAGCTCCTTCT
L287A	GCTTTTCTGGGAGAAGAGGGCACAA	TGATGCACTAAGTCCTTGTGCCCTCT
R286E_L2 87A	CACGTCAGCTTTTCTGGGAGAAGGA G	CATCTGATGCACTAAGTCCTTGTGCC
L290A	GGAGAAGAGGCTACAAGGAGCTAGT	TGTTACATCTGATGCACTAGCTCCTT
For full-length MBD2 mutagenesis of specific residues, following primers were used:		
R286E_L2 87A	TTT TCT GGG AGA AGG AGG CAC AAG	CTT TTC TGG GAG AAG GAG CTA CAA G
R286E	CTT TTC TGG GAG AAG GAG CTA CAA G	CTT TTC TGG GAG AAG GAG CTA CAA G

Table 2. List of primers used for qPCR in Chapter 2

	FORWARD	REVERSE	PROBE
MBD2	AAG AGC GAG TAC AGC AAG TAC GCA	TTC TGT ATC AGC AGC TCG CGA CAA	TGG AAG AAG CAC TGA TGG CAG ACA TC
GAPDH	TCG ACA GTC AGC CGC ATC TTC TTT	ACC AAA TCC GTT GAC TCC GAC CTT	AGC CAC ATC GCT CAG ACA CCA TGC
PRSS8	ACGCCTTCATAGGT GATGCT	ATCTTGATTACTC CGGTCG	

III. Results:

A. The MBD2_{IDR} region is intrinsically disordered

We cloned, expressed and purified from bacteria the MBD2_{IDR} from both the human (amino acids 238-356) and mouse (amino acids 241-359) MBD2 proteins. The 2D ¹⁵N-HSQC spectra of each (Figure 9a) show features characteristic of a disordered peptide. The amide backbone resonances are quite sharp, the ¹H(N) chemical shifts fall between 7.5-8.5 ppm, and the sidechain amide resonances are highly degenerate. In particular, the ¹H-¹⁵N_ε resonances of two Trp residues completely overlap (red dashed oval, Figure 9a), while seven Asn and ten Gln sidechain ¹⁵N-¹H₂ resonances overlap extensively (blue dashed oval, Figure 9a). These findings indicate that the MBD2_{IDR} domain remains largely disordered in solution.

We further characterized this region by CD and analytical ultracentrifugation. The CD spectra are consistent with a highly disordered peptide comprised of only 7% helix at 298 K and 12% helix at 368 K (Figure 9b) based on mean residue ellipticity (MRE) at 222nm. Thermal melting did not show a cooperative transition typical of a folded domain (blue curve, Figure 9b). Sedimentation velocity by analytical ultracentrifugation analysis is consistent with a monomeric protein at 20 μM with a tendency to aggregate at higher concentrations (Figure 9c).

B. A polypeptide linkage stabilizes the MBD2-p66α coiled-coil complex

These observations raise the possibility that the MBD2_{IDR} depends on the MBD or coiled-coil domains to adopt a regular structure. A polypeptide including the MBD, MBD2_{IDR} and coiled-coil domains of MBD2 has proven difficult to purify from bacterial expression systems. We recently designed and tested a remarkably stable and well-

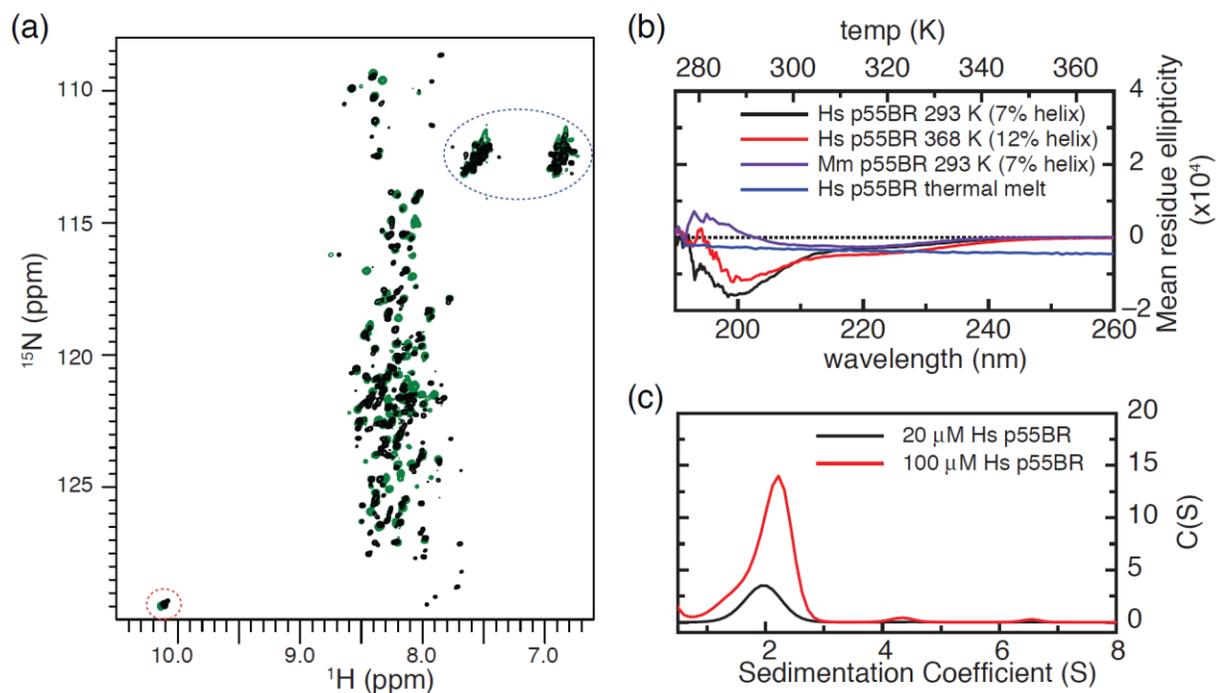


Figure 9. The MBD2_{IDR} of MBD2 is intrinsically disordered in isolation. (a) An overlay of ^{15}N -HSQC spectra for the MBD2_{IDR} of *Hs*MBD2 (black) and *Mm*MBD2 (green) shows a lack of chemical shift dispersion with highly degenerate Gln and Asn ^{15}N - $^1\text{H}_2$ (blue circle) and Trp ^{15}N - $^1\text{H}_\epsilon$ (red circle) resonances. (b) CD spectra of the MBD2_{IDR} at 293 K for *Hs*MBD2 (black) and *Mm*MBD2 (purple) and at 368 K for *Hs*MBD2 (red) shows little evidence of secondary structure formation while a thermal melt for *Hs*MBD2 MBD2_{IDR} fails to reveal a cooperative transition characteristic of folded domains. (c) Analytical ultracentrifugation sedimentation velocity analysis for 20 μM (black) and 100 μM (red) *Hs*MBD2 MBD2_{IDR} indicates that the MBD2_{IDR} is largely monomeric at 20 μM but does show evidence of aggregation at higher concentration.

(By Dr. David Williams et al.)

behaved single chain construct of the coiled-coil domains from MBD2 and p66 α (Figure 10a). This single-chain construct (scMBD2-p66 α) comprises the MBD2 coiled-coil domain (amino acids 361-393), a short GGSG linker, and the p66 α coiled-coil domain (amino acids 137-178). The scMBD2-p66 α retains helical content as measured by CD analysis up to 95°C (Figure 10b) and the 2D ^{15}N -HSQC spectrum remains dispersed and similar to that of the p66 α -MBD2 coiled-coil complex previously studied, even at 80°C (Figure 10c). Adding the scMBD2-p66 α to the MBD and MBD2_{IDR} of MBD2 (MBD2FLsc, amino acids 212-356) stabilizes the protein for bacterial expression and purification.

C. NMR analysis of intrinsically disordered MBD2_{IDR}

A 2D ^{15}N -HSQC spectrum of the MBD2FLsc protein bound to methylated DNA contains broadened and dispersed resonances traceable to the MBD and coiled-coil domains, while most resonances from the MBD2_{IDR} remain quite sharp and restricted in ^1H chemical shift similar to the isolated MBD2_{IDR} (Figure 11a). These findings show that the MBD2_{IDR} region remains largely disordered even in the context of full-length protein and while bound to DNA and does not appear to affect the structures of the MBD and coiled-coil domains. Although the backbone resonances of the MBD2_{IDR} lack significant chemical shift dispersion, we were able to assign these resonances using standard 3D heteronuclear techniques. The mouse domain was particularly well-behaved, such that we first assigned mouse MBD2_{IDR} and then carried those assignments over to the human protein where possible. Predicting the backbone order parameter and

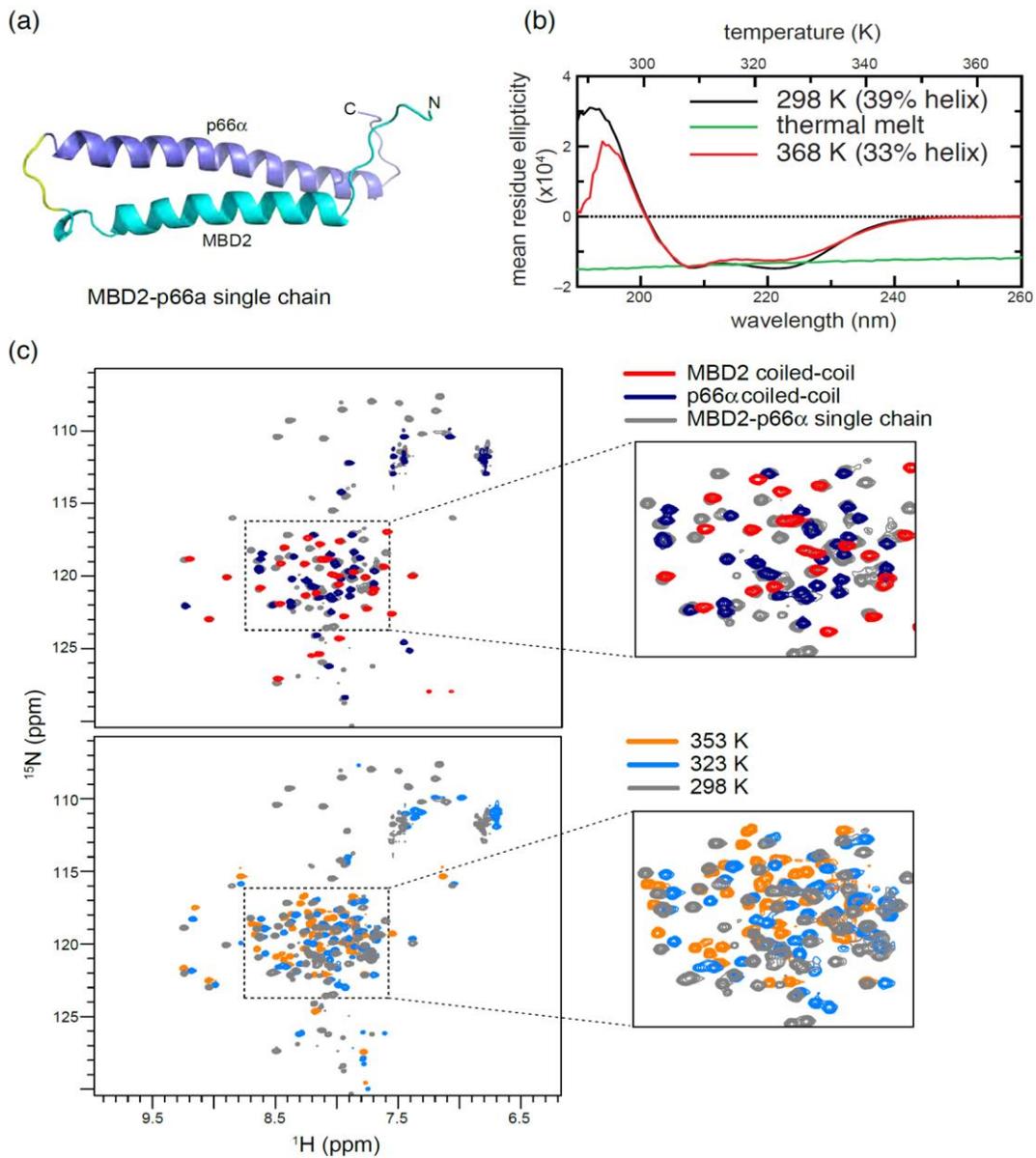


Figure 10. A polypeptide linkage stabilizes the MBD2-p66 α coiled-coil complex. (a) A ribbon diagram depicts the the scMBD2-p66 α construct comprised of the MBD2 coiled-coil domain (cyan), a short GGSG linker (yellow), and the p66 α coiled-coil domain (blue). (b) CD analysis shows that the scMBD2-p66 α contains similar helical content at 298 K and 368 K, while a thermal melt shows that the coiled-coil complex does not undergo a cooperative unfolding transition. (c) An overlay of ¹⁵N-HSQC spectra for the scMBD2-p66 α (gray) with the MBD2 (red) and p66 α (blue) coiled-coil complex peptides shows similar dispersion and chemical shifts. This finding indicates that the single-chain peptide adopts a similar structure to the coiled-coil complex of the individual peptides. (d) The ¹⁵N-HSQC spectra of the scMBD2-p66 α remains well dispersed at 298 K (gray), 323 K (blue) and 353 K (orange) showing that the scMBD2-p66 α resists thermal denaturation. (By Dr. David Williams et al.)

secondary structure by chemical shift indexing as implemented in Talos-N (Shen and Bax 2013) (Figure 11b) shows that the entire MBD2_{IDR} lacks a propensity to form structure in isolation. Several small regions (amino acids 275-280, 290-300, and 320-330) do show chemical shift perturbations suggestive of helical content (Figure 11b), but the predicted order parameters remain below 0.6 consistent with a lack of regular structure formation.

D. The MBD2_{IDR} modifies DNA binding kinetics and overall binding affinity:

To determine if the MBD2_{IDR} affected interaction with DNA, we measured binding to methylated DNA by surface plasmon resonance. As described previously (Scarsdale et al. 2011a), the isolated MBD shows rapid on and off-rates, which requires steady state analysis to determine the overall binding affinity ($K_D = 330$ nM, Figure 12a). In contrast, the MBD2^{FLsc} shows a dramatic reduction in off-rate leading to a 100-fold increase in affinity ($K_D = 2$ nM). Inspection of the MBD2_{IDR} reveals that the region just C-terminal to the MBD contains positively charged Arg and Lys residues, which could contribute to DNA binding through non-specific electrostatic interactions. Therefore we calculated a theoretical isoelectric point for a sliding window of 15 amino acids along the length of the protein. The results of this analysis for the MBD2_{IDR} homologous regions of human (*Hs*) MBD2 paralogs (MBD2, MBD3, MBD3-L1, and MBD3-L2) as well as *Bombyx mori* (*Bm*) and *Amphimedon queenslandica* (*Aq*) MBD2 orthologs are plotted in Figure 12b. Each of the MBD2 orthologs contains a positively charged region just following the MBD and at the N-terminus of the MBD2_{IDR} (Figure 12b).

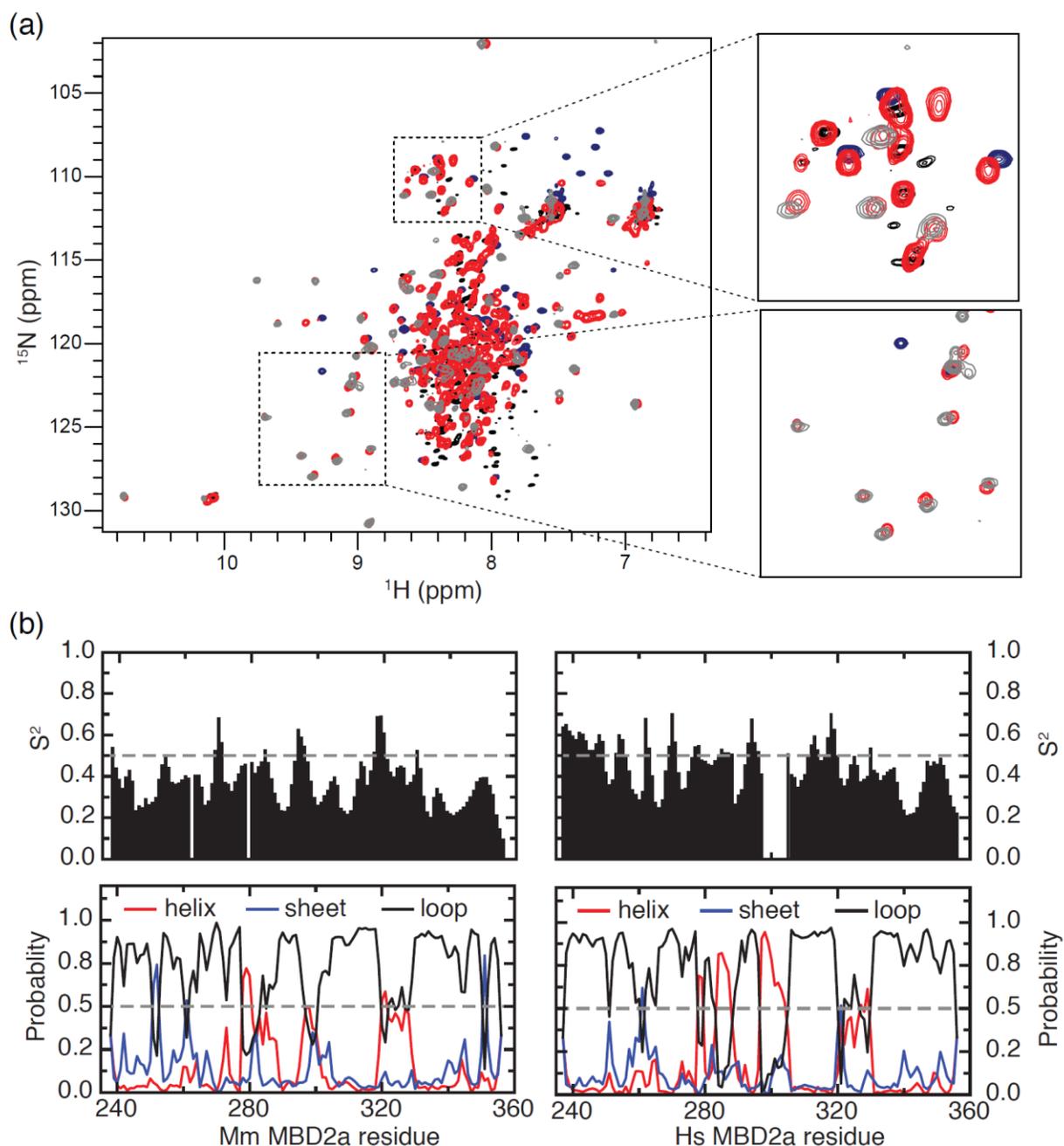


Figure 11 NMR analysis of intrinsically disordered MBD2IDR. (a) An overlay of ^{15}N -HSQC spectra for the MBD2 MBD (gray), scMBD2-p66 α (blue), and MBD2FLsc (red) indicates the MBD2_{IDR} remains disordered in the context of full-length protein and while bound to DNA. As highlighted in the expanded regions, broadened and dispersed resonances in the MBD2FLsc spectrum align with resonances from the MBD and coiled-coil domains. Resonances from the MBD2_{IDR} lack chemical shift dispersion and are highly degenerate consistent with a lack of regular structure. (b) Based on NMR chemical shift indexing(36, 37, 38), order parameters (S^2 , upper bar graphs) and secondary structure probabilities (lower graphs) for helix (red), sheet (blue) and loop (black) are plotted for the MBD2_{IDR} of *Mm*MBD2a and *Hs*MBD2a. The NMR chemical shifts indicate a lack of regular structure formation throughout the region. (By Dr. David Williams et al.)

To test whether this region is sufficient to modify DNA binding, we measured binding kinetics by surface plasmon resonance for an MBD2 construct that includes the MBD and the first 20 amino acids of the MBD2_{IDR} (residues 150-260). Similar to MBD2FLsc, this protein shows reduced on- and off-rates and increased binding affinity ($K_D = 20$ nM) compared to the isolated MBD (Figure 12a). Therefore the MBD2_{IDR} contains a positively charged region that modifies binding kinetics and affinity despite a lack of regular structure formation. Similar positively charged disordered regions adjacent to DNA binding domains have been described previously and referred to as “fuzzy” domains (Fuxreiter 2012; Fuxreiter and Tompa 2012; Uversky 2011).

E. The MBD2_{IDR} binds the histone deacetylase core complex of NuRD:

In previous studies we showed that the coiled-coil domain of MBD2 recruits the p66 and Mi2 proteins to NuRD. To investigate whether the MBD2_{IDR} and coiled-coil domains interact with the remaining core components of NuRD (MTA1/2/3, HDAC1/2, and RbAp46/48), the MBD2_{IDR} was expressed in 293T cells as a flag-tagged construct, with and without the scMBD2-p66 α coiled-coil domains (Figure 13a and 13b respectively). Cell lysates were immunoprecipitated with an anti-flag antibody followed by western blot analysis to identify the NuRD components interacting with the flag-tagged MBD2_{IDR}. Figure 13a shows that the MBD2_{IDR} of MBD2 strongly interacts with the RbAp48, HDAC2, and MTA2 components of NuRD while the scMBD2-p66 α does not. The MBD2_{IDR} does not interact with p66 α and Mi2 β proteins (Figure 14) as expected, which uniquely maps the interaction of MBD2 with the histone deacetylase and chromatin remodeling components of the NuRD complex into two distinct functional domains, the MBD2_{IDR} and the MBD2 coiled-coil domain, respectively.

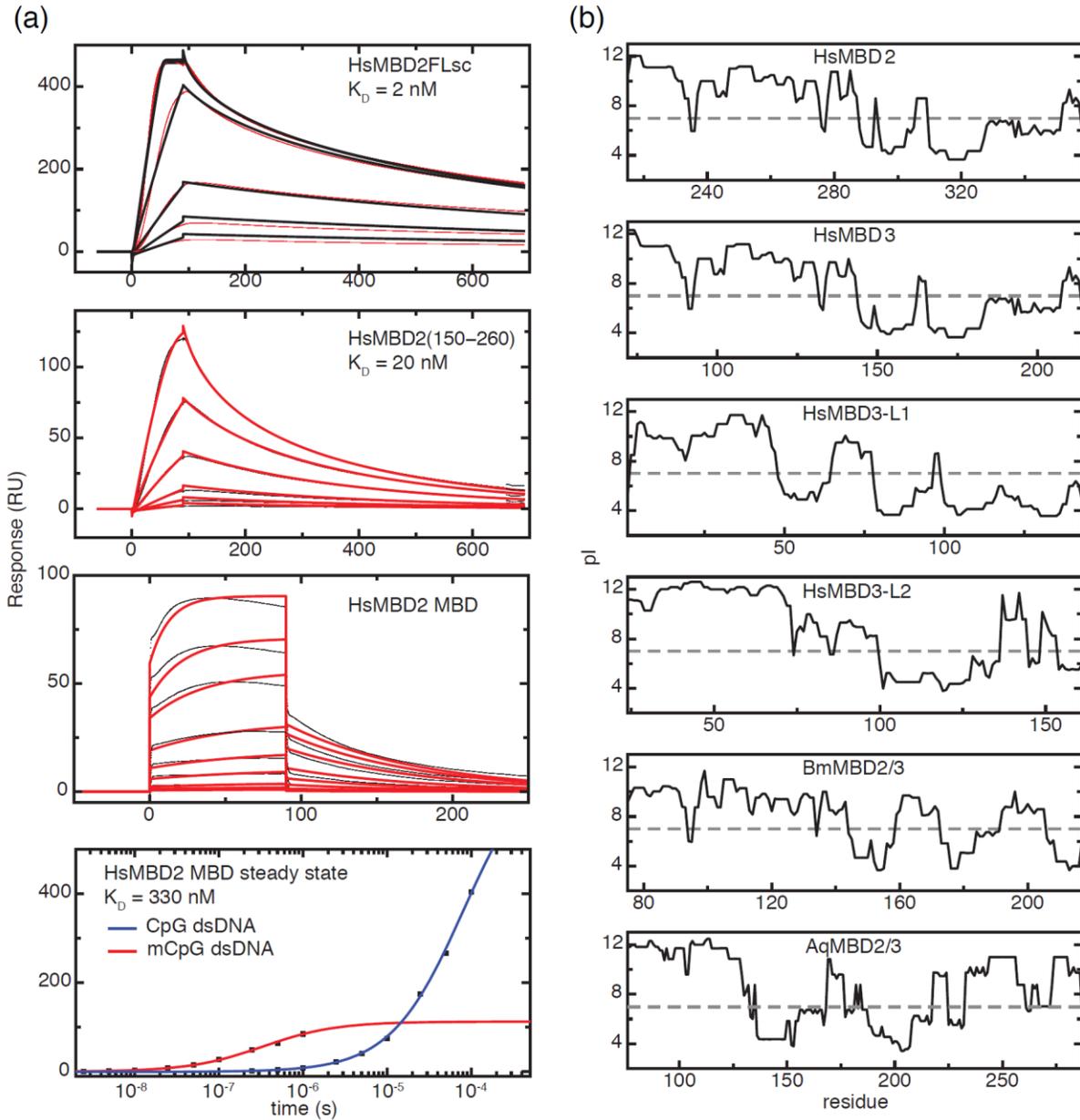


Figure 12 The MBD2IDR modifies DNA binding kinetics and affinity. (a) Surface plasmon analysis of DNA binding shows that both MBD2FLsc (upper panel), and MBD2(150-260) (second panel) bind with slower on and off rates and higher affinity than the isolated MBD (third panel). As seen previously(41), the rapid on and off-rates for the MBD preclude accurate fitting of the kinetic data. Steady-state analysis of the isolated MBD binding to methylated (red) and unmethylated (blue) DNA is shown in the fourth panel. (b) The local isoelectric point, calculated as described in the text, is plotted for the MBD2_{IDR} from *HsMBD2*, *HsMBD3*, *HsMBD3-L1*, *HsMBD3-L2*, *BmMBD2/3*, and *AqMBD2/3*. This analysis shows that a positively charged region at the N-terminus of the MBD2_{IDR} is conserved across the animal kingdom. (By Dr. David Williams et al.)

Segments of intrinsically disordered proteins can fold upon interacting with target proteins to adopt regular secondary structure. These segments, referred to as molecular recognition features (MoRFs) (Mohan et al. 2006; Uversky 2011; Vacic et al. 2007), may be identified as regions with low disorder propensity by disorder prediction algorithms. Based on analysis from the disorder prediction algorithm PONDR® VLXT (Predictor of Naturally Disordered Regions) (Li et al. 1999; Romero, Obradovic, Dunker 1997), three separate sub-regions with low disorder propensity were identified within MBD2_{IDR} (amino acids 212-273, 274-316 and 317-360, Figure 13c). We tested each of these ordered regions separately and in combination for binding to components of the histone deacetylase core complex of NuRD. As can be seen in Figure 13d, none of these three sub-regions interact with the core complex in isolation. However, the first and second sub-regions of MBD2_{IDR} in combination (amino acids 212-316) were sufficient to bind RbAp48, HDAC2, and MTA2, albeit with a somewhat weaker interaction with MTA2 as compared to the entire MBD2_{IDR} (Figure 13b).

F. Key residues within minimal MBD2_{IDR} are Evolutionarily Conserved:

Full-length protein sequences of *Hs*MBD2 orthologs from across the animal kingdom, (*Aq*, *Trichoplax adhaerans* (*Ta*), *Bm*, *Danio rerio* (*Dr*), *Gallus gallus* (*Gg*), *Mus musculus* (*Mm*)), and *Hs* paralogs MBD3, MBD3-like 1 and 2 (MBD3-L1, MBD3-L2) were obtained from the NCBI protein database. The protein sequences of individual domains of MBD2 were aligned using the default settings of the PRALINE online server. An alignment of the MBD (Figure 15a) for orthologs (*Aq*, *Bm*, *Dr*, *Gg*, and *Mm*) and a paralog (MBD3) that contain a recognizable MBD shows a high degree of conservation

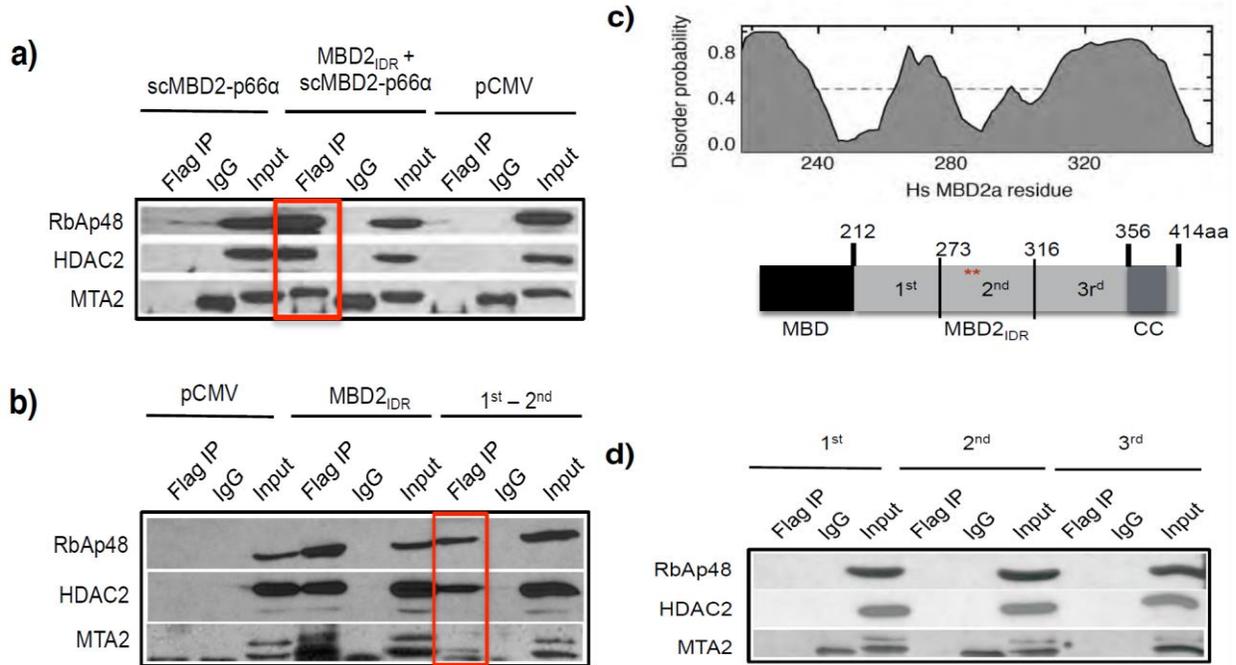


Figure 13. The MBD2_{IDR} binds the histone deacetylase core complex of NuRD. (a) The MBD2_{IDR} with scMBD2-p66α was transfected in high-transfection-efficiency HEK 293T cells. Immunoprecipitation and western blot analysis of the transfected cells indicates that the MBD2_{IDR} binds the histone deacetylase core components RbAp48, HDAC2 and MTA2, whereas the scMBD2-p66α construct does not. The flag-IP lane shows pull down of the histone deacetylase core components by immunoprecipitation using an anti-flag antibody directed against the flag-tagged MBD2_{IDR}. IgG and expression vector pCMV serve as negative controls while the input lane shows 2% of the input. (b) The region of MBD2_{IDR} from amino acids 240-316 comprising the first and second ordered sub-regions in combination can bind to RbAp48, HDAC2 and MTA2, although a weaker interaction with MTA2 was observed. (c) PONDR® VLXT analysis of the MBD2_{IDR} identified three sub-regions of low disorder propensity within this intrinsically disordered region (amino acids 240-262, 280-308 and 347-364). (d) The three ordered sub-regions of MBD2_{IDR} were expressed individually in HEK 293T cells and immunoprecipitated, but failed to bind to either of the histone deacetylase core complex components. Note that a non-specific band appears in the IgG lane when blotted with anti-MTA2 that runs just below the MTA2 protein in the input and flag IP lanes.

from sponge to human (*Aq*MBD2 and *Hs*MBD2 proteins share 64% identity in their MBDs). Critical residues known to influence DNA binding affinity and methylation selectivity (Tyr¹⁷⁸ and Lys¹⁷⁴ of *Hs*MBD2) (Cramer et al. 2014; Fraga et al. 2003; Saito and Ishikawa 2002) are highly conserved indicating that methylation specificity developed with the first multi-cellular organisms. Notably, several homologs (*Ta* MBD2/3, *Hs* MBD3-L1 and L2) lack an MBD, yet previous studies have shown that both the MBD3-L1 and L2 paralogs interact with NuRD and compete with MBD2 and MBD3 for complex formation (Jiang et al. 2002; Jiang, Jin, Pfeifer 2004; Jin et al. 2005). Therefore, NuRD complex recruitment and DNA binding represent distinct and separable functions of MBD2.

An alignment of the MBD2_{IDR} and coiled-coil regions (Figure 15b), shows much less conservation with large insertions (i.e. *Ta*MBD2/3 and *Aq*MBD2/3) and smaller deletions as compared to *Hs*MBD2. Nonetheless, specific individual residues are absolutely conserved across species suggesting that these amino acids are critical to function. Highly conserved residues were identified within the first two ordered regions of MBD2_{IDR} implicated in binding to the NuRD histone deacetylase core complex. We mutated these conserved residues individually or in pairs as follows: 1) P244G, 2) R246E, 3) P255A, 4) V256A, 5) P278G, 6) Q280A, 7) R286E, 8) L287A, 9) R246E/T248A and 10) R286E/L287A (see asterisks below sequence alignment in Figure 15b). Flag-tagged MBD2_{IDR} mutants were expressed in 293T cells followed by immunoprecipitation and blotting for the histone deacetylase core components. Of the ten conserved residues tested, only two contiguous amino acids from the second ordered region of MBD2_{IDR}, Arg²⁸⁶ and Leu²⁸⁷, when mutated effectively eliminated

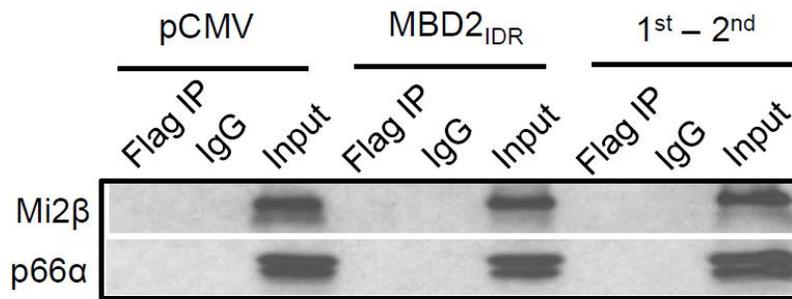


Figure 14. MBD2_{IDR} does not interact with p66α and Mi2β subunits of NuRD. Immunoprecipitation of MBD2_{IDR} and the MBD2_{IDR} minimal binding region in transfected HEK293T cells shows that the MBD2_{IDR} does not interact with p66α and Mi2β subunits of the NuRD complex, thus delineating unique points of contact associating MBD2_{IDR} with the histone deacetylase core subunits and in conjunction with our previous data, the coiled-coil domain with the chromatin remodeling component(s) of the NuRD complex

(a)

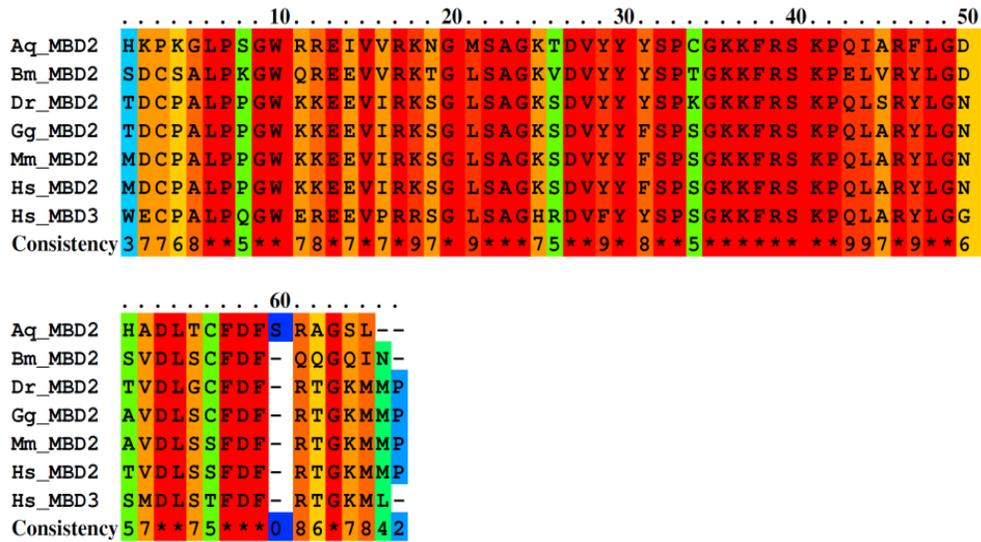


Figure 15. Evolutionary conservation of key residues within minimal MBD2_{IDR}. (a) Alignment of the protein sequence of the MBD of MBD2 for orthologs (*Aq*, *Bm*, *Dr*, *Gg* and *Mm*) and a paralog (MBD3) was done using the PRALINE program online. The software color codes the most conserved residues on a scale of 0 to 10 with 10 being the most conserved residue. Most residues within the MBD show high degree of conservation including critical residues known to influence DNA binding affinity and methylation selectivity. (b) Alignment of the protein sequence of MBD2_{IDR} and the coiled-coil domain shows much less conservation with large insertions (*Ta*MBD2/3 and *Aq*MBD2) and smaller deletions as compared to *Hs*MBD2. However, absolute conservation of specific residues in the minimal MBD2_{IDR} suggests functional importance. These conserved residues identified for functional analysis (with a conserved score of 8 out of 10) are depicted by an (*) under the sequence alignment and are as follows: Pro²⁴⁴, Arg²⁴⁶, Thr²⁴⁸, Pro²⁵⁵, Val²⁵⁶, Pro²⁷⁸, Gln²⁸⁰, Trp²⁸³, Arg²⁸⁶, Leu²⁸⁷, and Leu²⁹⁰. **(Part B on next page)**

binding to the core complex (Figure 16a). Interestingly neither of these mutations in isolation nor any of the other mutations disrupt binding to the histone deacetylase core components. Next, we introduced the double mutation R286E/L287A into the full-length human MBD2 gene by site-directed mutagenesis. Since MBD2-NuRD contains several potential DNA binding proteins, the immunoprecipitation protocol for full length MBD2Flag constructs was modified to eliminate contamination through non-specific DNA binding. The cell lysate from transfected cells was treated with micrococcal nuclease, which cleaves both single- and double-stranded DNA and RNA, and with ethidium bromide, which intercalates between stacked bases in the double helix. As expected, immunoprecipitation of the NuRD complex by the mutant full length MBD2 revealed impaired interaction with the histone deacetylase core components: RbAp48, HDAC2 and MTA2, with the effect being most pronounced for MTA2 (Figure 16b). Together these observations indicate that Arg²⁸⁶ and Leu²⁸⁷ comprise a critical interaction surface necessary but not sufficient for recruiting the histone deacetylase core complex of NuRD.

G. Functional biological assay of double mutant MBD2:

To determine if interaction with the histone deacetylase core components is necessary for methylation dependent gene silencing by MBD2, we expressed the R286E/L287A double mutant MBD2Flag construct in a background of partial knockdown of endogenous MBD2 in MDA-MB-435 breast cancer cells. The RNA expression level of the endogenous gene *PRSS8* was used as a reporter of MBD2-NuRD mediated transcriptional repression. We previously showed that stable knockdown of MBD2 by

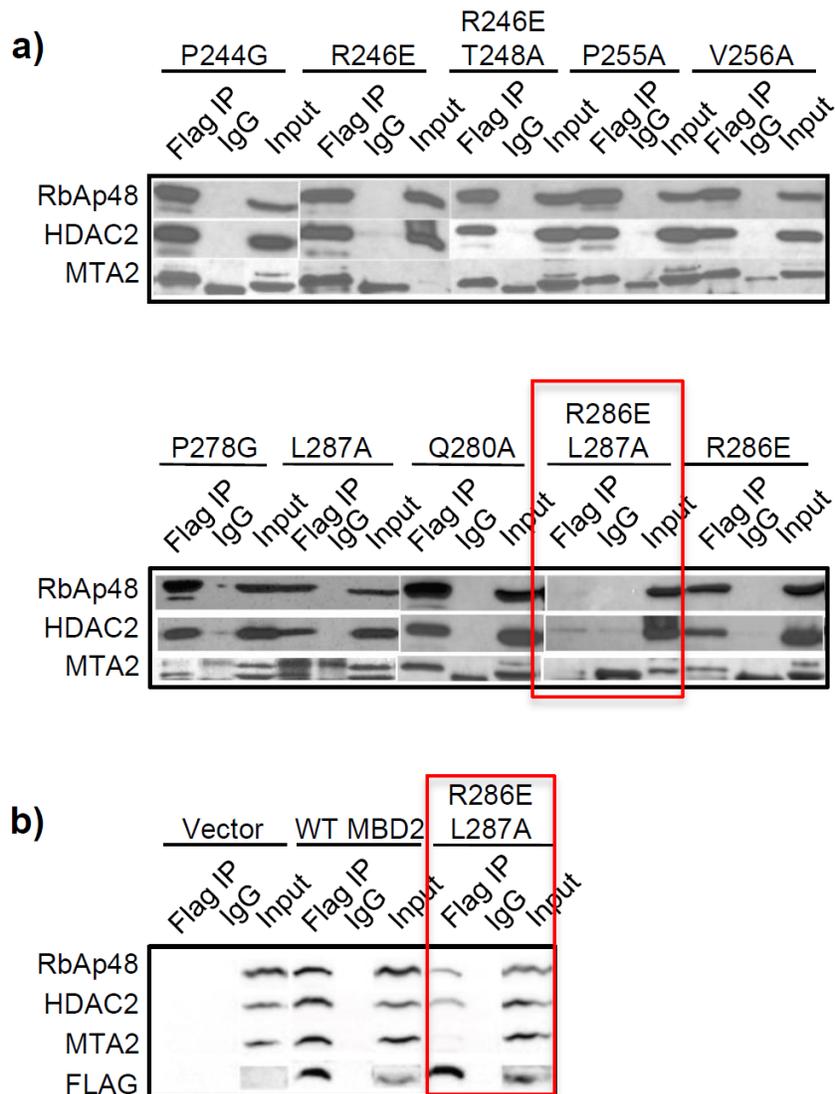


Figure 16. Two conserved residues from minimal MBD2_{IDR} are necessary to bind the histone deacetylase core of NuRD. (a) Mutating five residues from the first region of minimal MBD2_{IDR} (P244G, R246E, P255A, V256A, and R246E/T248A) does not affect immunoprecipitation of the histone deacetylase core complex. Similarly, mutating four residues from the second ordered region of minimal MBD2_{IDR} (P278G, Q280A, R286E, and L287A) does not affect immunoprecipitation of the histone deacetylase core complex. However, combined mutation of two adjacent residues (R286E/L287A) significantly abrogates the ability of MBD2_{IDR} to recruit the histone deacetylase core components. Strikingly, the individual mutants R286E and L287A have no effect on interaction of the MBD2_{IDR} with the histone deacetylase core. Three additional mutations (T248A, W283A and L290A) do not affect binding of MBD2_{IDR} to the histone deacetylase core complex (data not shown). (b) Full length flag-tagged MBD2 carrying the double mutation R286E/L287A also displays a disrupted interaction with RbAp48, HDAC2 and MTA2. The interaction of the double mutant MBD2 with MTA2 is completely lost whereas its interaction with RbAp/HDAC shows significant reduction.

shRNA in MDA-MB-435 cells increased *PRSS8* expression, demonstrating that MBD2 mediates transcriptional repression of *PRSS8* in these cells (Mian et al. 2011). Simultaneous knockdown of endogenous MBD2 and enforced expression of wild-type or R286E/L287A double mutant MBD2Flag was achieved by lentivirus mediated delivery of anti-MBD2 shRNA and wild-type and R286E/L287A double mutant MBD2Flag mRNA resistant to knockdown (Mian et al. 2011). Blotting with anti-MBD2 antibody (Figure 17a), which recognizes both endogenous (lower band) and lentivirally expressed MBD2Flag (top band), shows about 50% knockdown of the endogenous protein in SH+WT and SH+Double mutant cells as compared to the negative control (SC + LV). Detection of the exogenously expressed MBD2Flag proteins with an anti-flag antibody demonstrates similar expression levels in SH+WT and SH+Double mutant transduced cells (Figure 17b). Taking into account both knockdown of endogenous MBD2 and expression of MBD2Flag proteins, the total amount of MBD2 protein in both groups (SH+WT and SH+Double mutant) remains constant and represents ~50% more MBD2 protein than that in the negative control. Knockdown of endogenous MBD2 and add back of wild type MBD2, shows repression of *PRSS8* expression beyond the baseline due to increased total expression of wild-type MBD2. However, in the case of enforced expression of the R286E/L287A double mutant MBD2, *PRSS8* gene expression is not statistically significantly reduced compared to the baseline level although there is a suggestion of a slight decrease (Figure 17c). *PRSS8* gene expression is consistently ≥ 2 fold lower (p-value<0.01) in SH+WT cells than in SH+Double mutant cells, which we attribute to the inability of mutant MBD2 to effectively recruit the histone deacetylase core components of the NuRD complex.

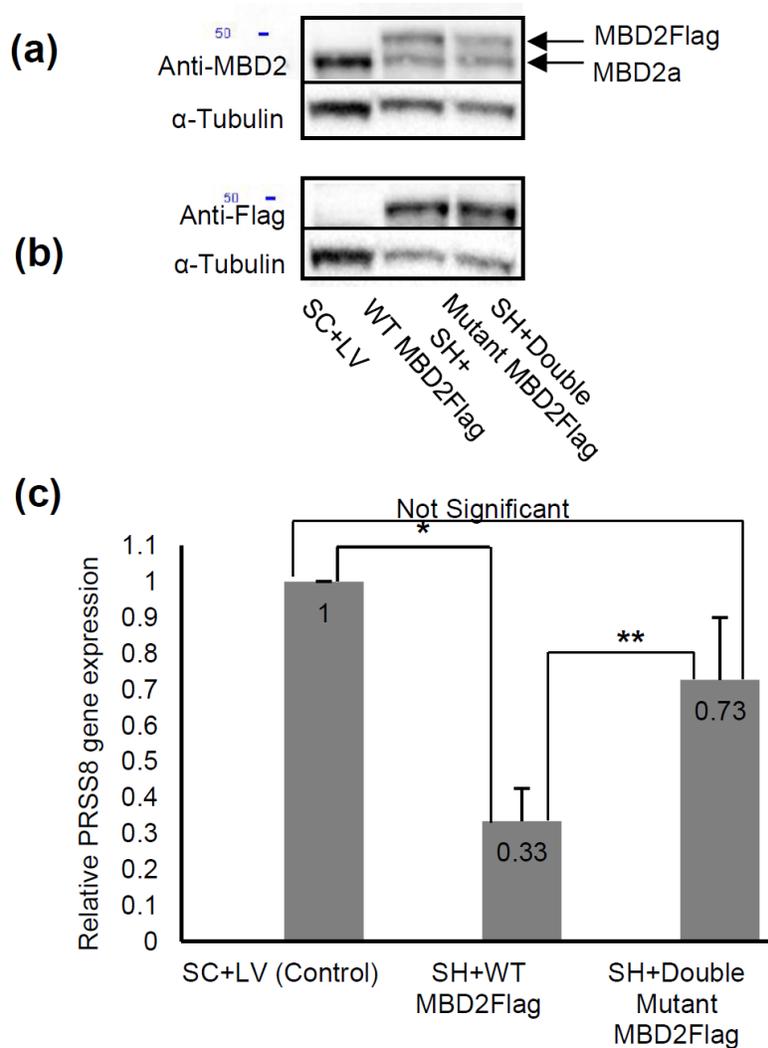


Figure 17. Double mutant MBD2 shows reduced transcriptional repression of its methylated target gene PRSS8. Western blot analysis of MDA-MB-435 cells infected with MBD2 knockdown and expression lentiviruses shows (a) ~50% and ~40% knockdown of endogenous MBD2 in SH+WT and SH+Double mutant cells respectively and (b) Equivalent expression of wild-type MBD2Flag and R286E/L287A double mutant MBD2Flag across both SH+WT and SH+Double mutant cells respectively. Therefore, SH+Double mutant cells contain 60% mutant MBD2Flag as compared to SH+WT cells. (c) SH+WT cells show increased repression of *PRSS8* as compared to the baseline expression level set by SC+LV control. However, SH+Double mutant cells express *PRSS8* at levels close to the baseline, due to the inability of double mutant MBD2 to effectively repress its target gene *PRSS8*. SC+LV control represents cells transduced with empty backbone lentiviral vectors for both shRNA knockdown (SC) and gene expression (LV). *PRSS8* gene expression is normalized to GAPDH expression. Error bars denote standard deviation of 3 independent experiments. *signifies p-value<0.005, ** signifies p-value<0.01.

IV. Discussion:

Research over the past two decades has challenged the traditional structure-function paradigm of proteins. Instead of adopting a well-folded three-dimensional structure in solution, many proteins are entirely disordered or contain functionally important intrinsically disordered regions (IDR). Estimates indicate that from 15-45% of eukaryotic proteins have long (>30 amino acid) regions of disorder (Tompa 2012). These regions can function either by folding upon binding to a target ligand or through dynamic interactions involving rapidly exchanging conformations of the disordered state (Fuxreiter and Tompa 2012; Fuxreiter 2012; Tompa 2012; Uversky 2011).

Here we show that a large segment (~145 amino acids) between the conserved MBD and coiled-coil domains of MBD2 behaves as an intrinsically disordered region in isolation and in the context of the full-length protein. This region does not adopt a regular structure even when bound to methylated DNA. Nonetheless incorporating the MBD2IDR affects the binding kinetics and overall affinity for DNA, functioning as a "fuzzy" domain (Fuxreiter 2012). The N-terminal portion of the MBD2IDR has a net positive charge that is conserved across evolution and likely contributes to this fuzzy interaction with the DNA (see Figure 12b). IDRs flanking functional DNA-binding domains have been shown to influence DNA recognition and potentially rates of association and lifetime of the DNA-protein interactions (Ghosh et al. 2010; Pontius 1993).

IDRs are often found in proteins that serve as hubs of one-to-many protein interactions (Dunker et al. 2005; Haynes et al. 2006; Patil and Nakamura 2006). In this report we demonstrate that the MBD2_{IDR} of MBD2 binds the histone deacetylase core complex of

NuRD comprised of the MTA1/2, HDAC1/2, and RbAp46/48 proteins. To map this interaction, we divided the MBD2IDR into three potential MoRFs based on disorder propensity as predicted by the PONDR® VLXT algorithm (Li et al. 1999; Romero, Obradovic, Dunker 1997). We found that the first two segments of the MBD2_{IDR} are necessary and sufficient to bind the histone deacetylase core complex (Figure 13b). Interestingly, this region also incorporates the positively charged segment that contributes to DNA binding. We were unable to identify sub-regions of MBD2IDR that selectively bind to different components of the histone deacetylase core complex, which suggests cooperative all-or-none formation of this complex (Figure 13c). This observation indicates that either the MBD2IDR forms a large contact surface for binding each of the proteins cooperatively or that the MBD2IDR binds to one of the core components which in turn interacts with other components. In support of the latter interpretation, recent structural analyses have shown that the ELM2 and SANT domains of MTA1 bind directly to HDAC1 (Millard et al., 2013) and a small peptide from the C-terminal region of MTA1 binds RbAp48 (Alqarni et al. 2014). The MTA proteins may provide a common point of contact such that MBD2IDR binding directly to either of RbAp, HDAC, or MTA would be sufficient to recruit all of the histone deacetylase core components.

Intrinsic disorder in MBD2 appears to be under selective pressure as evidenced by conservation of numerous residues within the MBD2IDR across several paralogs and distantly related orthologs, despite large insertions and sequence variability (Figure 15). Using this evolutionary conservation as a guide, we designed point mutations in the first and second segments of the MBD2_{IDR} and found that two highly conserved residues

(Arg²⁸⁶ and Leu²⁸⁷) are critical for binding to the histone deacetylase core complex. Mutating these critical residues even in the context of full length MBD2 diminishes interaction with the histone deacetylase core complex. A recent analysis of MBD2 alternative splice variants revealed that dominant isoform in embryonic stem cells, MBD2c, lacks most of the C-terminus including the MBD2_{IDR} and the coiled-coil domains. Consistent with the role of these two domains in recruiting Mi2 and the histone deacetylase core complex, MBD2c fails to interact with any of the NuRD components and lacks repressive activity (Lu et al. 2014).

MBD2-NuRD has been established as a transcriptional repressor of embryonic and fetal β -type globin genes in primary adult erythroid cells across different species (Gnanapragasam et al. 2011b; Kransdorf et al. 2006; Rupon et al. 2006a; Singal et al. 2002) and of tumor suppressor genes in breast cancer cells (Esteller 2008; Mian et al. 2011). We hypothesized that the Arg²⁸⁶ and Leu²⁸⁷ mutations of MBD2 that impair recruitment of the histone deacetylase core components of the NuRD complex would in turn abrogate repression of methylated target genes. To test this hypothesis, we assayed the expression of a previously identified highly methylated MBD2 target gene, PRSS8, in MDA-MB-435 breast cancer cells (Mian et al. 2011). PRSS8 encodes for prostaticin, a member of the trypsin family of serine proteases that has been implicated in inhibition of metastasis of both breast and prostate cancer cells (Chen et al. 2001; Chen and Chai 2002). A CpG-rich region in the promoter and exon 1 of PRSS8 is heavily methylated when this gene is repressed in MDA-MB-435 breast cancer cells (Chen and Chai 2002), while demethylation by 5-aza-2'-deoxycytidine combined with histone deacetylase inhibitor treatment reactivates expression of PRSS8. In our studies,

R286E/L287A double mutant MBD2 protein exerts diminished repression of PRSS8 when compared to an equivalent amount of wild-type MBD2 protein, (Figure 17c) thus confirming the importance of these two amino acid residues in MBD2-NuRD function. Our results can be summarized into a dynamic working model for the architecture of the NuRD complex wherein the subunits carrying the chromatin remodeling enzymatic activities unique to the NuRD complex can be mapped to two separate domains of MBD2: the MBD2_{IDR} which recruits the histone deacetylase core components and the coiled-coil domain which recruits the chromatin remodeling subunit Mi2 β through its interaction with p66 α (Figure 18). Delineating the multi-protein interactions within the MBD2-NuRD complex should enable future attempts to decouple the distinct enzymatic activities of the complex, in isolation or combination, to disrupt the repressive functions of each and dissect their potentially independent roles in regulating specific sets of genes.

Understanding the role of IDRs in the assembly of multi-protein complexes like MBD2-NuRD adds valuable insight to their function in epigenetic regulation. IDRs have been well-recognized as hubs for interaction of many proteins due to their malleability and fluctuating structures. MBD1-c was recently shown to selectively interact with different binding partners through its intrinsically disordered transcriptional repression domain (Hameed et al. 2014). Structural characterization of intrinsically disordered regions of other proteins such as p53 (Dawson et al. 2003; Wells et al. 2008), PTP1B (Krishnan et al. 2014), androgen receptor (Myung et al. 2013), α -synuclein (Toth et al. 2014; Uversky, Li, Fink 2001) and oncoprotein c-myc (Berg et al. 2002; Hammoudeh et al. 2009) has elucidated the function of these proteins in pathological pathways and lead to

successful efforts developing small molecule inhibitors for therapeutic drug targeting. The work presented here sets the stage for further biochemical and functional characterization of the MBD2-NuRD complex and raises the possibility of selectively abrogating NuRD dependent silencing of specific sets of genes by disrupting the MBD2_{IDR} mediated interactions with the HDAC complex. Exploitation of such mechanistic findings could ultimately lead to improved treatment of hemoglobin disorders and cancer.

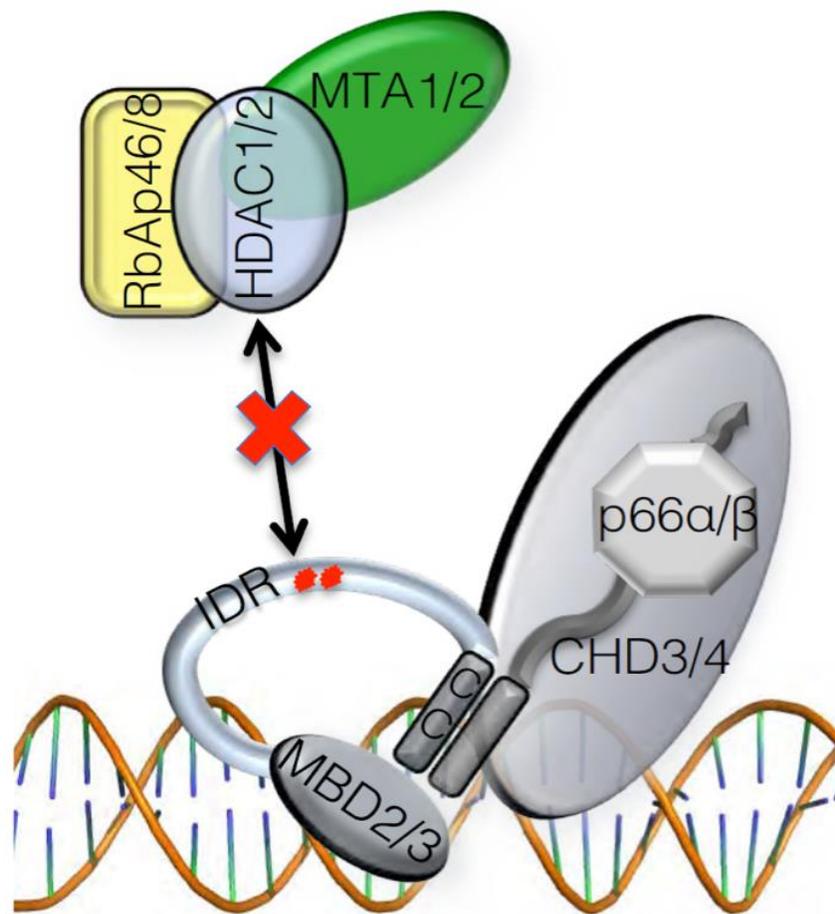


Figure 18. A working model of the architecture of the MBD2-NuRD co-repressor complex. Our model maps interactions between MBD2 and the histone deacetylase core subunits to the MBD2_{IDR}, wherein mutation of the two contiguous residues comprising the critical interaction site (shown as red asterisks in the IDR) impairs binding. The previously characterized coiled-coil domain of MBD2 recruits the chromatin remodeler Mi-2 β through p66 α and the MBD recruits the co-repressor complex to sites of dense CpG methylation.

CHAPTER 3: Role of the MBD2-NuRD complex in developmental regulation of the fetal γ -globin gene

I. Introduction:

The γ -globin gene lies on chromosome 11 within the β -globin gene cluster, which contains the genes 5'-LCR- ϵ -G γ -A γ - δ - β -3'. These β -type globin genes are expressed in ontogeny in the order in that they appear on the chromosome. Several cis-acting sequences and trans-acting factors have been identified to regulate the fetal-to-adult hemoglobin switch (Sankaran, Xu, Orkin 2010b). Our lab previously identified MBD2, an epigenetic regulator, to be involved in developmental silencing of the fetal γ -globin gene in adult erythroid cells (Amaya et al. 2013; Gnanapragasam et al. 2011b; Rupon et al. 2006a). Methyl CpG binding domain protein 2 (MBD2) belongs to the family of proteins characterized by the presence of the MBD domain. This domain enables these proteins to preferentially bind to methylated DNA. MBD2 binds to densely methylated CpG islands and recruits the Nucleosomal Remodeling and Histone Deacetylase (NuRD) co-repressor complex, to cause chromatin compaction and transcriptional silencing (Meehan et al. 1989; Wade 2001).

We previously demonstrated that MBD2 knockout in β -YAC transgenic mice leads to at least a ten-fold induction of human fetal γ -globin in adult erythroid cells (Rupon et al.

2006a). This level of induction is very similar to that which is obtained with the treatment of a DNA methylation inhibitor, 5-Azacytidine. During the switch to definitive erythropoiesis in the β -YAC transgenic mice embryos, the absence of MBD2 causes an inability to fully silence the fetal γ -globin. Knockdown of MBD2 in primary human adult erythroid cells caused significantly increased expression of the fetal γ -globin gene by ~9-fold as compared to the scramble shRNA control cells, without affecting their normal terminal differentiation into red cells. Although MBD2 was shown to have a role in the silencing of the γ -globin gene, it did not bind to the gene to mediate its silencing (Rupon et al. 2006a). This was not entirely surprising because there are no CpG islands within 6 kb of the γ -globin gene. In fact, no region in the promoter contains more than 4 CpGs in a span of 70 bp DNA. While MBD2 has been shown to bind to a sequence containing as few as three CpGs *in vitro* (Fraga et al. 2003), MeCP1 complex as a whole seems to require at least 15 CpGs per the complex *in vivo* (Meehan et al. 1989; Wade 2001). Moreover, *in vivo*, MBD2 binds only to methylated CpG island sequences (Ginder, Gnanapragasam, Mian 2008).

Because MBD2 does not bind to the γ -globin gene to cause its silencing in adult erythrocytes, we postulated that the loss of MBD2 results in transcriptional activation of a gene or genes that are normally silent in adult erythroid cells. The product of this gene(s) could in turn result in the transcriptional activation of the γ -globin gene. To recognize indirect downstream effectors of MBD2, we conducted gene and microRNA expression microarrays in MBD2^{-/-} and wild-type mice adult erythroblasts to identify genes/miRNAs that would be differentially regulated in the absence of MBD2.

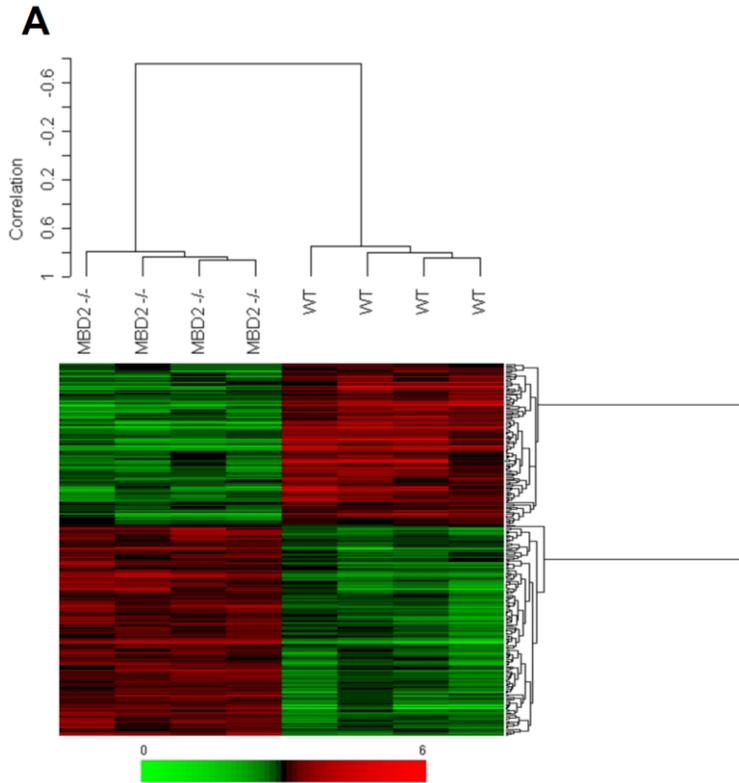
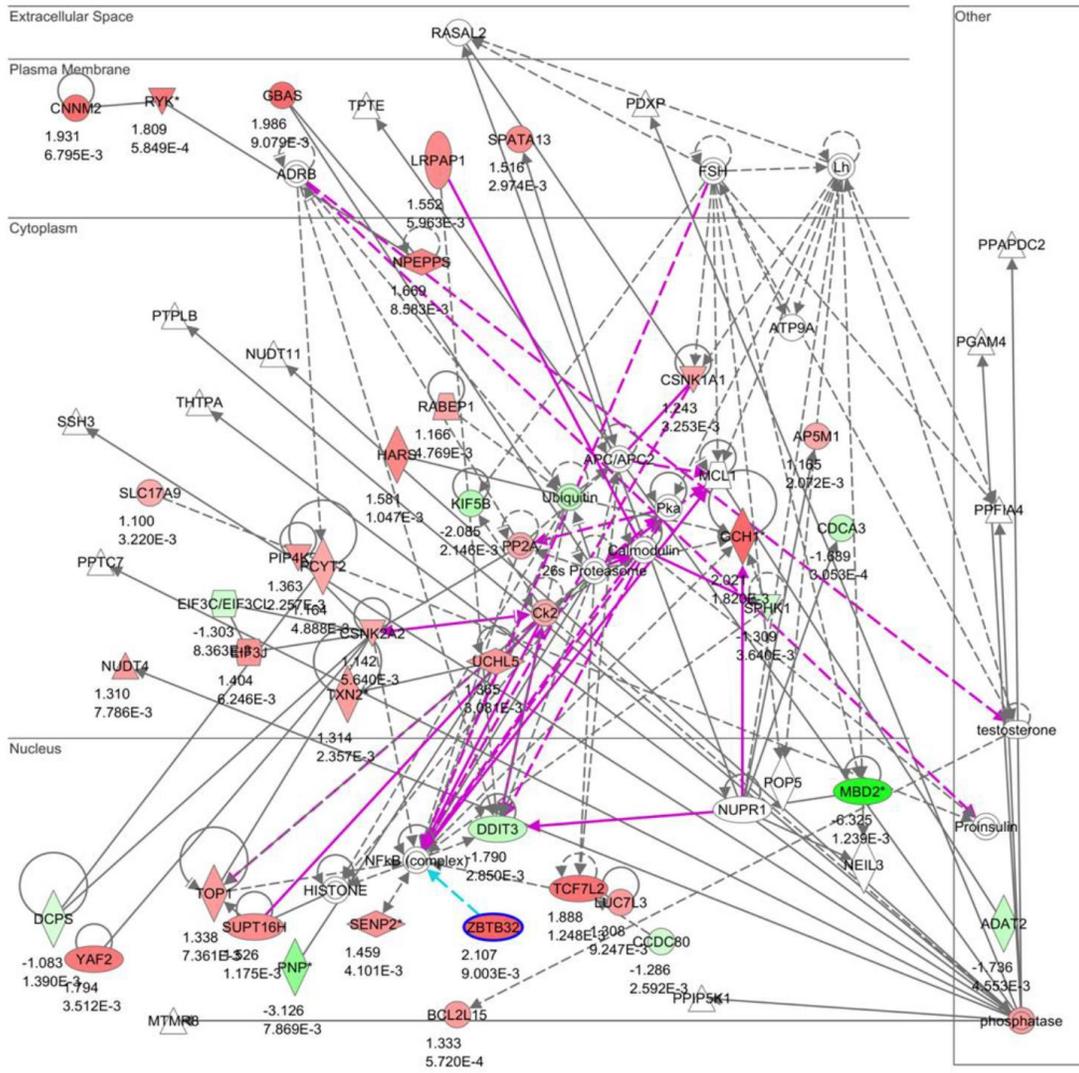


Figure 19. Gene expression microarray in erythroid cells of MBD2^{-/-} mice (A) Heat map showing differential expression of genes in erythroid cells of MBD2^{-/-} and wild-type mice. 93 were found to be up-regulated and 50 genes were down-regulated in MBD2^{-/-} mice compared to the control wild-type mice. (B) Ingenuity Pathway Analysis of the 93 differentially regulated genes to identify genes involved in hematological pathways yielded the lists of genes listed here. Prioritization based on genes having a CpG island identified these four genes as candidates: Zinc-finger and BTB domain containing 32 (Zbtb32), YY1-associated factor 2 (YAF2), BRG1-associated factor 57 (Baf57) and Itchy E3 ubiquitin protein ligase homolog (ITCH). (Part B on next page)
(Unpublished data by Rupon JW and Gnanapragasam MN)

B

The gene expression microarray was performed on an Affymetrix Genechip 430A 2.0 to recognize differentially regulated genes in MBD2^{-/-} and wild type adult erythroid cells. To this end, MBD2^{-/-} and wild type mice were intraperitoneally injected with 1-acetyl-2-phenylhydrazine to induce hemolytic anemia, causing extra-medullary hematopoiesis. This makes the spleen the primary site of erythropoiesis and it populates to have >90% erythroid cells. Total RNA was isolated from the spleens and sent for microarray analysis on an Affymetrix 430A2.0 platform with probe-sets for over 14,000 well-characterized mouse genes. To generate the list of genes which are differentially expressed between MBD2^{-/-} and wild type adult mouse erythroid cells, a statistical cut off p-value of ≤ 0.01 was used. 93 genes were found to be up-regulated and 50 genes were down-regulated in MBD2^{-/-} mice compared to wild type mice adult erythroid cells (Figure 19A). For the statistical analysis, we collaborated with Dr. Kellie J. Archer (Department of Biostatistics, VCU). Since MBD2 serves as an epigenetic repressor of its target genes, we focused on identifying genes with increased expression in MBD2^{-/-} erythroid cells. The 93 up-regulated genes were prioritized as candidate genes by Ingenuity Pathway Analysis of genes relevant in hematological pathways and we focussed on those genes with a CpG island. Ingenuity Pathway Analysis dynamically computes a large “global” molecular network based on published physical interactions (direct as well as indirect) as well as functional interactions between orthologous genes using an extensive database. From the 93 genes studied, the following four hits populated our analyses: Zinc-finger and BTB domain containing 32 (Zbtb32), YY1-associated factor 2 (YAF2), BRG1-associated factor 57 (Baf57) and Itchy E3 ubiquitin protein ligase homolog (ITCH) (Figure 19B).

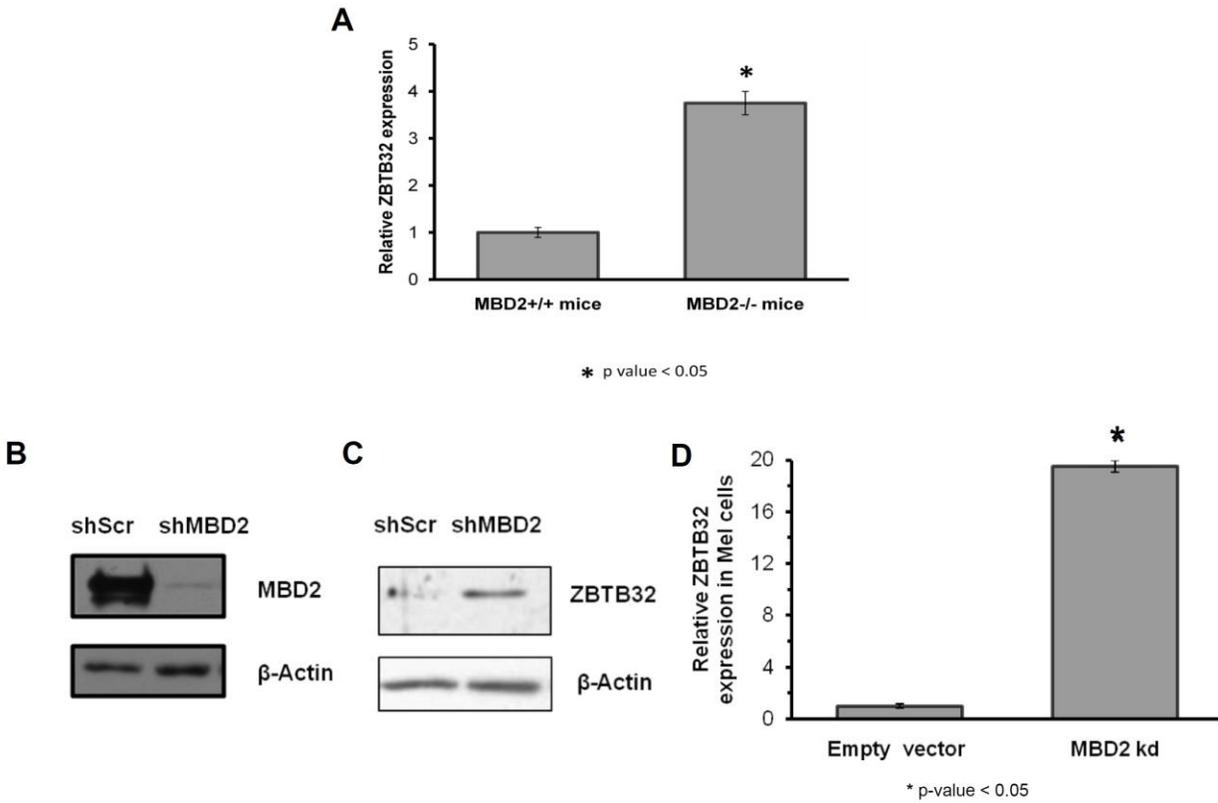


Figure 20. Validation of candidate genes by Q-RT PCR. (A) RNA from spleen of three anemic MBD2 knock down (MBD2^{-/-}) and wild type (MBD2^{+/+}) mice was analyzed by Q-RT PCR to verify the differential expression of Zbtb32; n=3, Error bars indicate mean \pm SD. Mouse erythroleukemia cell line was transfected with pSuperior-MBD2 shRNA and empty vector control to validate expression of the candidate gene Zbtb32. Western analysis was performed to determine of (B) MBD2 knockdown levels; (C) ZBTB32 protein levels in the MBD2 knockdown cells. Q-RT PCR was performed to determine (D) Zbtb32 mRNA levels in the MBD2 knockdown cells. CyclophilinA was used for normalization in Q-RT PCR. Error bars indicate mean \pm SE. (Unpublished data by Gnanapragasam MN)

Of the 4 genes mentioned above, *Zbtb32* was the only gene successfully validated by qPCR, with a 4-fold increased expression in adult erythrocytes of *MBD2*^{-/-} mice as compared to wild type mice (Figure 20A). For in vitro validation of the candidate gene *Zbtb32*, we used the mouse erythroleukemia (MEL) cell line. *Mbd2* was stably knocked down using the shRNA system in MEL cells by upto 90%, both at the mRNA and protein level (Figure 20B). In the absence of *Mbd2*, gene expression of *Zbtb32* protein was increased by ~2.5 fold as compared to the shScramble control cells (Figure 20C).

ZBTB32 contains a CpG island in the body of the gene which could potentially be bound by MBD2 when methylated. Studies have identified the regions to be preferentially occupied by MBD2 which include -700 to +200 base-pairs relative to a transcription start site (TSS) when regulating the expression of the target gene through binding methylated DNA (Chatagnon et al. 2011). We investigated whether MBD2 binds this CpG island through chromatin immunoprecipitation (ChIP). ChIP assays were performed on adult erythroid cells from spleen of anemic *MBD2*^{-/-} and wild-type mice. MBD2 showed enrichment at the CpG islands of the *Zbtb32* and *Ugt8* genes in wild-type mice but not in *MBD2*^{-/-} mice (Figure 21). The CpG island in the promoter of the *Ugt8* gene serves as a positive control for the binding of MBD2. This observation demonstrates that MBD2 binds directly to the *Zbtb32* gene in adult erythrocytes. To study the biological functional effect of increased expression of ZBTB32 as observed in cells deficient for MBD2, our experimental strategy involved enforced expression of ZBTB32 in adult hematopoietic cells. We hypothesized that enforced expression of ZBTB32 would reactivate the silenced γ -globin gene. Chemical inducer of dimerization (CID) dependent multipotential β -YAC mouse bone marrow cells were used as a

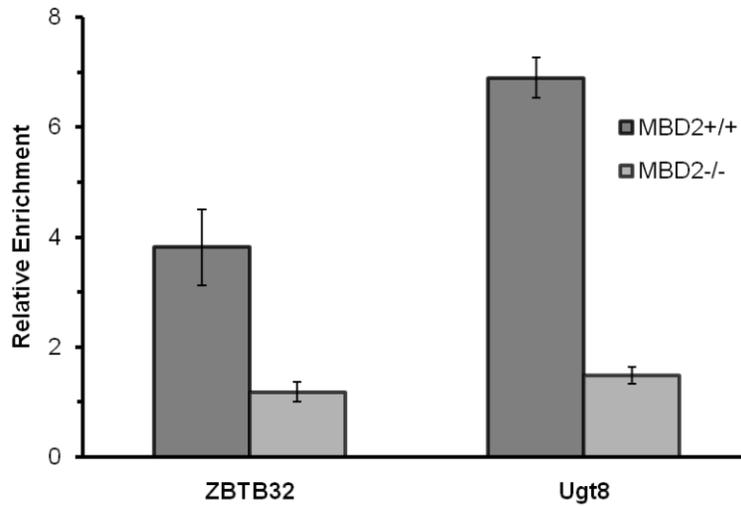


Figure 21. Chromatin Immunoprecipitation of MBD2 in ZBTB32 gene. ChIP assays were performed using anti-MBD2 and goat IgG control antibodies in adult erythroblasts from MBD2 wild type and knockout mice. DNA was analyzed using real-time PCR the Ugt8 gene CpG island and ZBTB32 gene CpG island. Data was normalized to endogenous amylase gene levels to account for loading errors and IgG controls to account for non-specific pull-downs. Error bars indicate mean \pm SD. (Unpublished data by Rupon JW and Gnanapragasam MN)

murine cell system of functional validation. These cells primarily express adult human and murine β -type globin genes and hence serve as a valuable system to study activators of the fetal γ -globin gene (Blau et al. 2005). Knockdown of Mbd2 in CID cells lead to elevated expression of Zbtb32 (Figure 22B), which makes it a valid system to study functional effects of Zbtb32. Transient over-expression of Zbtb32 from a plasmid carrying the coding sequence of human ZBTB32 was achieved by nucleofection and desired level of expression was observed (Figure 22C). Over-expression of ZBTB32 caused elevation of the γ -globin gene expression by ~ 2.5 fold at 72 hours post-transfection (Figure 22D).

In the studies presented in Chapter 3, we validate the function of ZBTB32 in γ -globin gene activation in primary adult human erythroid cells. Since miRNAs are also important epigenetic regulators of gene expression through post-transcriptional mechanisms, we carried out a miRNA microarray analysis of differentially regulated miRNAs in erythroid cells of adult MBD2 knockout mice. We identify a single miRNA, miR-210, which regulates the expression of the fetal γ -globin gene through a direct or indirect mechanism. Understanding the mechanism of MBD2-mediated fetal globin gene regulation gives insight into the role of epigenetics in the fetal-to-adult globin switch and hematopoietic development in general.

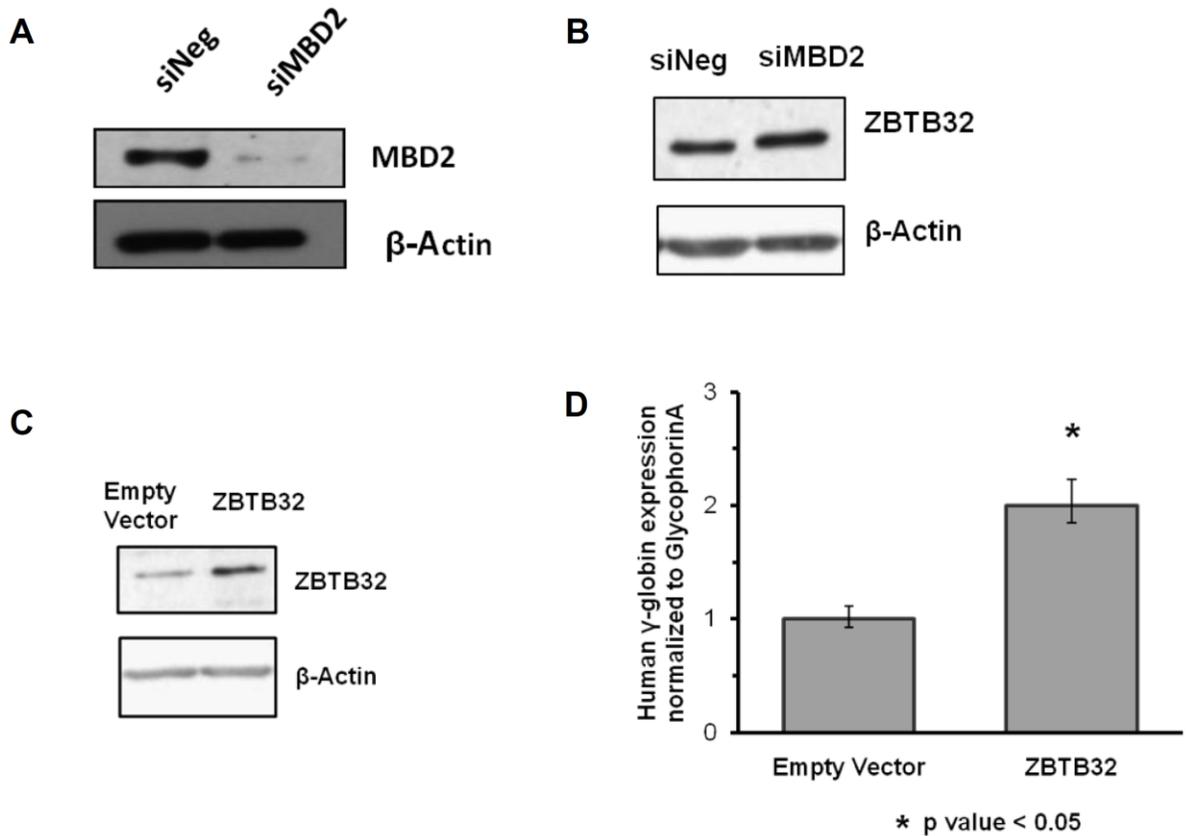


Figure 22. Validation of candidate genes in CID β-YAC cells. CID dependent β-YAC bone marrow cells were transiently transfected with siMBD2 and siNeg control (n=3). They were analyzed for (A) MBD2 protein levels by western analysis; and (B) ZBTB32 protein level by western analysis. All the experiments were performed at 72h post transfection. Error bars represent mean ±SE. (C) Functional studies in CID β-YAC cells: CID dependent β-YAC bone marrow cells were transiently transfected with the coding region of ZBTB32 (n=3) along with their corresponding empty vector control. The over-expression of ZBTB32 was verified using western analysis. Error bars indicate mean ± SE (D) mRNA levels of γ-globin was calculated using Q-RT PCR in the ZBTB32 over-expressing cells. Gamma globin expression was normalized to GlycophorinA. (Unpublished data by Gnanapragasam MN)

II. Methods:

Lentiviral knockdown and expression:

shMBD2 (GGGTAAACCAGACTTGAA), hsa-miR-210 (CTGTGCGTGTGACAGCGGCTGA) and shScramble sequences were cloned into the pRRL.H1 shRNA vector and packaged into lentivirus by calcium phosphate transfections in HEK 293T cells as described earlier. The coding sequence of human TZFP cloned in the pCEP-Flag plasmid was a generous gift from Dr. Maureen Hoatlin (Oregon Health and Science University, Portland, Oregon, USA). To be used for stable over expression studies, this sequence was later cloned into the pTRIPZ lentiviral vector with an enhanced green fluorescent protein (eGFP) as the reporter gene. This sequence of TZFP was subcloned into a TA cloning vector using PCR amplification to introduce restriction cut sites for Mlu1-Cla1 and followed by ligation into the TA vector using the TA Cloning® kit (Life Technologies). A clone verified by sequencing was cloned into the pTRIPZ lentiviral expression vector (ThermoScientific) by Mlu1-Cla1 restriction digestion of the respective TA vector, gel purification of the desired band and ligation into pTRIPZ lentiviral vector. This stable expression vector was packaged into lentivirus as mentioned above and used to infect CD34+ hematopoietic progenitor cells.

CD34+ cell isolation:

CD34+ cells were isolated from de-identified apheresis packs donated by the VCU Bone Marrow Transplant Clinic. Cells were thawed in a 37°C incubator and then poured into a 50 mL tubes. Cells were mixed with 1 volume of 1X PBS containing 2% FBS. ~30 mL volume of cells were then added to a 50 mL tube containing 15 mL of Ficoll-paque-plus

(StemCell Technologies Inc). The mix was spun at 400xg for 30 minutes at room temperature with no brakes. The middle layer containing the mononuclear cells was extracted and placed on a fresh tube, where 40 mL of 1X PBS with 2% FBS was then added. Mononuclear cells were then filtered through a 70 μ M nylon filter () and spun at 250xg for 6 minutes at room temperature. The epaulet was resuspended in 4-5 mL of DNaseI solution (StemCell Technologies Inc.) and incubated at room temperature for 15 minutes. Following DNaseI incubation, 40 mL of 1X PBS with 2% FBS and 1mM EDTA was added to dilute the solution, followed by spinning the cells at 250xg for 6 minutes at room temperature. The cell pellet was resuspended in 1X PBS with 2% FBS and 1mM EDTA to a density of 2×10^8 cells/mL. EasySep kit (StemCell Technologies Inc.) was used for positive selection of CD34+ cells as per the manufacturer's protocol. Cells were then grown in an expansion medium consisting of StemSpan SFEM II medium (StemCell Technologies Inc.) with 1X CC100 cytokine cocktail (StemCell Technologies Inc.) and 2% penicillin and streptomycin.

CD34+ cell infection and culture:

2×10^5 CD34+ cells were plated per well on a 12-well tissue culture plate in 50 μ L of expansion medium on the day of infection. The medium containing lentivirus was then mixed with 4 μ g/mL of polybrene and 650 μ L of the lentiviral solution was added per well. The cells were incubated overnight in 37°C and 5% CO₂ and the following day, 2X expansion medium was added to the cells. 96 hours post-infection, GFP+ cells were selected by flow cytometry on a BD FACSAria™ II flow cytometer instrument and maintained after in a growth medium containing IMDM supplemented with 20% FBS

(Hyclone), 10 ng/mL stem cell factor (SCF), 1 u/mL of erythropoietin (EPO), 1 ng/mL IL-3, 1 uM dexamethasone, 1 uM estradiol and 2% penicillin and streptomycin for three days. The cells were then washed 2 times with IMDM and grown in a differentiation medium containing IMDM supplemented with 20% FBS (Hyclone), 1 u/mL of erythropoietin (EPO), 10 ng/mL insulin, and 2% penicillin and streptomycin. Cells were grown in the differentiation medium for 10 days and changed to a fresh medium once in every two days.

Cell surface marker analysis by flow cytometry:

On day 10 of differentiation, the differentiated erythroid cells were washed with upto 10 mL of cell staining buffer containing 1X PBS with 10% FBS and spun at 300xg for 5 minutes at room temperature. The cell pellet was then resuspended in 100 uL of the cell staining buffer and incubated with 0.06 ug of CD71 (Transferrin receptor) antibody and 0.015 ug of CD235a (Glycophorin A) antibody for 30 minutes on ice in the dark. Cells were then washed twice with 1.5 mL of cell staining buffer by spinning at 300xg for 5 minutes and resuspending in fresh buffer each time. After the final wash, cells were resuspended in 400 uL of buffer and analyzed and sorted by a BD FACSAria™ II machine. These sorted cells were used for gene and protein expression.

RNA Isolation and Quantitative PCR (qPCR):

Total RNA was isolated from transduced cells X days post-infection by phenol-chloroform extractions (TRIzol, Life Technologies) as per the manufacturer's protocol. Samples were then subjected to DNaseI digestion as follows:

1 uL of 10X DNaseI Buffer (Ambion)

0.25 uL of DNaseI

0.25 uL of SUPERase Inhibitor

1 ug of total RNA

The volume was made unto 10 uL with DEPC treated H₂O.

The DNaseI reaction was incubated at 37°C for 30 minutes followed by 75°C for 10 minutes.

5 uL (500 ng) of this RNA product was then converted to cDNA using the iScript cDNA synthesis kit (BioRad) as follows:

2.0 uL of 5X reaction buffer

0.5 uL of reverse transcriptase enzyme and

2.5 uL of DEPC teated H₂O

This 10 uL reaction mix was subjected to following thermal cycling conditions: 25°C for 5 minutes, 42°C for 30 minutes and 85°C for 5 minutes. This cDNA was used for gene expression analysis using quantitative PCR on an ABI 7300 Real Time PCR system; FastStart Universal SYBR Green Master mix (Roche Diagnostics, Germany) or Taqman FastStart Universal Probe master mix (Roche Diagnostics, Germany) were used.

Thermal cycling conditions for the PCR were: 1 cycle at 50°C for 2 minutes and 95°C for 10 minutes followed by 40 cycles at 95°C for 15 seconds and 60°C for 1 minute. The 2- $\Delta\Delta$ Ct method was used for analysis. Relative quantification was determined using the SDS 1000 software. Primers used for qPCR are listed in Table 3.

MiRNA microarray analysis:

MBD2^{-/-} mice and wild type mice were made anemic by treatment for two days with intra-peritoneal injection of 1-acetyl-2-phenylhydrazine (10 mg/ml; Sigma) at a dose of 0.4 mg/10 g weight of mice. This was done so that spleen becomes greater than 90% erythroid. On the fifth day the spleens from four MBD2^{-/-} and wild type mice were harvested and RNA was extracted using Trizol (Sigma) as explained earlier. These RNA samples were sent to Dr. Catherine Dumur in the Department of Pathology at Virginia Commonwealth University for further processing. The Affymetrix GeneChip® miRNA array was used. This array covers miRNAs from with 71 organisms on a single array, including 609 murine miRNAs from the Sanger miRNA database v11. The FlashTag™ Biotin HSR RNA Labeling Kit (Genisphere Inc., Hatfield, PA) was used to label total RNA, containing low molecular weight RNA species. Briefly, 500 ng of total RNA from your samples was subjected to a brief Poly(A) tailing reaction followed by ligation of a biotinylated signal molecule to the target RNA sample. The labeled RNA was then hybridized onto GeneChip® miRNA arrays, and scanned on an Affymetrix GeneChip® Scanner 3000 G7 according to the GeneChip® Expression Analysis Technical Manual procedures (Affymetrix, Santa Clara, CA).

For the statistical analysis, we used the following workflow for miRNA microarray data summarization: detection of each probe set above background (using a Wilcoxon Test); background adjustment (using the Robust Multiarray Analysis - RMA – algorithm); quantile normalization; and median polish probe set summarization. Prior to further statistical analysis, the dataset was restricted to probe sets representing murine miRNAs (609 mmu-miRNAs). I identified probe sets having average signals below

background (in grey). Thereafter, for each probe set, a two sample t-test will be used to compare MBD2^{+/+} to MBD2^{-/-} samples. Resulting p-values were used to estimate the false discovery rate (FDR) using the Benjamini-Hochberg method.

Quantitative-PCR validation of the miRNA, miR-210:

TaqMan® miRNA assay kit was obtained from Life Technologies designed specifically to detect and quantify the mature form of miR-210 (hsa-miR-210). Briefly, total RNA was isolated from erythroid cells using TRIzol® reagent as previously described. In the reverse transcription step, cDNA was transcribed from total RNA samples using primers specific for miR-210 and reagents from the TaqMan® MicroRNA Reverse Transcription kit. In the PCR step, PCR products were amplified from the cDNA samples using the TaqMan MicroRNA Assay together with the TaqMan® Universal PCR master mix (Figure 23). U6-snRNA is a small non-coding RNA and is a component of the U6 small nuclear ribonucleoprotein complex, was used as the endogenous control for expression of small RNAs.

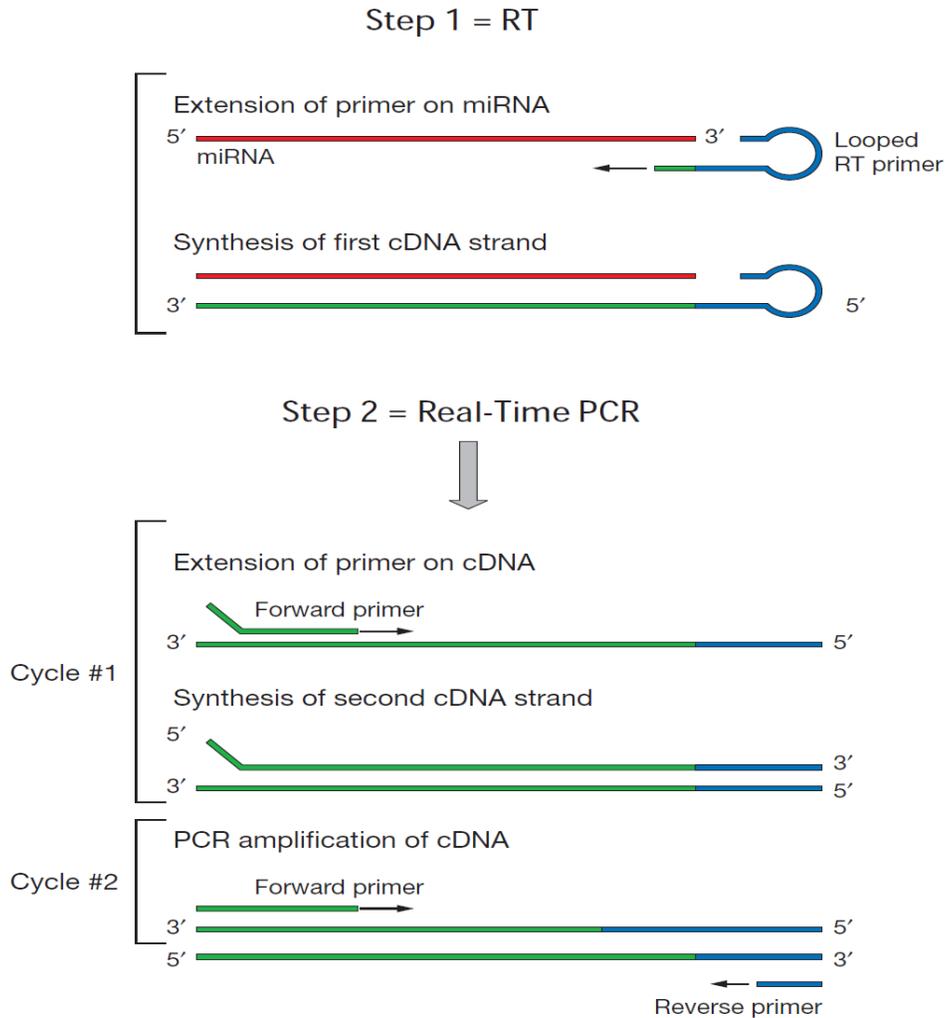


Figure 23 Two step miRNA qPCR. 1. In the reverse transcription (RT) step, cDNA is reverse transcribed from total RNA samples using specific miRNA primers from the TaqMan MicroRNA Assays and reagents from the TaqMan® MicroRNA Reverse Transcription Kit. 2. In the PCR step, PCR products are amplified from cDNA samples using the TaqMan MicroRNA Assay together with the TaqMan® Universal PCR Master Mix.

Western Blot Analysis:

Cell pellets were resuspended in 4% SDS and sonicated for 5 minutes (30 seconds ON, 30 seconds OFF setting). Total protein was quantitated using BioRad DC Protein assay (BioRad) followed by UV-Vis spectrophotometry. 100 ug of protein mixed with 2X laemmli buffer was run on a 10% SDS-PAGE gel. The protein was transferred onto a PVDF membrane (0.45 uM, Millipore) for 1 hour at a constant voltage of 100V. The membrane was then blocked in a blocking buffer consisting of 5% milk in 1X PBST for one hour at room temperature. Following this, the membrane was incubated with the primary antibody of interest at room temperature for 1 hour. The membrane was then washed 3 times by 1X PBST for 5 minutes each and incubated with the secondary antibody in a blocking buffer solution for 45 minutes at room temperature. The membrane was again washed 3 times with 1X PBST for 5 minutes each and then incubated with SuperSignal® West Pico Chemiluminescence substrate or SuperSignal® West Dura Extended duration substrate (Thermo Scientific) for 5 minutes according to the manufacturer's protocol. The membrane was then exposed to an autoradiography film (Bio Express) and developed to study protein bands of interest. Alternatively, the membrane incubated with the chemiluminescent substrate was exposed in the Odyssey Infrared Imaging System (*LI-COR* Biosciences).

Antibodies: Anti-MBD2 (Santa Cruz sc-1244), anti-TZFP (Abcam 26066), anti-CD235a (eBioscience # 12-9987-82), anti-CD71 (eBioscience # 14-0719-82).

Primers: The sequences of the primers used in chapter 3 are in Table 3.

Table 3. List of primers used for qPCR in Chapter 3

	Forward	Reverse	Probe
Human γ -globin	GTG GAA GAT GCT GGA GGA GAA A	TGC CAT GTG CCT TGA CTT TG	AGG CTC CTG GTT GTC TAC CCA TGG ACC
Human β -globin	GCA AGG TGA ACG TGG ATG AAG T	TAA CAG CAT CAG GAG TGG ACA GA	CAG GCT GCT GGT GGT CTA CCC TTG GAC CC
Human MBD2	AAG AGC GAG TAC AGC AAG TAC GCA	TTC TGT ATC AGC AGC TCG CGA CAA	TGG AAG AAG CAC TGA TGG CAG ACA TC
Human TZFP	GGA CAA AGG CTA CAG GAG AAG	TTC ACT CAA GCA TCA GAT GGA G	CAC TAC CGA GTC CAC ACA GGA GAG AA
Human GAPDH	TCG ACA GTC AGC CGC ATC TTC TTT	ACC AAA TCC GTT GAC TCC GAC CTT	AGC CAC ATC GCT CAG ACA CCA TGC

III. Results:

A. Over-expression of ZBTB32 in CD34+ adult human erythroid cells activates the fetal γ -globin gene:

A better model than the mouse bone marrow CID cells to study regulators of fetal γ -globin gene silencing is an adult human progenitor derived erythroid cell system. As described earlier, these cells express <1-2% fetal hemoglobin out of the total globin at the end of their *in vitro* erythroid differentiation process. To validate this system to study the function of ZBTB32, we knocked down MBD2 in these cells and quantitated the increase in ZBTB32 expression. Knockdown of MBD2 by ~80% (Figure 24A) causes ~1.5-fold increase in the expression of ZBTB32 in these cells (Figure 24B). Therefore, in line with our previous data, we hypothesized that over-expression of ZBTB32 in these primary adult cells will activate expression of the fetal γ -globin gene. The coding sequence of ZBTB32 was stably over-expressed in these cells through lentiviral transduction, followed by growth and erythroid differentiation of the cells. Over-expression of ZBTB32 at the mRNA level was confirmed (Figure 24C). On day 10 of differentiation, erythroid cells were selected by sorting for CD71/CD235a double positive cells by flow cytometry. Analyses demonstrated that over-expression of ZBTB32 increased the expression of the γ -globin gene by 2-fold compared to the empty vector control (Figure 24D). At the same time, over-expression of ZBTB32 did not cause a significant increase in adult β -globin gene expression, which suggests that the effect of ZBTB32 on fetal γ -globin gene expression is specific to that gene and not a non-specific effect on the β -globin gene locus (Figure 24E). Also, over-expression of ZBTB32 did not interfere with the normal differentiation of the erythroid

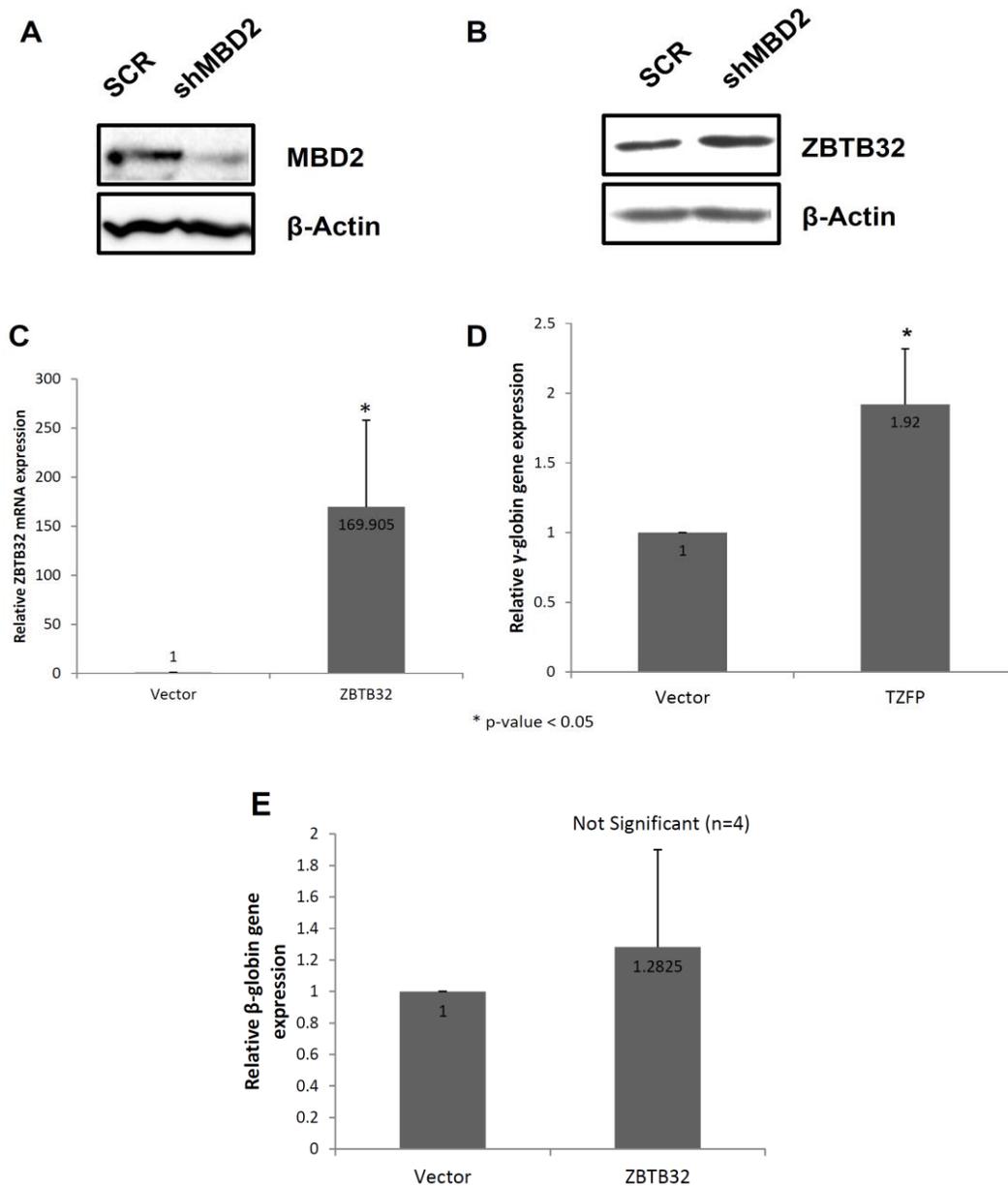


Figure 24. Validation and functional analysis of ZBTB32 in CD34⁺ adult erythroid cells. CD34⁺ primary adult human CD34⁺ erythroid progenitor cells were transduced with anti-MBD2 shRNA and SCR control expressing lentivirus. Western blot analysis shows (A) ~75% knockdown of MBD2 in these cells at the protein level and (B) increased expression of ZBTB32 at the protein level in the MBD2 knocked down CD34⁺ erythroid cells. The western blot is representative of three independent experiments. (C) CD34⁺ erythroid progenitor cells were transduced with lentivirus expressing the coding sequence of hZBTB32 and the empty vector control. qPCR analysis shows over-expression of hZBTB32 mRNA in these cells as compared to the empty vector control. (D) Human fetal γ -globin gene mRNA expression was quantitated using qPCR and normalized to GAPDH. hZBTB32 over-expressing CD34⁺ cells show ~2-fold increased expression of the γ -globin gene (p-value < 0.05). (E) Human adult β -globin gene mRNA expression was quantitated using qPCR and normalized to GAPDH. hZBTB32 over-expressing CD34⁺ do not show a significant increase in the expression of β -globin gene (p-value > 0.05). Error bars denote standard deviation of three independent experiments.

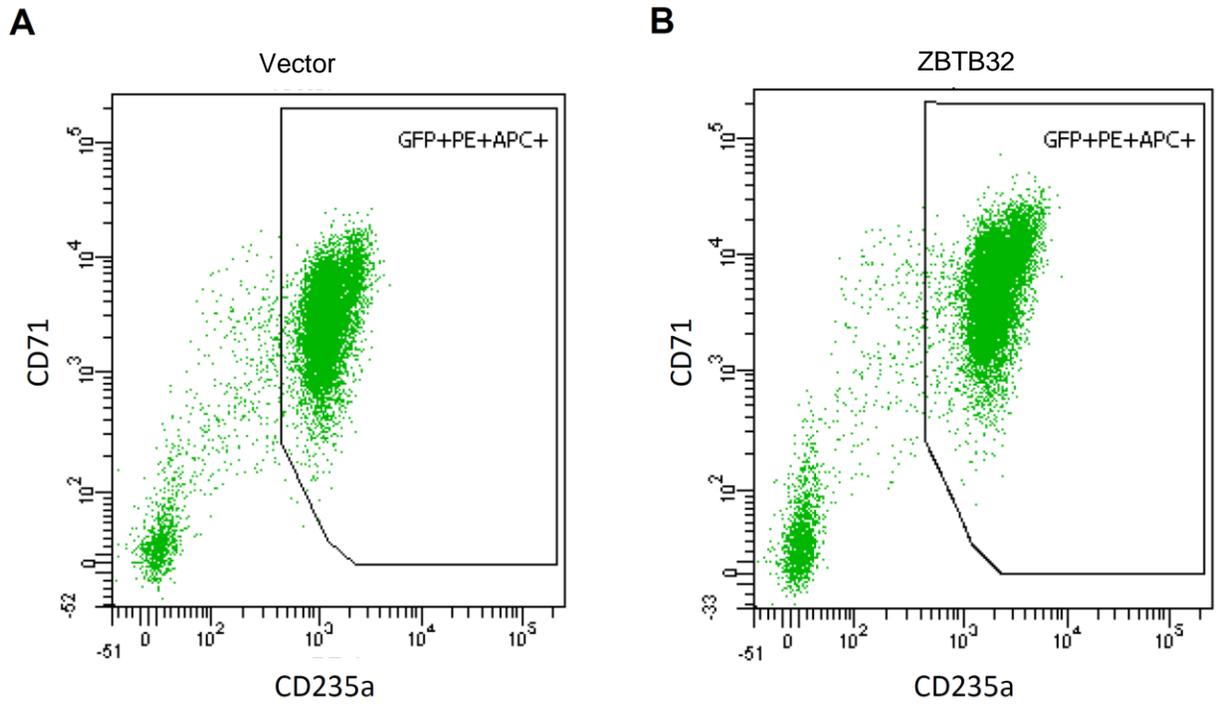


Figure 25. ZBTB32 over-expressing CD34+ cells show normal erythroid differentiation profile. Expression of cell surface markers CD71 and CD235a was analyzed using flow cytometry to study the erythroid differentiation profile of the transduced cells. Both empty vector expressing control cells and hZBTB32 expressing cells show normal erythroid differentiation. Level of erythroid differentiation is similar in both the groups.

cells as can be seen from the CD71/CD235a cell-surface marker analysis. Up to 95% of the cell population consisted of double positive erythroid cells (Figure 25).

B. MicroRNA expression profile of MBD2 knockout mice:

Three wild-type and three MBD2^{-/-} mice were injected intra-peritoneally with 1-acetyl-2-phenylhydrazine in order to induce hemolytic anemia, an event which results in extra medullary hematopoiesis in the spleen, making the spleen >90% erythroid. Total RNA from the spleen was harvested and sent over to our collaborator Dr. Catherine Dumur (Dept. of Pathology, VCU) for murine miRNA microarray analysis. This array covers miRNAs from 71 organisms on a single array, including 609 murine miRNAs from the Sanger miRNA database v11. Total RNA from our samples was ligated with a biotin label followed by hybridization onto GeneChip® miRNA arrays, and it was scanned on an Affymetrix GeneChip® Scanner 3000 G7.

Statistical analysis was restricted to the 609 murine miRNA probe sets. A two sample student t-test was used to compare wild-type and MBD2^{-/-} samples. The resulting p-values were used to calculate the false discovery rate (FDR) using the Benjamini-Hochberg method. Only one miRNA probe set, miR-210, showed a significant change (FDR<5%) with an ~3 fold induction in MBD2^{-/-} compared to wild-type samples (Figure 26).

C. Validation of miR-210 expression in adult erythrocytes from β -YAC transgenic mice:

To validate the differentially expressed miR-210 identified from the microarray screen, we conducted miRNA quantitative PCR on the total RNA obtained from spleens of wild-

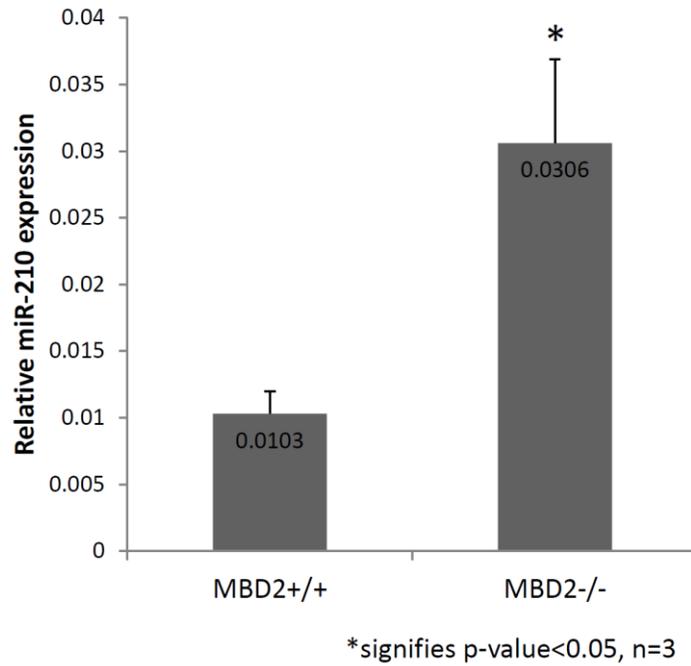


Figure 26. Validation of miR-210 expression in mouse splenic erythroblasts. Total RNA from three wild-type and three MBD2 knockout (MBD2^{-/-}) mice was analyzed for mature miR-210 expression by qPCR. MBD2^{-/-} mice show ~3-fold (p-value<0.05) increased expression of miR-210 in the splenic erythroblasts, validating the results obtained from the microarray screen. Error bars denoted standard deviation of three different mice for each group.

type and MBD2^{-/-} mice. qPCR verification did show a 3-fold increase in the expression of miR-210, using U6 snRNA as an endogenous control (p-value<0.05, n=3). U6 snRNA is a non-coding small nuclear RNA (snRNA) component of the U6 snRNP (small nuclear ribonucleoprotein) RNA-protein complex which is involved in the splicing of pre-mRNAs.

miR-210 is located on chromosome 11p15.5 and is transcribed from the intron 1 of a non-protein coding processed RNA, the miR-210 host gene. The predicted size of the pri-miR-210 is 2927 bp which is almost the same as that of miR-210 host gene. The upstream regulatory region contains HIF-1 α binding sites and several studies have reported miR-210 as a robust hypoxia-inducible miRNA. Analysis of the upstream regulatory region of the murine miR-210 host gene by the UCSC genome browser also shows the presence of a CpG-island between -586 and -19 bp of the promoter (Zhou et al. 2007) and differential methylation of this region has been reported in various pathophysiological conditions (Kiga et al. 2014; Lages et al. 2011; Xiong et al. 2012). This information suggests a potential direct role of MBD2 in epigenetic regulation of miR-210.

D. Functional analyses of miR-210 expression in adult mouse bone marrow cells:

Given that MBD2^{-/-} mice show elevated γ -globin expression in adult erythrocytes as well as a concomitant increase in expression of the miR-210, we hypothesized that MBD2 might regulate γ -globin gene expression in part through miR-210. Hence we postulated that over-expression of miR-210 in mouse bone marrow CID cells and primary adult human erythroid cells will be able to induce γ -globin gene expression. Chemical inducer of dimerization (CID) dependent β -YAC mouse bone marrow cells are

widely used to study human γ -globin gene regulation. These cells are derived from the bone marrow of β -YAC transgenic mice and can be propagated in culture by liquid media containing an artificial chemical inducer of dimerization (Blau et al., 2005). This chemical induces thrombopoietin signaling through homodimerization of FKBP (FK506 binding protein) binding domains which are in turn fused to thrombopoietin signaling domain. The resultant signaling pathway enables indefinite proliferation of these bone marrow cells. These cells predominantly express adult human β -globin and very low levels of fetal γ -globin (<1% of β -globin), akin to adult human erythroid cells, thus making it a good *in vitro* system to study regulators of the fetal γ -globin gene.

Stable over-expression of miR-210 was achieved by lentiviral delivery of the mature miR-210 in these cells. Up to 2-fold ectopic expression of mature miR-210 was achieved in these cells as quantitated by qPCR. Quantitation of human fetal γ -globin and adult β -globin mRNA levels was further carried out in these cells. Fetal γ -globin gene showed an increased trend in expression by ~ 2.4 fold when normalized to glycophorin A gene expression, however this was not statistically significant ($n=3$, p -value >0.05) (Figure 27A). Interestingly, quantitation of the human β -globin gene also showed a similar but significant increase in expression by ~ 3.1 -fold ($n=3$, p -value <0.01), thus indicating an overall increased β -type globin gene expression, possibly brought about by induced erythroid differentiation of the multi-potent bone marrow cells (Figure 27B).

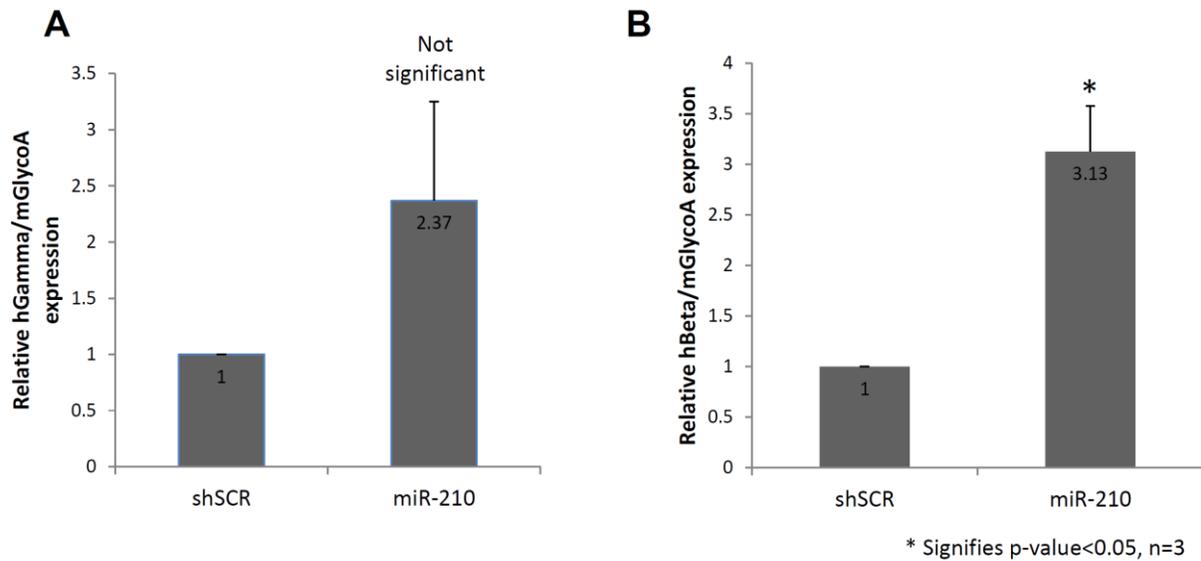


Figure 27. miR-210 regulates human β -type globin gene expression in β -YAC mouse bone marrow cells. CID-dependent β -YAC mouse bone marrow cells were transduced with lentivirus expressing miR-210 and SCR control. Human fetal γ -globin and adult β -globin gene mRNA expression was quantitated by qPCR and normalized to mouse Glycophorin A expression. (A) Human γ -globin gene shows an increased expression with miR-210 expression, although it is not statistically significant (n=3, error bars denoted standard deviation). (B) miR-210 over-expressing CID cells show ~3-fold increased expression of human adult β -globin gene as compared to SCR control infected cells (p-value<0.05, error bars denote standard deviation of 3 independent experiments).

E. Validation of miR-210 expression in MBD2 knockdown CD34⁺ hematopoietic progenitor cells:

Following a preliminary functional analysis of miR-210 in the murine erythroid system, we investigated the role of miR-210 in the human erythroid system. To this end, we quantitated mature miR-210 expression in stable MBD2 knocked down CD34⁺ cells by qPCR. Knockdown of MBD2 in CD34⁺ cells causes a modest 2-fold increase in expression of mature miR-210. However, due to a large variability in the fold-change, the increase in miR-210 expression was not found to be statistically significant over multiple repeats (n=8, p-value>0.05) (Figure 28). This observation may arise from the fact that erythroid cells harvested for mRNA analysis are often a heterogeneous population. Although the differentiated cells are sorted for CD71/CD235a double positive populations only, within this erythroid population one may envision cells at different stages of erythroid differentiation and hence have different miRNA expression levels.

Alternately, we hypothesized that a significant increase in miR-210 expression may require very high levels of MBD2 knockdown, which would also indicate high level of γ -globin gene expression. For each experiment conducted on the same sample of CD34⁺ cells, in Table 4 we have plotted values of the percentage of MBD2 knockdown, $\gamma/(\gamma+\beta)$ globin fold change in MBD2 knockdown vs. SCR cells and the increase in miR-210 expression as compared to the SCR controls. It can be observed that with highest level of MBD2 knockdown and γ -globin gene induction, miR-210 expression is significantly higher as compared to its SCR control. As the percentage of

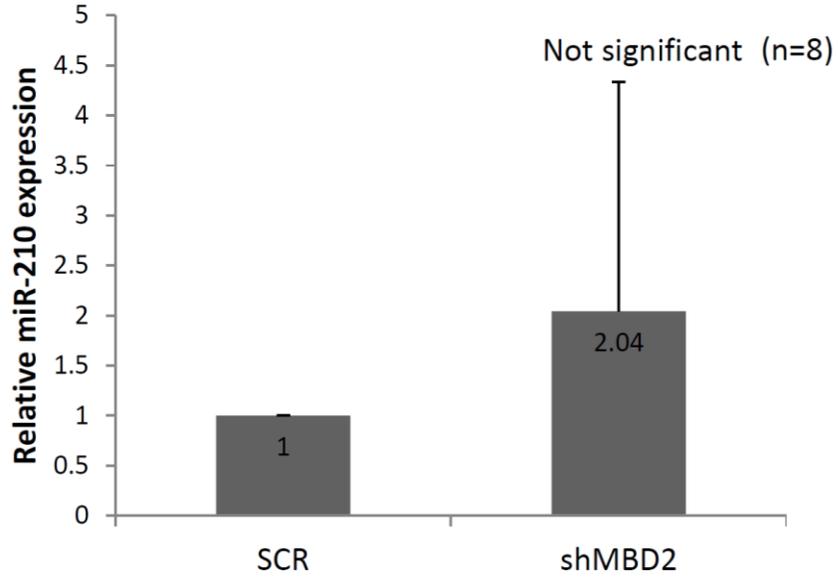


Figure 28. miR-210 expression in MBD2 knockdown CD34⁺ erythroid cells. MBD2 was knocked down by about 75% in CD34⁺ erythroid cells, followed by their *in vitro* erythroid differentiation. miR-210 expression was quantitated by qPCR and normalized to U6RNA expression. miR-210 expression was not statistically significant in MBD2 knockdown erythroid cells (n=8, error bars denote standard deviation).

MBD2 knock down and γ -globin gene expression reduces, the fold-change in expression of miR-210 becomes variable and less consistent. Since the most robust induction of miR-210 was observed in MBD2 knockout erythroid cells, our data suggests that increased expression of miR-210 in erythroid cells may require >90% elimination of MBD2.

Table 4. miR-210 expression in a background of varying MBD2 knockdown and γ -globin gene expression in erythroid cells. Expression of miR-210, MBD2 and γ -globin gene was measured by qPCR. Percentage of MBD2 knockdown positively correlates with percentage of absolute γ -globin gene expression i.e. \uparrow MBD2 knockdown \propto \uparrow γ -globin gene expression and this is statistically significant. miR-210 expression does not consistently show increased expression with higher MBD2 knockdown and higher γ -globin gene expression.

% MBD2 Knockdown	Gamma/(gamma+beta) fold change compared to SCR	Increase in miR210 expression compared to SCR
90%	135 fold (14.29% HbF)	9.43
66%	15.65	2.14
51%	14.14	0.97
45%	7.62	1.03
—	9.28	0.67
no kd	4.84	no change
11%	2.87	no change

F. Functional analysis of miR-210 in CD34⁺ hematopoietic progenitor cells:

Although we were limited in validating the expression of miR-210 in MBD2 knockdown CD34⁺ cells, we nevertheless proceeded with validating the function of miR-210 in primary human erythroid cells. We over-expressed mature miR-210 in the primary human erythroid cells through lentiviral expression. CD34⁺ human erythroid cells were transduced with SCR and miR-210 expressing lentivirus and transduced cells were selected by flow cytometry based on their expression of green fluorescent protein (GFP). We achieved a very robust expression of mature miR-210 in the erythroid cells as quantitated by qPCR (n=6, p-value<0.01) (Figure 29A). Following growth and differentiation of the transduced CD34⁺ cells *in vitro* to erythrocytes, we quantitated the expression of the fetal γ -globin gene on day 10 of differentiation. We observe an increased trend in expression of the fetal γ -globin gene relative to total γ - and β -globin gene expression in miR-210 over-expressing cells as compared to SCR control cells (n=6, p-value>0.05) (Figure 29B). Given the role of miR-210 in erythroid differentiation and its effect on both γ - and β -globin gene induction in mouse erythroid cells, we analyzed absolute expression of γ - and β -globin gene expression by normalizing its expression to an endogenous gene - GAPDH in SCR and miR-210 over-expressing cells. MiR-210 over-expressing erythroid cells show ~ 2 fold increase in γ -globin gene (n=3, p-value>0.05) and a parallel 3-fold increase in β -globin gene expression (n=3, p-value<0.05) (Figure 29C,D), similar to the observation in CID mouse bone marrow cells. Over-expression of miR-210 does not affect the differentiation profile of the cells, as can be seen from the similar cell surface marker expression by shSCR and miR-210 expressing cells (Figure 30). These results lead us to an understanding of the function

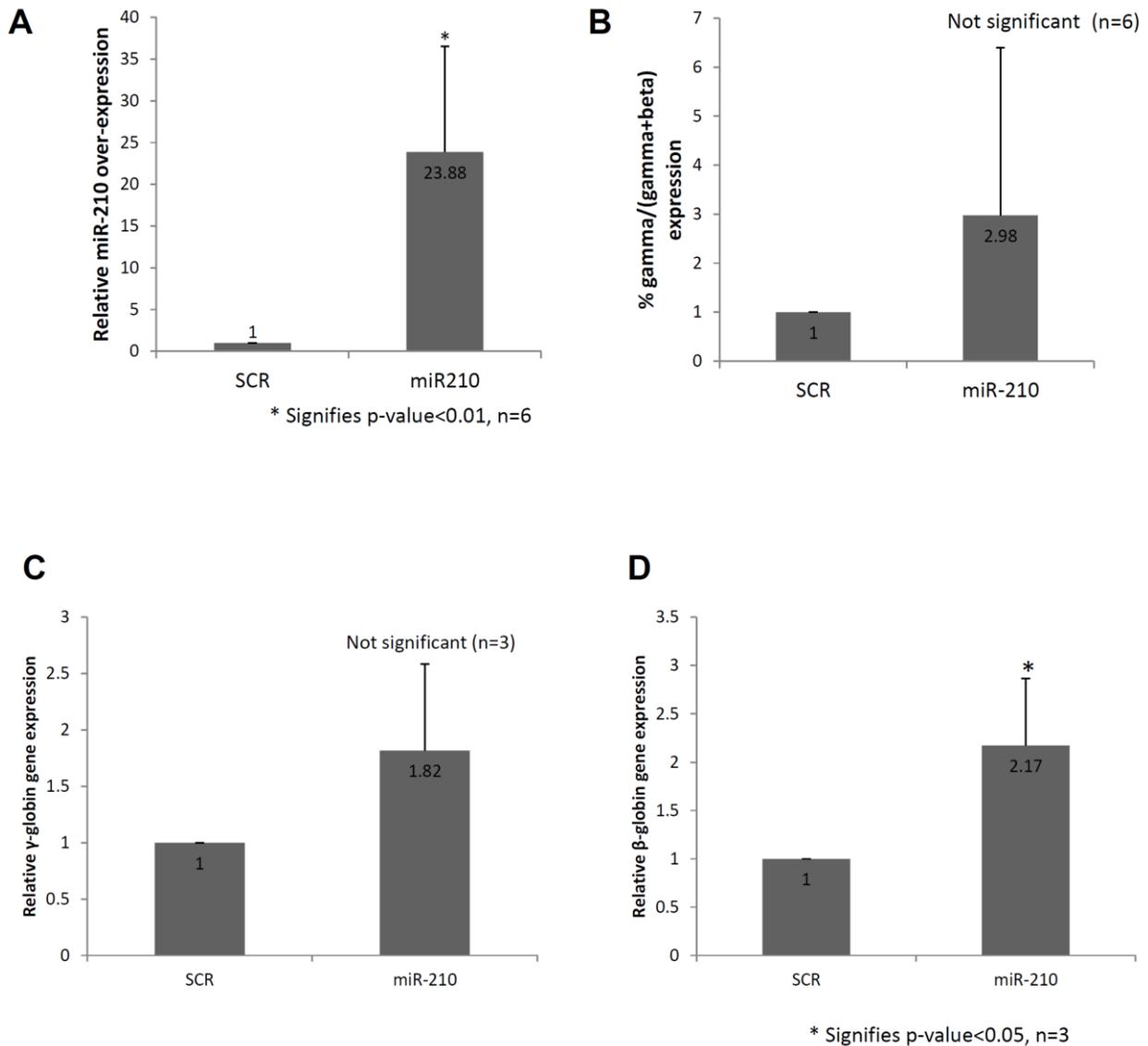


Figure 29 Functional analysis of miR-210 over-expression in CD34⁺ cells. miR-210 was stably over-expressed in adult erythroid progenitors followed by their *in vitro* erythroid differentiation. miR-210 and globin gene expression was quantitated using qPCR. (A) Over-expression of miR-210 by upto 23-fold as compared to SCR controls was achieved by lentiviral delivery of mature miR-210 (p-value < 0.01, n=6, error bars denote standard deviation). (B) Absolute gamma globin gene expression (gamma/gamma+beta) shows an increased expression trend in miR-210 over-expressing erythroid cells as compared to SCR controls. However, the fold change is not statistically significant (n=6, error bars denote standard deviation). (C) Human fetal γ -globin gene mRNA expression was quantitated using qPCR and normalized to GAPDH. miR-210 over-expressing CD34⁺ cells show increased expression of the γ -globin gene, but it is not statistically significant (p-value > 0.05). (D). Human adult β -globin gene mRNA expression was quantitated using qPCR and normalized to GAPDH. miR-210 over-expressing CD34⁺ show a significant increase in the expression of β -globin gene by ~2-fold as compared to SCR control (p-value < 0.05). Error bars denote standard deviation of three independent experiments.

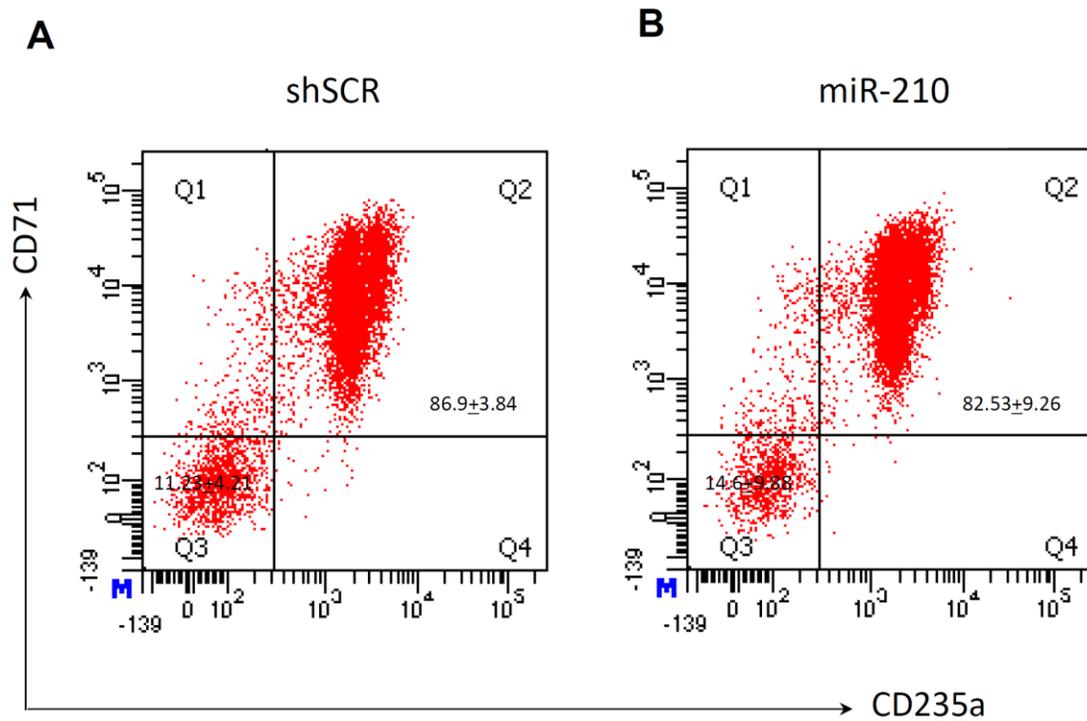


Figure 30. Over-expression of miR-210 in erythroid cells does not affect their normal differentiation profile. Percentage of CD71+/CD235a+ erythroid cell population (Quadrant Q2) is similar in (A) SCR expressing control cells and (B) miR-210 over-expressing cells. (Percentage values denote average and standard deviation of 6 independent experiments).

of miR-210 in β -type globin gene expression in erythroid cells, possibly through modulation of erythroid differentiation and a general effect on transcription of genes belonging to the β -globin gene locus. It also sheds light on a novel downstream effector of MBD2 through which it may mediate its effect on γ -globin gene expression.

IV. Discussion:

Methyl CpG-binding domain protein 2 (MBD2) mediates developmental silencing of the fetal γ -globin gene, although without direct interaction with cis-regulatory sequences near the γ -globin gene promoter (Rupon et al. 2006a). In fact, MBD2 does not bind to any promoter sequences within the β -globin gene locus. This result could be anticipated given that the entire human β -globin gene locus is CpG sparse and the MBD2-NuRD complex requires at least 12 methylated CpG dinucleotides for binding methylated DNA *in vitro* (Ginder, Gnanapragasam, Mian 2008; Meehan et al. 1989; Mian et al. 2011; Wade 2001). *In vivo* MBD2 binds only to densely methylated CpG islands. Hence, we set out to define the mechanism of action for developmental silencing of the fetal γ -globin gene by MBD2.

We hypothesized that MBD2 may regulate silencing of γ -globin gene expression in adult erythrocytes by an indirect mechanism which may involve other gene(s) or miRNA(s). MiRNAs have recently evolved as vital small non coding RNAs that play important roles in regulating gene expression. Loss of MBD2 would activate expression of a gene or a miRNA, which would in turn activate the expression of the fetal γ -globin gene. To test our hypothesis and identify downstream effectors of MBD2, we carried out gene and miRNA expression microarrays in splenic erythrocytes of MBD2^{-/-} and wild

type mice. 93 genes were found to be up-regulated in MBD2^{-/-} mice compared to wild-type mice; a single miRNA, miR-210 was found to be increased in expression in MBD2^{-/-} erythrocytes.

So far MBD2 has been shown to regulate a limited set of genes in normal development. MBD2^{-/-} mice show disrupted expression of IL-4 and interferon- γ in helper T cells (Kersh 2006), abnormal expression of exocrine pancreatic genes in the colon (Berger et al. 2007), elevated expression of the Xist gene (Barr et al. 2007) and dysregulation of olfactory receptor neuron differentiation (Macdonald et al. 2010). Mbd2^{-/-} mice also show reduced adenoma formation in *Apc*^{min/+} mice by attenuating the Wnt signaling pathway (Phesse et al. 2008). Limited studies have been performed to understand the function of MBD2 in normal development. In the erythroid system in mice, our studies identify a new set of protein coding genes that are regulated by MBD2 in normal physiology. They indicate that loss of MBD2 disrupts the expression of just a few hundred genes rather than a large set of thousands of genes in the erythroid compartment. Similarly, miRNA microarray results showed a single differentially regulated miRNA, miR-210, in MBD2^{-/-} adult erythrocytes. MBD2 has been shown to regulate expression of miR-373 in hilar cholangiocarcinoma (Chen et al. 2011). However, miR-210 is so far the only miRNA recognized to be regulated by MBD2 in normal tissues.

We used Ingenuity pathway analysis (IPA) to identify candidate genes for a potential role in regulation of the γ -globin gene in adult erythrocytes. IPA analysis helped to categorize the differentially regulated genes based on their functions and we focused on genes involved in hematological pathways. Four genes were identified as candidates

based on increased expression in MBD2^{-/-} erythrocytes and presence of a CpG Island in its promoter region: Zbtb32, YAF2, Baf57 and ITCH. Of these, only Zbtb32 was validated for its increased expression in the MBD2^{-/-} adult erythrocytes by qPCR. Zbtb32 was further validated for its increased expression in MBD2 knock down mouse erythroleukemia cell line (MEL), CID-dependent β -YAC mouse bone marrow cells and primary adult human CD34⁺ erythroid cells. A high number of false positives may arise due to variability in modifier genes brought about by the mixed background of the mice used in this study. Although this can be a limitation, the advantage of using mixed background mice as compared to inbred mice is that the former better represent the variation present in human populations.

We further validated the functional role of ZBTB32 in human γ -globin gene expression in CID β -YAC mouse bone marrow cells and primary adult human CD34⁺ erythroid cells. Both of these *in vitro* cell systems have an adult β -globin gene expression profile and minimal expression (<1%) of fetal hemoglobin (Blau et al. 2005; Pope et al. 2000). Over-expression of ZBTB32 showed a modest increase in the expression of γ -globin gene (~2-fold), indicating that ZBTB32 may comprise one of the many different ways in which MBD2 regulates the γ -globin gene. Also, as described earlier, the repression of the γ -globin gene in adult erythrocytes is a complex interplay between many different cis-acting sequences and trans-acting factors. Over-expression of a single positive regulator may not be able to offset effects of other repressor proteins present in these cells. At this stage, the mechanism of ZBTB32 mediated activation of the γ -globin gene in adult erythrocytes is conjectural. ZBTB32 has not been shown to be directly involved in the regulation of the γ -globin gene. However, it has been reported to interact with

GATA2 which is known to be involved in γ -globin gene regulation in human erythroleukemia K562 cells (Ikonomi et al. 2000; Tsuzuki and Enver 2002). Study of the expression profile of ZBTB32 in human erythroid progenitor cells displayed high expression levels of ZBTB32 in the progenitor stage when the γ -globin gene expression is also higher, followed by a significant decline in expression of ZBTB32 as well as γ -globin in the differentiated erythroid cells (Dai et al. 2002). This could be consistent with the role of ZBTB32 as a potential activator of the γ -globin gene.

We validated miR-210 expression in MBD2^{-/-} mouse splenic erythrocytes and also carried out functional analyses in CID mouse bone marrow cells by over-expressing miR-210. Our results showed a modest increase in γ -globin gene expression, but a more significant increase in adult β -globin gene expression, suggesting induced erythroid differentiation of these cells caused by miR-210 over-expression. In primary adult human erythroid cells, we were unable to report a consistent increase in miR-210 expression in MBD2 knocked down cells. However, our experimental strategy may be limited by the fact that the differentiated cells used for analysis are a heterogeneous population of erythroid cells. Although these cells are sorted for their expression of both CD71 and CD235a cell surface markers, they may represent several different stages of erythroid differentiation. This could in turn dictate the expression of miR-210 which has been implicated in erythroid maturation of progenitor cells (Bianchi et al. 2009; Bianchi et al. 2012). To address this concern, we could alternately sort different stages of erythroblasts by categorizing the CD71⁺/CD235a⁺ cells in the R1-R5 sub groups, of which R3-R5 sub-groups represent later stages of erythroid differentiation. Recent studies have also introduced band 3 and $\alpha 4$ integrin expression as markers to the

studies of erythropoiesis to isolated specific stages of erythroblasts through flow cytometry (Hu et al. 2013).

Functional analysis of γ -globin gene expression in miR-210 over-expressing cells showed an increase in expression of the fetal γ -globin as well adult β -globin gene compared to the scramble expressing control cells, similar to results observed in mouse bone marrow CID cells. The increase in γ -globin gene expression however was not statistically significant which may be due to the fact that each experiment in primary human erythroid cells is conducted on cells from an individual bone marrow donor. Hence part of the variability in our results represents the variability seen in a mixed human population with various genetic backgrounds. Increasing the sample size of the experiments might bring the increased fold change of γ -globin gene expression in miR-210 over-expressing cells in a statistically significant range.

Thus, taken together, both the murine and the human erythroid cell systems display an effect of miR-210 over-expression on γ - and β -globin gene expression. Several other groups have reported a role of miR-210 in erythroid differentiation. miR-210 expression is increased in a cell culture model which induces γ -globin gene expression and erythroid differentiation (Bianchi et al. 2009). Moreover, miR-210 levels are also elevated during mouse fetal liver erythroid cell differentiation *in vitro* (Kosaka et al. 2008). miR-210 has been shown to be up-regulated in erythroid cells of patients with β -thalassemia and high HbF levels (Bianchi et al. 2009). Hence, miR-210 expression has been correlated with increased γ -globin gene expression and erythroid maturation, however the mechanism has not yet been identified. No known targets of miR-210 have been shown to be involved in erythropoiesis, nor have bioinformatic analyses through

miRNA target prediction algorithms identified any known regulators of β -type globin expression as targets of miR-210.

However, miR-210 is also predominantly a hypoxia induced miRNA and hypoxia response of erythroid cells has been shown to induce erythroid differentiation in part through miR-210 (Sarakul et al. 2013). Thus miR-210 may be an indirect regulator of β -type globin gene expression through pathways that involve induced erythroid differentiation or hypoxia mediated stress erythropoiesis. The role of miR-210 in erythroid differentiation could be further evaluated in conditions of ineffective erythropoiesis caused by a maturation blockade example congenital dyserythropoietic anemia and β -thalassemias. It would be worthwhile to investigate whether over-expression of miR-210 can restore normal erythroid differentiation of immature erythroid cells.

In conclusion, our studies have identified two novel targets of MBD2 in the adult erythroid system: ZBTB32 and miR-210. We tested the function of both of these factors individually in adult mouse bone marrow cells and primary adult human erythroid cells. ZBTB32 and miR-210 modestly contribute to the induction of γ -globin gene expression in MBD2 deficient erythroid cells. Given the current literature and our data, we believe that ZBTB32 and miR-210 function in independent pathways to induce γ -globin gene expression. Co-expression of ZBTB32 and miR-210 in primary adult human erythroid cells would be anticipated to produce at least an additive increase in γ -globin gene expression. Optimization of the lentiviral transduction method will be required to perform co-transduction of both downstream effectors of MBD2 in primary erythroid cells.

CHAPTER 4: Summary, Future Directions and Perspectives

Sickle cell disease and β -thalassemia are among the most common monogenic disorders worldwide. Sickle cell disease is characterized by “sickle” shaped red blood cells caused by polymerization of deoxygenated HbS. Sickling of red cells leads to vaso-occlusion and ischemia in small vessels causing painful crises and multiple organ damage. Hydroxyurea is the only therapeutic option for patients with sickle cell disease, however, its effects are often variable. β -thalassemia is characterized by reduced synthesis of the β -globin chains. Cooley’s anemia or β -thalassemia major is the most severe form of the disease and blood transfusions are the standard course of treatment for these patients. Iron chelation therapy is required to deplete the iron deposits in organs due to regular blood transfusions. Reactivation of fetal hemoglobin in patients with sickle cell disease and β -thalassemia can alleviate symptoms of the disease. Understanding the regulators involved in the fetal-to-adult hemoglobin switch can provide new avenues for search of molecular drug targets with higher efficacy and lower side-effects.

Research over the past 6 decades has added valuable information to the understanding of the hemoglobin switching process. DNA methylation has emerged as

an important regulator of embryonic and fetal globin gene expression (Gnanapragasam et al. 2011b; Rupon et al. 2006b; Rupon et al. 2011b). The MBD2-NuRD co-repressor complex directly represses the expression of the chicken ρ -globin gene in avian adult erythroid cells whereas it indirectly regulates γ -globin gene expression in human cells (Rupon et al. 2006b). MBD2 knockout mice carrying the human β -globin gene locus (β -YAC) display higher than normal expression of the human fetal γ -globin gene in adult erythroid cells (Rupon et al. 2006b). Primary adult human erythroid cells deficient for MBD2 also exhibit significant reactivation of the γ -globin gene (Gnanapragasam et al. 2011b). Partial knockdown of the Mi2 β subunit of the NuRD complex has an even larger effect on de-repression of the fetal γ -globin gene, through additionally down-regulating BCL11A and KLF1 proteins which are repressors of the γ -globin gene in adult erythroid cells. KLF1 and BCL11A have been shown to play a major role in γ -globin gene expression. KLF1 is a positive regulator of the adult β -globin gene and the BCL11A gene. BCL11A binds to the β -globin gene in the γ - δ intergenic region and brings about repression of the γ -globin gene, through its interaction with co-repressor complexes that include the MBD3-NuRD complex, LSD1/CoREST complex, Sin3A, NCoR/SMRT, and DNMT1 (Bauer, Kamran, Orkin 2012). The NuRD complex with its various binding partners has been identified as an important network point in regulation of fetal γ -globin gene expression. However, further work is required to elucidate the assembly and functional organization of the various subunits of the NuRD complex.

In our studies presented in Chapter 2, we have pursued the structural and functional characterization of a previously uncharacterized domain of MBD2, MBD2_{IDR}. This domain of MBD2 lies between the N-terminus methyl-binding domain (MBD) and

the C-terminus coiled-coil (CC) interaction domain. MBD2, as a whole protein, has been difficult to express *in vitro*, which limits its biophysical characterization. Therefore, we strategized to characterize the structure of MBD2 domain by domain. We have previously described the structure of chicken MBD2, which is 96% identical to human MBD2 in protein sequence (Scarsdale et al. 2011b). We also studied the coiled-coil domain of MBD2 and described its interaction with the conserved region 1 (CR1) of p66 α component of the NuRD complex (Gnanapragasam et al. 2011b). In collaboration with the research group lead by Dr. David Williams, we conducted NMR studies that identified the MBD2_{IDR} to be an intrinsically disordered region. Our results outline that despite the inherent disorder, MBD2_{IDR} is involved in two important functions: 1) modifying the kinetics of DNA binding and MBD2's affinity for methylated DNA through ionic interactions mediated by the positively charged N-terminal region of MBD2_{IDR} and 2) mediating interaction of MBD2 with the histone deacetylase core of the NuRD complex. In pursuit of refining the former function, we mapped two contiguous amino acid residues, Arg²⁸⁶ and Leu²⁸⁷ to form a critical region of interaction between MBD2_{IDR} and the histone deacetylase core. Mutation of these key residues in full length MBD2 protein significantly disrupted its interaction with RbAp48, HDAC2 and MTA2 subunits. To test the role of these critical amino acids in the repressive activity of MBD2, we mutated the two contiguous amino acids in the full length MBD2 protein and investigated its function *in vivo*. We designed an *in vivo* gene repression assay in a breast cancer cell line such that a tumor suppressor gene *PRSS8* silenced by MBD2 mediated repression was used as a reporter of MBD2 function (Chen et al. 2001). To this end, endogenous MBD2 was stably knocked down in MDA-MB-435 breast cancer

cells and wild type as well as double mutant MBD2 was re-introduced in the cells through a lentiviral expression system. In this system, we could attain about 50% knockdown of endogenous MBD2 protein and a comparable level of enforced expression of wild-type and double mutant MBD2 protein. We observed that add back of wild-type MBD2 repressed the *PRSS8* gene beyond the baseline expression level set by the control cells, whereas the double mutant MBD2 failed to do so. This resulted in a 2-fold higher expression of *PRSS8* in the double mutant MBD2 expressing cells as compared to the wild-type MBD2 expressing cells. This observation reflected an impaired repressor function of the MBD2 protein due to its disrupted interaction with the histone deacetylase core of the NuRD complex.

For de-repression of a gene silenced by MBD2, at least a 75% knockdown of the MBD2 protein is required (Amaya et al. 2013; Gnanapragasam et al. 2011b). Since we were able to achieve only a 50% knockdown of endogenous MBD2 in our assay due to technical limitations, we did not observe a complete de-repression of the gene *PRSS8*. Also, genes known to be repressed by MBD2 have been identified either in MBD2 knockout mice or through knockdown of MBD2 in various cell systems. Functionally, this entails loss of function of all the domains of MBD2. While our assay interferes with the interaction of MBD2 with the histone deacetylase core, the double mutant MBD2 still carries a functional MBD and CC domain. These domains will allow for the mutant MBD2 to occupy sites of CpG methylation as well as recruit the p66 α and Mi2 β proteins, which bring about chromatin remodeling activities at the respective locus. A combined effect of both of these events may prevent complete de-repression of the target gene of MBD2, even in the absence of interaction with the histone deacetylase core

components. Thus, complete de-repression of a methylated target gene may require decoupling of both histone deacetylase and chromatin remodeling activities from the MBD2-NuRD complex. This hypothesis will be tested by introducing mutations in the key residues of coiled-coil domain of MBD2, which are critical for its interaction with p66 α and Mi2 β proteins, in addition to the key residue mutations in the IDR.

An ideal experimental design to study the function of mutant MBD2 protein will be to introduce these critical mutations in the endogenous MBD2 gene in a cell line of interest, followed by testing gene expression changes. Although this idea would have been perceived as too far-fetched in the past, the field of genomic editing has been revolutionized with the advent of the CRISPR/Cas9 technology. The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas9 system has been adapted from the prokaryotic immune system which allows them resistance against foreign genetic elements such as plasmids and phages in the form of adaptive immunity. This technology has enabled generation of gene knockout cell lines and transgenic animals at an unprecedented speed and greatly empowered genetic research (Maddalo et al. 2014; Platt et al. 2014; Swiech et al. 2014). A stunning application of the CRISPR/Cas9 system was recently reported in the correction of HBB mutations in β -thalassemia patient-derived induced pluripotent stem cells (iPSCs). These “corrected” iPSCs showed restored expression of the HBB gene post the gene editing when differentiated into erythroblasts *in vitro* (Xie et al. 2014).

In the context of MBD2 function, the CRISPR/Cas9 system could be employed to introduce the critical amino acid mutations in the endogenous MBD2 gene in a cell line such as MDA-MB-435 where the function of MBD2 has been well characterized (Mian et

al. 2011). Introducing mutations at critical sites in the MBD2_{IDR} or the coiled-coil domain or in combination will enable us to dissect the specific role(s) of the histone deacetylase core and the chromatin remodeling components in MBD2 mediated repression of methylated target genes. Gene expression microarray analysis will be employed in such a system to understand regulation of specific sets of genes by the histone deacetylase core or the chromatin remodeling components of NuRD. Such an experiment where the interaction of MBD2 with both histone deacetylase and chromatin remodeling components of the NuRD complex is disrupted will also shed light on the function of MBD2 as an independent methylation-mediated transcriptional repressor.

The fetal γ -globin gene has been extensively studied by our lab as a target of the MBD2-NuRD co-repressor complex in murine as well as human systems. Studying the function of critical residues of the MBD2_{IDR} in the context of developmental silencing of the γ -globin gene by MBD2 is an ongoing project in the lab. Knockdown of MBD2 in adult human erythroid cells causes a robust de-repression of the fetal γ -globin gene, which provides a bigger dynamic range to study the function of mutant MBD2 protein in these cells. A functional assay similar to that in the MDA-MB-435 breast cancer cells is being pursued; endogenous MBD2 is knocked down followed by introduction of wild-type or double mutant MBD2 in sequential lentiviral transductions. Preliminary studies in CD34⁺ adult erythroid cells have generated encouraging results. However, the feasibility of such an experiment is significantly limited by the low infection efficiency of primary cells and reduced viability of the cells due to double lentiviral infections. If the design of experiment works successfully in CD34⁺ cells, future studies will be planned to introduce critical residue mutations in the coiled-coil domain of MBD2 in combination with the

mutant MBD2_{IDR} and study its effect on γ -globin gene expression. We anticipate that expression of MBD2 carrying mutations in both IDR and CC domains may lead to de-repression of the γ -globin gene at levels comparable to those in MBD2 knockdown CD34⁺ cells.

A long standing goal of the lab has been to inhibit the repressor activity of the MBD2/NuRD complex. The repressor activity of MBD2 is mediated through interactions with other co-repressor proteins, suggesting that disruption of these protein-protein interactions can serve as a potential avenue to inhibit its function. Intrinsically disordered proteins have recently been identified as attractive drug targets for inhibiting protein-protein interactions because of their over-representation in disease pathways and characteristic high specificity-low affinity interaction networks (Metallo 2010). However, leads to such inhibitors have been limited by the fact that protein-protein interactions involving unstructured regions often span large flat contact areas and lack short binding pockets which are ideal for drug targets. Alternately, inhibitors targeting disordered regions can be designed to prevent interaction between the unstructured region and its structured ligand/interaction partner. Recent work has indicated that protein-protein interactions based on disorder-to-order transition of an intrinsically disordered protein can be disrupted (Dawson et al. 2003; Klein and Vassilev 2004; Stauffer 2007). Few studies of intrinsically disordered proteins have led to the discovery of successful small molecule inhibitors. Examples include, inhibitor of interaction between tumor suppressor p53 and its binding partner Mdm2; Mdm2 by binding to an intrinsically disordered region of p53 targets p53 for ubiquitination and causes it to be transported out of the nucleus. Small molecule inhibitors called "nutlins" have been

identified which associate with Mdm2 and thus prevent its interaction with p53 (Klein and Vassilev 2004). The interaction between oncoprotein c-myc and its binding partner Max has been inhibited by small molecules which bind c-myc and stabilize its disordered conformation. This in turn prevents its interaction with Max protein (Hammoudeh et al. 2009). Characterization of the MBD2_{IDR} adds valuable insight to understanding interactions within the multi-protein NuRD complex. Identification of the two contiguous amino acid residues critical for the interaction between MBD2_{IDR} and the histone deacetylase core subunits provides an exciting avenue to target these protein-protein interactions within the NuRD complex through small molecule/peptide inhibitors.

An important step towards identifying small molecule/peptide inhibitors of the MBD2-NuRD complex will be to better define the structure of MBD2 protein. We have previously reported the structure of the chicken MBD bound to methylated DNA sequence (Scarsdale et al. 2011a) and the helical coiled-coil interaction between CC domain of MBD2 and the conserved region 1 (CR1) of p66 α protein (Gnanapragasam et al. 2011b). We have now also characterized the structure of the MBD2_{IDR} as an intrinsically disordered region in isolation and in the context of its interaction with methylated DNA. However, many intrinsically disordered regions often fold into a well-defined conformation in the presence of their binding partners (Tompa 2002). This refolding of the region is guided by principles of thermodynamics that dictate that binding will stabilize and strengthen binding interactions. Our studies suggest an interaction between MBD2_{IDR} and RbAp48, HDAC2 and MTA2 proteins. To understand the structure of the MBD2_{IDR} in presence of its binding partners, MBD2_{IDR} will need to be co-expressed along with RbAp/HDAC/MTA proteins and purified as a complex. NMR

spectroscopy of this complex will better define the structure of MBD2_{IDR} in the presence of its functional interactions.

MBD2 and MBD3 are highly similar proteins and are both involved in regulatory functions through interactions with the NuRD complex. MBD2_{IDR} and the homologous region of MBD3 are 79% identical in protein sequence and this amino acid sequence can be a strong determinant of the protein structure. To extend the information gained through our work on MBD2, we will hypothesize that the region of MBD3 homologous to MBD2_{IDR} will also be intrinsically disordered and be involved in recruiting the histone deacetylase core of the NuRD complex. In the absence of the functional methyl-binding domain, the positively charged N-terminal region of MBD3, which is homologous to MBD2_{IDR}, may facilitate non-specific ionic interactions of MBD3 protein with the DNA. The MBD3/NuRD complex is involved in many important biological processes, in particular embryonic development as well as reprogramming of iPSCs and pluripotency of embryonic stem cells (dos Santos et al. 2014; Hendrich et al. 2001; Rais et al. 2013).

In the work detailed in Chapter 3, we investigated the mechanism of MBD2 mediated repression of the fetal γ -globin gene. Through microarray analysis of protein-coding genes and miRNAs differentially regulated in MBD2^{-/-} mice, we identified two new targets of MBD2: ZBTB32 and miR-210. We successfully validated the function of ZBTB32 in various murine and human erythroid cell systems, wherein over-expression of ZBTB32 modestly induced the expression of the fetal γ -globin gene in adult erythroid cells. Similarly, enforced expression of miR-210 in adult erythroid cells displayed an effect on increasing the expression of the fetal γ -globin gene. Taking into account our current understanding of the biological functions of ZBTB32 and miR-210, we believe

that both of these factors regulate γ -globin gene expression through independent pathways. Simultaneous enforced expression of ZBTB32 and miR-210 will be hypothesized to have an additive effect on the induction of the fetal γ -globin gene in adult erythroid cells. The execution of this experiment has been limited by technical difficulties involved in maintaining optimum transduction efficiency and viability of CD34⁺ primary human erythroid cells transduced with two lentiviral vectors. Two approaches will be used to overcome this limitation: 1. Improving lentiviral transduction efficiency 2. Use of a lentiviral expression vector allowing high expression of the gene of interest followed by selection of transduced cells through a method other than FACS. Rapamycin, an inhibitor of mammalian target of rapamycin (mTOR) complexes, can improve lentiviral transduction of hematopoietic stem cells by increasing cytoplasmic entry of lentiviruses by endocytic events (Wang et al. 2014). Through careful design of experiments to ensure minimum cytotoxicity, introduction of rapamycin in the lentiviral transduction mixture will be used to improve transduction efficiency of CD34⁺ cells. For the second approach, lentiviral expression vectors that carry a cell surface antigen gene instead of GFP or a drug selection marker will enable selection of transduced cells through magnetic bead sorting, a technique that is easier to perform and does not affect viability of the transduced CD34⁺ cells.

Given the modest fold changes in γ -globin gene expression by over-expression of either ZBTB32 or miR-210, we anticipate that there exist other effectors of MBD2, which also contribute to its repressive effect. Long non-coding RNAs have gained increasing significance in transcriptional regulation and various biological processes (Wilusz, Sunwoo, Spector 2009). Regulation of non-coding RNAs by MBD2 is an unexplored

area, which holds great potential and will enable identification of yet unknown mechanisms of transcriptional regulation by MBD2.

Perspectives:

Several potential molecular targets for reactivation of fetal hemoglobin in adult cells have been identified. Of these, BCL11A is one of the prime candidates that have been pursued due to its impressive therapeutic potential as observed in *Bcl11a* knockout sickle cell disease transgenic mice. These transgenic mice show a dramatic elevation of fetal hemoglobin expression in adult erythroid cells and an almost curative effect on sickle cell disease pathophysiology. Though initial studies were encouraging, studies in primary human adult erythroid cells showed a relatively less dominant effect of BCL11A on γ -globin gene expression. This result arises from the fact that regulation of the transgenic human γ -globin gene in mice has differences in regulation in mice compared to humans due to the absence of a fetal β -type globin gene in mice. Nevertheless, targeting BCL11A has been a challenge since its exact mechanism of action is not very clear and it does not harbor an enzymatically active functional domain, which could be targeted through small molecule inhibitors. BCL11A may function through a multi-protein complex and targeting the protein-protein interactions may be imperative to inhibit its activity as a repressor of fetal γ -globin gene. Besides, given the role of BCL11A in normal lymphoid development, systemic effects of pharmacologically targeting BCL11A remains a concern.

The MBD2-NuRD co-repressor complex serves as an attractive epigenetic target for induction of fetal hemoglobin in adult erythrocytes of patients with β -hemoglobinopathies. In contrast to other known regulators of fetal hemoglobin, MBD2 stands out as a target of clinical significance due to its minimal involvement in processes of normal development and erythroid differentiation. MBD2^{-/-} knockout mice are viable and fertile and display a minimally abnormal phenotype. MBD2 deficient primary human erythroid cells also show normal cellular proliferation and differentiation. This underscores the therapeutic potential of targeting MBD2 since it would not affect normal physiological processes. However, the road to developing inhibitors of the MBD2-NuRD complex has been wrought with many obstacles.

The first challenge is the need for a robust *in vitro* system that can provide a dynamic read-out for the repressor function of the MBD2-NuRD complex. The most well defined role of MBD2 is in developmental silencing of the fetal γ -globin gene. Many *in vitro* systems are available today to study the hemoglobin switch in murine and human adult erythroid cells, but these systems have their limitations. β -YAC mouse bone marrow CID cells primarily express adult mouse and human globin genes; the embryonic and fetal globin genes are very tightly repressed in these cells (Blau et al. 2005). Although MBD2 knockdown increases γ -globin gene expression in these cells, the tight transcriptional repression provides a small dynamic range to study the function of MBD2. Recently, CID cells engineered with a γ -globin promoter-firefly luciferase (γ -luc) and β -globin promoter-Renilla luciferase (β -luc) fusions from γ -luc β -luc β -YAC transgenic mice were developed for high-throughput screening of inducers of fetal γ -globin gene (Peterson et al. 2014). This system would allow the identification of

compounds that specifically induce the firefly luciferase (γ -globin), but not Renilla luciferase (β -globin). A similar system would be required to screen large number of compounds that inhibit the function of MBD2 and activate the γ -globin gene promoter. Primary adult human erythroid cells offer the advantage of a large dynamic range of γ -globin gene repression by MBD2; however the technical difficulties involved in genetic manipulation of these cells are a challenge. Additionally, although primary adult human erythroid cells have become a valuable and standard system to identify and characterize regulators of the hemoglobin switch, there is lack of a consistency in the extent of erythroid differentiation and relative $\gamma/(\gamma+\beta)$ expression level achieved in studies performed by different groups. Having a comparable quantitative end-point for experiments involving primary human erythroid cells will add value to this system as well as aide comparison of effects of various factors on the hemoglobin switch.

The feasibility of targeting the function of MBD2 is another challenge. Unlike most “druggable” targets that possess enzymatic activities or well-defined binding pockets for ligands that can be bound by small molecules, MBD2 functions as a repressor in co-operation with the multiple proteins of NuRD complex. Protein-protein interactions and protein-DNA interactions are challenging drug targets. However, recent advances have been made in developing covalently stapled peptides and small molecules, which can disrupt protein-protein interactions. A stabilized alpha-helical peptide was synthesized to interfere with the assembly of the active NOTCH transcriptional complex (Moellering et al. 2009); a small-molecule inhibitor that disrupted the BCL6/corepressor complex by binding at the binding groove of BCL6 BTB domain (Cerchiatti et al. 2010), and a hydrocarbon-stapled peptide modeled after the BIM BH3 helix which broadly targeted all

BCL2 family proteins (LaBelle et al. 2012) are all examples of successful applications of therapeutic targeting of protein-protein interactions. Our studies have identified two critical interaction surfaces on the MBD2 protein, which can be targeted to disrupt binding with the histone deacetylase and chromatin remodeling components of the NuRD complex. In the future, this information will be employed to test a library of small molecule inhibitors or to design synthetic peptides that can specifically target the critical regions of MBD2. Targeting the function of MBD2 will have therapeutic value in treatment of β -hemoglobinopathies as well as various cancers where tumor suppressor genes are silenced by MBD2.

Literature Cited

- Abdulmalik O, Safo MK, Chen Q, Yang J, Brugnara C, Ohene-Frempong K, Abraham DJ, Asakura T. 2005. 5-hydroxymethyl-2-furfural modifies intracellular sickle haemoglobin and inhibits sickling of red blood cells. *Br J Haematol* 128(4):552-61.
- Adams VH, McBryant SJ, Wade PA, Woodcock CL, Hansen JC. 2007. Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *J Biol Chem* 282(20):15057-64.
- Alqarni SS, Murthy A, Zhang W, Przewloka MR, Silva AP, Watson AA, Lejon S, Pei XY, Smits AH, Kloet SL, et al. 2014. Insight into the architecture of the NuRD complex: STRUCTURE OF THE RbAp48-MTA1 SUBCOMPLEX. *J Biol Chem* 289(32):21844-55.
- Amaya M, Desai M, Gnanapragasam MN, Wang SZ, Zu Zhu S, Williams DC, Jr, Ginder GD. 2013. Mi2beta-mediated silencing of the fetal gamma-globin gene in adult erythroid cells. *Blood* 121(17):3493-501.
- Angelucci E. 2010. Hematopoietic stem cell transplantation in thalassemia. *Hematology Am Soc Hematol Educ Program* 2010:456-62.
- Ardekani AM and Naeini MM. 2010. The role of MicroRNAs in human diseases. *Avicenna J Med Biotechnol* 2(4):161-79.
- Atweh G and Fathallah H. 2010. Pharmacologic induction of fetal hemoglobin production. *Hematol Oncol Clin North Am* 24(6):1131-44.
- Bader S, Walker M, Harrison D. 2000. Most microsatellite unstable sporadic colorectal carcinomas carry MBD4 mutations. *Br J Cancer* 83(12):1646-9.
- Barr H, Hermann A, Berger J, Tsai HH, Adie K, Prokhortchouk A, Hendrich B, Bird A. 2007. Mbd2 contributes to DNA methylation-directed repression of the *xist* gene. *Mol Cell Biol* 27(10):3750-7.
- Baubec T, Ivanek R, Lienert F, Schubeler D. 2013. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* 153(2):480-92.
- Bauer DE, Kamran SC, Orkin SH. 2012. Reawakening fetal hemoglobin: Prospects for new therapies for the beta-globin disorders. *Blood* 120(15):2945-53.
- Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C, Shao Z, Canver MC, Smith EC, Pinello L, et al. 2013. An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* 342(6155):253-7.
- Behringer RR, Ryan TM, Palmiter RD, Brinster RL, Townes TM. 1990. Human gamma- to beta-globin gene switching in transgenic mice. *Genes Dev* 4(3):380-9.
- Bell SG. 1999. An introduction to hemoglobin physiology. *Neonatal Netw* 18(2):9-15.
- Bender MA, Bulger M, Close J, 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):387-93.
- Berg T, Cohen SB, Desharnais J, Sonderegger C, Maslyar DJ, Goldberg J, Boger DL, Vogt PK. 2002. Small-molecule antagonists of myc/max dimerization inhibit myc-induced transformation of chicken embryo fibroblasts. *Proc Natl Acad Sci U S A* 99(6):3830-5.
- Berger J, Sansom O, Clarke A, Bird A. 2007. MBD2 is required for correct spatial gene expression in the gut. *Mol Cell Biol* 27(11):4049-57.
- Bianchi N, Zuccato C, Lampronti I, Borgatti M, Gambari R. 2009. Expression of miR-210 during erythroid differentiation and induction of gamma-globin gene expression. *BMB Rep* 42(8):493-9.
- Bianchi N, Zuccato C, Finotti A, Lampronti I, Borgatti M, Gambari R. 2012. Involvement of miRNA in erythroid differentiation. *Epigenomics* 4(1):51-65.

- Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, Frutiger S, Hochstrasser D. 1993. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* 14(10):1023-31.
- Blau CA, Barbas CF, 3rd, Bomhoff AL, Neades R, Yan J, Navas PA, Peterson KR. 2005. {Gamma}-globin gene expression in chemical inducer of dimerization (CID)-dependent multipotential cells established from human {beta}-globin locus yeast artificial chromosome ({beta}-YAC) transgenic mice. *J Biol Chem* 280(44):36642-7.
- Borg J, Papadopoulos P, Georgitsi M, Gutierrez L, Grech G, Fanis P, Phylactides M, Verkerk AJ, van der Spek PJ, Scerri CA, et al. 2010. Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nat Genet* 42(9):801-5.
- Boyer SH, Belding TK, Margolet L, Noyes AN. 1975. Fetal hemoglobin restriction to a few erythrocytes (F cells) in normal human adults. *Science* 188(4186):361-3.
- Brackertz M, Gong Z, Leers J, Renkawitz R. 2006. p66alpha and p66beta of the mi-2/NuRD complex mediate MBD2 and histone interaction. *Nucleic Acids Res* 34(2):397-406.
- Bradner JE, Mak R, Tanguturi SK, Mazitschek R, Haggarty SJ, Ross K, Chang CY, Bosco J, West N, Morse E, et al. 2010. Chemical genetic strategy identifies histone deacetylase 1 (HDAC1) and HDAC2 as therapeutic targets in sickle cell disease. *Proc Natl Acad Sci U S A* 107(28):12617-22.
- Cai M, Williams DC, Jr, Wang G, Lee BR, Peterkofsky A, Clore GM. 2003. Solution structure of the phosphoryl transfer complex between the signal-transducing protein IIA_{glucose} and the cytoplasmic domain of the glucose transporter IICB_{glucose} of the escherichia coli glucose phosphotransferase system. *J Biol Chem* 278(27):25191-206.
- Campbell AD, Cui S, Shi L, Urbonya R, Mathias A, Bradley K, Bonsu KO, Douglas RR, Halford B, Schmidt L, et al. 2011. Forced TR2/TR4 expression in sickle cell disease mice confers enhanced fetal hemoglobin synthesis and alleviated disease phenotypes. *Proc Natl Acad Sci U S A* 108(46):18808-13.
- Cavazzana-Calvo M, Payen E, Negre O, Wang G, Hehir K, Fusil F, Down J, Denaro M, Brady T, Westerman K, et al. 2010. Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature* 467(7313):318-22.
- Cerchietti LC, Ghetu AF, Zhu X, Da Silva GF, Zhong S, Matthews M, Bunting KL, Polo JM, Fares C, Arrowsmith CH, et al. 2010. A small-molecule inhibitor of BCL6 kills DLBCL cells in vitro and in vivo. *Cancer Cell* 17(4):400-11.
- Chang CJ and Bouhassira EE. 2012. Zinc-finger nuclease-mediated correction of alpha-thalassemia in iPS cells. *Blood* 120(19):3906-14.
- Chang KH, Nelson AM, Cao H, Wang L, Nakamoto B, Ware CB, Papayannopoulou T. 2006. Definitive-like erythroid cells derived from human embryonic stem cells coexpress high levels of embryonic and fetal globins with little or no adult globin. *Blood* 108(5):1515-23.
- Charache S, Dover G, Smith K, Talbot CC, Jr, Moyer M, Boyer S. 1983. Treatment of sickle cell anemia with 5-azacytidine results in increased fetal hemoglobin production and is associated with nonrandom hypomethylation of DNA around the gamma-delta-beta-globin gene complex. *Proc Natl Acad Sci U S A* 80(15):4842-6.
- Charache S, Terrin ML, Moore RD, Dover GJ, Barton FB, Eckert SV, McMahon RP, Bonds DR. 1995. Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. investigators of the multicenter study of hydroxyurea in sickle cell anemia. *N Engl J Med* 332(20):1317-22.
- Chatagnon A, Perriaud L, Nazaret N, Croze S, Benhattar J, Lachuer J, Dante R. 2011. Preferential binding of the methyl-CpG binding domain protein 2 at methylated transcriptional start site regions. *Epigenetics* 6(11):1295-307.
- Chen LM and Chai KX. 2002. Prostaticin serine protease inhibits breast cancer invasiveness and is transcriptionally regulated by promoter DNA methylation. *Int J Cancer* 97(3):323-9.

- Chen LM, Hodge GB, Guarda LA, Welch JL, Greenberg NM, Chai KX. 2001. Down-regulation of prostatic serine protease: A potential invasion suppressor in prostate cancer. *Prostate* 48(2):93-103.
- Chen Y, Gao W, Luo J, Tian R, Sun H, Zou S. 2011. Methyl-CpG binding protein MBD2 is implicated in methylation-mediated suppression of miR-373 in hilar cholangiocarcinoma. *Oncol Rep* 25(2):443-51.
- Chen Z, Luo HY, Basran RK, Hsu TH, Mang DW, Nuntakarn L, Rosenfield CG, Patrinos GP, Hardison RC, Steinberg MH, et al. 2008. A T-to-G transversion at nucleotide -567 upstream of HBG2 in a GATA-1 binding motif is associated with elevated hemoglobin F. *Mol Cell Biol* 28(13):4386-93.
- Collis P, Antoniou M, Grosveld F. 1990. Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *EMBO J* 9(1):233-40.
- Cornilescu G, Delaglio F, Bax A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13(3):289-302.
- Costa FC, Fedosyuk H, Chazelle AM, Neades RY, Peterson KR. 2012. Mi2beta is required for gamma-globin gene silencing: Temporal assembly of a GATA-1-FOG-1-Mi2 repressor complex in beta-YAC transgenic mice. *PLoS Genet* 8(12):e1003155.
- Cramer JM, Scarsdale JN, Walavalkar NM, Buchwald WA, Ginder GD, Williams DC, Jr. 2014. Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNA. *J Biol Chem* 289(3):1294-302.
- Cui S, Kolodziej KE, Obara N, Amaral-Psarris A, Demmers J, Shi L, Engel JD, Grosveld F, Strouboulis J, Tanabe O. 2011. Nuclear receptors TR2 and TR4 recruit multiple epigenetic transcriptional corepressors that associate specifically with the embryonic beta-type globin promoters in differentiated adult erythroid cells. *Mol Cell Biol* 31(16):3298-311.
- Dai MS, Chevallier N, Stone S, Heinrich MC, McConnell M, Reuter T, Broxmeyer HE, Licht JD, Lu L, Hoatlin ME. 2002. The effects of the fanconi anemia zinc finger (FAZF) on cell cycle, apoptosis, and proliferation are differentiation stage-specific. *J Biol Chem* 277(29):26327-34.
- David G. Nathan, David Ginsburg, Stuart H. Orkin, A. Thomas Look. 2003. **Nathan and oski's hematology of infancy and childhood.** 0721693172.
- Dawson R, Muller L, Dehner A, Klein C, Kessler H, Buchner J. 2003. The N-terminal domain of p53 is natively unfolded. *J Mol Biol* 332(5):1131-41.
- Deaton AM and Bird A. 2011. CpG islands and the regulation of transcription. *Genes Dev* 25(10):1010-22.
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277-93.
- Dijon M, Bardin F, Murati A, Batoz M, Chabannon C, Tonnelle C. 2008. The role of ikaros in human erythroid differentiation. *Blood* 111(3):1138-46.
- Dillon N and Grosveld F. 1991. Human gamma-globin genes: silenced independently of other genes in the beta-globin locus. *Nature* 350(6315):252-4.
- dos Santos RL, Tosti L, Radzishewska A, Caballero IM, Kaji K, Hendrich B, Silva JC. 2014. MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. *Cell Stem Cell* 15(1):102-10.
- Dover GJ and Charache S. 1992. Hydroxyurea induction of fetal hemoglobin synthesis in sickle-cell disease. *Semin Oncol* 19(3 Suppl 9):61-6.
- Dumitriu B, Patrick MR, Petschek JP, Cherukuri S, Klingmuller U, Fox PL, Lefebvre V. 2006. Sox6 cell-autonomously stimulates erythroid cell survival, proliferation, and terminal maturation and is thereby an important enhancer of definitive erythropoiesis during mouse development. *Blood* 108(4):1198-207.
- Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. 2005. Flexible nets. the roles of intrinsic disorder in protein interaction networks. *FEBS J* 272(20):5129-48.

- EDINGTON GM and LEHMANN H. 1955. Expression of the sickle-cell gene in africa. *Br Med J* 1(4925):1308-11.
- Eisen JA, Sweder KS, Hanawalt PC. 1995. Evolution of the SNF2 family of proteins: Subfamilies with distinct sequences and functions. *Nucleic Acids Res* 23(14):2715-23.
- Ellis J, Tan-Un KC, Harper A, Michalovich D, Yannoutsos N, Philipson S, 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):562-8.
- Enver T, Raich N, Ebens AJ, Papayannopoulou T, Costantini F, Stamatoyannopoulos G. 1990. Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature* 344(6264):309-13.
- Esteghamat F, Gillemans N, Bilic I, van den Akker E, Cantu I, van Gent T, Klingmuller U, van Lom K, von Lindern M, Grosveld F, et al. 2013. Erythropoiesis and globin switching in compound Klf1::Bcl11a mutant mice. *Blood* 121(13):2553-62.
- Esteller M. 2008. Epigenetics in cancer. *N Engl J Med* 358(11):1148-59.
- Evans T and Felsenfeld G. 1989. The erythroid-specific transcription factor Eryf1: A new finger protein. *Cell* 58(5):877-85.
- Feng Q and Zhang Y. 2001. The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes Dev* 15(7):827-32.
- Fibach E and Prus E. 2005. Differentiation of human erythroid cells in culture. *Curr Protoc Immunol* Chapter 22:Unit 22F.7.
- Field JJ, Lin G, Okam MM, Majerus E, Keefer J, Onyekwere O, Ross A, Campigotto F, Neuberger D, Linden J, et al. 2013. Sickle cell vaso-occlusion causes activation of iNKT cells that is decreased by the adenosine A2A receptor agonist regadenoson. *Blood* 121(17):3329-34.
- Filion GJ, Zhenilo S, Salozhin S, Yamada D, Prokhortchouk E, Defossez PA. 2006. A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol Cell Biol* 26(1):169-81.
- Fraga MF, Ballestar E, Montoya G, Taysavang P, Wade PA, Esteller M. 2003. The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res* 31(6):1765-74.
- Fucharoen S, Inati A, Siritanaratku N, Thein SL, Wargin WC, Koussa S, Taher A, Chaneim N, Boosalis M, Berenson R, et al. 2013. A randomized phase I/II trial of HQK-1001, an oral fetal globin gene inducer, in beta-thalassaemia intermedia and HbE/beta-thalassaemia. *Br J Haematol* 161(4):587-93.
- Fujita N, Watanabe S, Ichimura T, Ohkuma Y, Chiba T, Saya H, Nakao M. 2003. MCAF mediates MBD1-dependent transcriptional repression. *Mol Cell Biol* 23(8):2834-43.
- Fujiwara Y, Browne CP, Cunniff K, Goff SC, Orkin SH. 1996. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A* 93(22):12355-8.
- Fuxreiter M. 2012. Fuzziness: Linking regulation to protein dynamics. *Mol Biosyst* 8(1):168-77.
- Fuxreiter M and Tompa P. 2012. Fuzzy complexes: A more stochastic view of protein function. *Adv Exp Med Biol* 725:1-14.
- Gallienne AE, Dreau HM, Schuh A, Old JM, Henderson S. 2012. Ten novel mutations in the erythroid transcription factor KLF1 gene associated with increased fetal hemoglobin levels in adults. *Haematologica* 97(3):340-3.
- Ghosh RP, Nikitina T, Horowitz-Scherer RA, Gierasch LM, Uversky VN, Hite K, Hansen JC, Woodcock CL. 2010. Unique physical properties and interactions of the domains of methylated DNA binding protein 2. *Biochemistry* 49(20):4395-410.
- Ginder GD, Gnanapragasam MN, Mian OY. 2008. The role of the epigenetic signal, DNA methylation, in gene regulation during erythroid development. *Curr Top Dev Biol* 82:85-116.

- Gnanapragasam MN, Scarsdale JN, Amaya ML, Webb HD, Desai MA, Walavalkar NM, Wang SZ, Zu Zhu S, Ginder GD, Williams DC, Jr. 2011a. p66Alpha-MBD2 coiled-coil interaction and recruitment of mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc Natl Acad Sci U S A* 108(18):7487-92.
- Gnanapragasam MN, Scarsdale JN, Amaya ML, Webb HD, Desai MA, Walavalkar NM, Wang SZ, Zu Zhu S, Ginder GD, Williams DC, Jr. 2011b. p66Alpha-MBD2 coiled-coil interaction and recruitment of mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc Natl Acad Sci U S A* 108(18):7487-92.
- Grosveld F, van Assendelft GB, Greaves DR, Kollias G. 1987. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* 51(6):975-85.
- Guy J, Hendrich B, Holmes M, Martin JE, Bird A. 2001. A mouse *Mecp2*-null mutation causes neurological symptoms that mimic rett syndrome. *Nat Genet* 27(3):322-6.
- Hameed UF, Lim J, Zhang Q, Wasik MA, Yang D, Swaminathan K. 2014. Transcriptional repressor domain of MBD1 is intrinsically disordered and interacts with its binding partners in a selective manner. *Sci Rep* 4:4896.
- Hammoudeh DI, Follis AV, Prochownik EV, Metallo SJ. 2009. Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-myc. *J Am Chem Soc* 131(21):7390-401.
- Harju-Baker S, Costa FC, Fedosyuk H, Neades R, Peterson KR. 2008. Silencing of agamma-globin gene expression during adult definitive erythropoiesis mediated by GATA-1-FOG-1-Mi2 complex binding at the -566 GATA site. *Mol Cell Biol* 28(10):3101-13.
- Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM. 2006. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol* 2(8):e100.
- Hendrich B and Bird A. 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* 18(11):6538-47.
- Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A. 2001. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev* 15(6):710-23.
- Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A. 1999. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* 401(6750):301-4.
- Hong W, Nakazawa M, Chen YY, Kori R, Vakoc CR, Rakowski C, Blobel GA. 2005. FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *EMBO J* 24(13):2367-78.
- Hu J, Liu J, Xue F, Halverson G, Reid M, Guo A, Chen L, Raza A, Galili N, Jaffray J, et al. 2013. Isolation and functional characterization of human erythroblasts at distinct stages: Implications for understanding of normal and disordered erythropoiesis in vivo. *Blood* 121(16):3246-53.
- Ikonomi P, Noguchi CT, Miller W, Kassahun H, Hardison R, Schechter AN. 2000. Levels of GATA-1/GATA-2 transcription factors modulate expression of embryonic and fetal hemoglobins. *Gene* 261(2):277-87.
- Illingworth RS and Bird AP. 2009. CpG islands--'a rough guide'. *FEBS Lett* 583(11):1713-20.
- INGRAM VM. 1956. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature* 178(4537):792-4.
- Jabbari K and Bernardi G. 2004. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* 333:143-9.
- Jiang CL, Jin SG, Pfeifer GP. 2004. MBD3L1 is a transcriptional repressor that interacts with methyl-CpG-binding protein 2 (MBD2) and components of the NuRD complex. *J Biol Chem* 279(50):52456-64.
- Jiang CL, Jin SG, Lee DH, Lan ZJ, Xu X, O'Connor TR, Szabo PE, Mann JR, Cooney AJ, Pfeifer GP. 2002. MBD3L1 and MBD3L2, two new proteins homologous to the methyl-CpG-binding proteins MBD2 and MBD3: Characterization of MBD3L1 as a testis-specific transcriptional repressor. *Genomics* 80(6):621-9.

- Jiang J, Best S, Menzel S, Silver N, Lai MI, Surdulescu GL, Spector TD, Thein SL. 2006. cMYB is involved in the regulation of fetal hemoglobin production in adults. *Blood* 108(3):1077-83.
- Jin SG, Jiang CL, Rauch T, Li H, Pfeifer GP. 2005. MBD3L2 interacts with MBD3 and components of the NuRD complex and can oppose MBD2-MeCP1-mediated methylation silencing. *J Biol Chem* 280(13):12700-9.
- Jing H, Vakoc CR, Ying L, Mandat S, Wang H, Zheng X, Blobel GA. 2008. Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Mol Cell* 29(2):232-42.
- Jones PA and Baylin SB. 2007. The epigenomics of cancer. *Cell* 128(4):683-92.
- Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP. 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 19(2):187-91.
- Kersh EN. 2006. Impaired memory CD8 T cell development in the absence of methyl-CpG-binding domain protein 2. *J Immunol* 177(6):3821-6.
- Kiga K, Mimuro H, Suzuki M, Shinozaki-Ushiku A, Kobayashi T, Sanada T, Kim M, Ogawa M, Iwasaki YW, Kayo H, et al. 2014. Epigenetic silencing of miR-210 increases the proliferation of gastric epithelium during chronic helicobacter pylori infection. *Nat Commun* 5:4497.
- Kim JK, Samaranyake M, Pradhan S. 2009. Epigenetic mechanisms in mammals. *Cell Mol Life Sci* 66(4):596-612.
- Klein C and Vassilev LT. 2004. Targeting the p53-MDM2 interaction to treat cancer. *Br J Cancer* 91(8):1415-9.
- Kloet SL, Baymaz HI, Makowski M, Groenewold V, Jansen PW, Berendsen M, Niazi H, Kops GJ, Vermeulen M. 2014. Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. *FEBS J* .
- Kosaka N, Sugiura K, Yamamoto Y, Yoshioka Y, Miyazaki H, Komatsu N, Ochiya T, Kato T. 2008. Identification of erythropoietin-induced microRNAs in haematopoietic cells during erythroid differentiation. *Br J Haematol* 142(2):293-300.
- Kransdorf EP, Wang SZ, Zhu SZ, Langston TB, Rupon JW, Ginder GD. 2006. MBD2 is a critical component of a methyl cytosine-binding protein complex isolated from primary erythroid cells. *Blood* 108(8):2836-45.
- Krishnan N, Koveal D, Miller DH, Xue B, Akshinthala SD, Kragelj J, Jensen MR, Gauss CM, Page R, Blackledge M, et al. 2014. Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. *Nat Chem Biol* 10(7):558-66.
- Kumita JR, Smart OS, Woolley GA. 2000. Photo-control of helix content in a short peptide. *Proc Natl Acad Sci U S A* 97(8):3803-8.
- LaBelle JL, Katz SG, Bird GH, Gavathiotis E, Stewart ML, Lawrence C, Fisher JK, Godes M, Pitter K, Kung AL, et al. 2012. A stapled BIM peptide overcomes apoptotic resistance in hematologic cancers. *J Clin Invest* 122(6):2018-31.
- Lages E, Guttin A, El Atifi M, Ramus C, Ipas H, Dupre I, Rolland D, Salon C, Godfraind C, deFraipont F, et al. 2011. MicroRNA and target protein patterns reveal physiopathological features of glioma subtypes. *PLoS One* 6(5):e20600.
- Le Guezennec X, Vermeulen M, Brinkman AB, Hoeijmakers WA, Cohen A, Lasonder E, Stunnenberg HG. 2006. MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol Cell Biol* 26(3):843-51.
- Lebowitz J, Lewis MS, Schuck P. 2002. Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Sci* 11(9):2067-79.
- Lette G, Sankaran VG, Bezerra MA, Araujo AS, Uda M, Sanna S, Cao A, Schlessinger D, Costa FF, Hirschhorn JN, et al. 2008. DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci

- associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc Natl Acad Sci U S A* 105(33):11869-74.
- Levasseur DN, Ryan TM, Pawlik KM, Townes TM. 2003. Correction of a mouse model of sickle cell disease: Lentiviral/antisickling beta-globin gene transduction of unmobilized, purified hematopoietic stem cells. *Blood* 102(13):4312-9.
- Levasseur DN, Ryan TM, Reilly MP, McCune SL, Asakura T, Townes TM. 2004. A recombinant human hemoglobin with anti-sickling properties greater than fetal hemoglobin. *J Biol Chem* 279(26):27518-24.
- Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A. 1992. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* 69(6):905-14.
- Ley TJ, Maloney KA, Gordon JI, Schwartz AL. 1989. Globin gene expression in erythroid human fetal liver cells. *J Clin Invest* 83(3):1032-8.
- Ley TJ, DeSimone J, Anagnou NP, Keller GH, Humphries RK, Turner PH, Young NS, Keller P, Nienhuis AW. 1982. 5-azacytidine selectively increases gamma-globin synthesis in a patient with beta+ thalassemia. *N Engl J Med* 307(24):1469-75.
- Li X, Romero P, Rani M, Dunker AK, Obradovic Z. 1999. Predicting protein disorder for N-, C-, and internal regions. *Genome Inform Ser Workshop Genome Inform* 10:30-40.
- Lin X and Nelson WG. 2003a. Methyl-CpG-binding domain protein-2 mediates transcriptional repression associated with hypermethylated GSTP1 CpG islands in MCF-7 breast cancer cells. *Cancer Res* 63(2):498-504.
- Lin X and Nelson WG. 2003b. Methyl-CpG-binding domain protein-2 mediates transcriptional repression associated with hypermethylated GSTP1 CpG islands in MCF-7 breast cancer cells. *Cancer Res* 63(2):498-504.
- Lopez RA, Schoetz S, DeAngelis K, O'Neill D, Bank A. 2002. Multiple hematopoietic defects and delayed globin switching in ikaros null mice. *Proc Natl Acad Sci U S A* 99(2):602-7.
- Lowrey CH and Nienhuis AW. 1993. Brief report: Treatment with azacitidine of patients with end-stage beta-thalassemia. *N Engl J Med* 329(12):845-8.
- Lu Y, Loh YH, Li H, Cesana M, Ficarro SB, Parikh JR, Salomonis N, Toh CX, Andreadis ST, Luckey CJ, et al. 2014. Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells. *Cell Stem Cell* 15(1):92-101.
- Ma Q, Wyszynski DF, Farrell JJ, Kutlar A, Farrer LA, Baldwin CT, Steinberg MH. 2007. Fetal hemoglobin in sickle cell anemia: Genetic determinants of response to hydroxyurea. *Pharmacogenomics J* 7(6):386-94.
- Macdonald JL, Verster A, Berndt A, Roskams AJ. 2010. MBD2 and MeCP2 regulate distinct transitions in the stage-specific differentiation of olfactory receptor neurons. *Mol Cell Neurosci* 44(1):55-67.
- Maddalo D, Machado E, Concepcion CP, Bonetti C, Vidigal JA, Han YC, Ogradowski P, Crippa A, Rekhman N, de Stanchina E, et al. 2014. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* .
- Magdinier F and Wolffe AP. 2001. Selective association of the methyl-CpG binding protein MBD2 with the silent p14/p16 locus in human neoplasia. *Proc Natl Acad Sci U S A* 98(9):4990-5.
- Martin DI, Tsai SF, Orkin SH. 1989. Increased gamma-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. *Nature* 338(6214):435-8.
- McGrath K and Palis J. 2008. Ontogeny of erythropoiesis in the mammalian embryo. *Curr Top Dev Biol* 82:1-22.
- Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. 1989. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 58(3):499-507.

- Menzel S, Garner C, Gut I, Matsuda F, Yamaguchi M, Heath S, Foglio M, Zelenika D, Boland A, Rooks H, et al. 2007. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat Genet* 39(10):1197-9.
- Metallo SJ. 2010. Intrinsically disordered proteins are potential drug targets. *Curr Opin Chem Biol* 14(4):481-8.
- Mian OY, Wang SZ, Zhu SZ, Gnanapragasam MN, Graham L, Bear HD, Ginder GD. 2011. Methyl-binding domain protein 2-dependent proliferation and survival of breast cancer cells. *Mol Cancer Res* 9(8):1152-62.
- Miccio A and Blobel GA. 2010. Role of the GATA-1/FOG-1/NuRD pathway in the expression of human beta-like globin genes. *Mol Cell Biol* 30(14):3460-70.
- Miccio A, Wang Y, Hong W, Gregory GD, Wang H, Yu X, Choi JK, Shelat S, Tong W, Poncz M, et al. 2010. NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development. *EMBO J* 29(2):442-56.
- Millar CB, Guy J, Sansom OJ, Selfridge J, MacDougall E, Hendrich B, Keightley PD, Bishop SM, Clarke AR, Bird A. 2002. Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* 297(5580):403-5.
- Miller IJ and Bieker JJ. 1993. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the kruppel family of nuclear proteins. *Mol Cell Biol* 13(5):2776-86.
- Miltenberger-Miltenyi G and Laccone F. 2003. Mutations and polymorphisms in the human methyl CpG-binding protein MECP2. *Hum Mutat* 22(2):107-15.
- Moellering RE, Cornejo M, Davis TN, Del Bianco C, Aster JC, Blacklow SC, Kung AL, Gilliland DG, Verdine GL, Bradner JE. 2009. Direct inhibition of the NOTCH transcription factor complex. *Nature* 462(7270):182-8.
- Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN. 2006. Analysis of molecular recognition features (MoRFs). *J Mol Biol* 362(5):1043-59.
- Myung JK, Banuelos CA, Fernandez JG, Mawji NR, Wang J, Tien AH, Yang YC, Tavakoli I, Haile S, Watt K, et al. 2013. An androgen receptor N-terminal domain antagonist for treating prostate cancer. *J Clin Invest* 123(7):2948-60.
- Nakao M, Matsui S, Yamamoto S, Okumura K, Shirakawa M, Fujita N. 2001. Regulation of transcription and chromatin by methyl-CpG binding protein MBD1. *Brain Dev* 23 Suppl 1:S174-6.
- Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393(6683):386-9.
- Ng HH, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D, Bird A. 1999. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet* 23(1):58-61.
- Nuez B, Michalovich D, Bygrave A, Ploemacher R, Grosveld F. 1995. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLf gene. *Nature* 375(6529):316-8.
- O'Neill DW, Schoetz SS, Lopez RA, Castle M, Rabinowitz L, Shor E, Krawchuk D, Goll MG, Renz M, Seelig HP, et al. 2000. An ikaros-containing chromatin-remodeling complex in adult-type erythroid cells. *Mol Cell Biol* 20(20):7572-82.
- Palis J. 2014. Primitive and definitive erythropoiesis in mammals. *Front Physiol* 5:3.
- Papapetrou EP, Lee G, Malani N, Setty M, Riviere I, Tirunagari LM, Kadota K, Roth SL, Giardina P, Viale A, et al. 2011. Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells. *Nat Biotechnol* 29(1):73-8.
- Patil A and Nakamura H. 2006. Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett* 580(8):2041-5.

- Perkins AC, Gaensler KM, Orkin SH. 1996. Silencing of human fetal globin expression is impaired in the absence of the adult beta-globin gene activator protein EKLF. *Proc Natl Acad Sci U S A* 93(22):12267-71.
- Perrine SP, Ginder GD, Faller DV, Dover GH, Ikuta T, Witkowska HE, Cai SP, Vichinsky EP, Olivieri NF. 1993. A short-term trial of butyrate to stimulate fetal-globin-gene expression in the beta-globin disorders. *N Engl J Med* 328(2):81-6.
- Perumbeti A, Higashimoto T, Urbinati F, Franco R, Meiselman HJ, Witte D, Malik P. 2009. A novel human gamma-globin gene vector for genetic correction of sickle cell anemia in a humanized sickle mouse model: Critical determinants for successful correction. *Blood* 114(6):1174-85.
- Peschle C, Mavilio F, Care A, Migliaccio G, Migliaccio AR, Salvo G, Samoggia P, Petti S, Guerriero R, Marinucci M. 1985. Haemoglobin switching in human embryos: Asynchrony of zeta----alpha and epsilon----gamma-globin switches in primitive and definite erythropoietic lineage. *Nature* 313(5999):235-8.
- Pestina TI, Hargrove PW, Jay D, Gray JT, Boyd KM, Persons DA. 2009. Correction of murine sickle cell disease using gamma-globin lentiviral vectors to mediate high-level expression of fetal hemoglobin. *Mol Ther* 17(2):245-52.
- Peterson KR, Costa FC, Fedosyuk H, Neades RY, Chazelle AM, Zelenchuk L, Fonteles AH, Dalal P, Roy A, Chaguturu R, et al. 2014. A cell-based high-throughput screen for novel chemical inducers of fetal hemoglobin for treatment of hemoglobinopathies. *PLoS One* 9(9):e107006.
- Petronzelli F, Riccio A, Markham GD, Seeholzer SH, Stoerker J, Genuardi M, Yeung AT, Matsumoto Y, Bellacosa A. 2000. Biphasic kinetics of the human DNA repair protein MED1 (MBD4), a mismatch-specific DNA N-glycosylase. *J Biol Chem* 275(42):32422-9.
- Pheesse TJ, Parry L, Reed KR, Ewan KB, Dale TC, Sansom OJ, Clarke AR. 2008. Deficiency of Mbd2 attenuates wnt signaling. *Mol Cell Biol* 28(19):6094-103.
- Platt OS, Brambilla DJ, Rosse WF, Milner PF, Castro O, Steinberg MH, Klug PP. 1994. Mortality in sickle cell disease. life expectancy and risk factors for early death. *N Engl J Med* 330(23):1639-44.
- Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M, et al. 2014. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* 159(2):440-55.
- Pontius BW. 1993. Close encounters: Why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends Biochem Sci* 18(5):181-6.
- Pope SH, Fibach E, Sun J, Chin K, Rodgers GP. 2000. Two-phase liquid culture system models normal human adult erythropoiesis at the molecular level. *Eur J Haematol* 64(5):292-303.
- Powars DR, Weiss JN, Chan LS, Schroeder WA. 1984. Is there a threshold level of fetal hemoglobin that ameliorates morbidity in sickle cell anemia? *Blood* 63(4):921-6.
- Prokhortchouk A, Hendrich B, Jorgensen H, Ruzov A, Wilm M, Georgiev G, Bird A, Prokhortchouk E. 2001. The p120 catenin partner kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev* 15(13):1613-8.
- Radmilovic M, Zukic B, Petrovic MS, Bartsakoulia M, Stankovic B, Kotur N, Dokmanovic L, Georgitsi M, Patrinos GP, Pavlovic S. 2013. Functional analysis of a novel KLF1 gene promoter variation associated with hereditary persistence of fetal hemoglobin. *Ann Hematol* 92(1):53-8.
- Raich N, Enver T, Nakamoto B, Josephson B, Papayannopoulou T, Stamatoyannopoulos G. 1990. Autonomous developmental control of human embryonic globin gene switching in transgenic mice. *Science* 250(4984):1147-9.
- Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, Mansour AA, Caspi I, Krupalnik V, Zerbib M, et al. 2013. Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* 502(7469):65-70.

- Rank G, Cerruti L, Simpson RJ, Moritz RL, Jane SM, Zhao Q. 2010a. Identification of a PRMT5-dependent repressor complex linked to silencing of human fetal globin gene expression. *Blood* 116(9):1585-92.
- Rank G, Cerruti L, Simpson RJ, Moritz RL, Jane SM, Zhao Q. 2010b. Identification of a PRMT5-dependent repressor complex linked to silencing of human fetal globin gene expression. *Blood* 116(9):1585-92.
- Riccio A, Aaltonen LA, Godwin AK, Loukola A, Percesepe A, Salovaara R, Masciullo V, Genuardi M, Paravatou-Petsotas M, Bassi DE, et al. 1999. The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. *Nat Genet* 23(3):266-8.
- Rodriguez-Paredes M and Esteller M. 2011. Cancer epigenetics reaches mainstream oncology. *Nat Med* 17(3):330-9.
- Romero, Obradovic, Dunker K. 1997. Sequence data analysis for long disordered regions prediction in the calcineurin family. *Genome Inform Ser Workshop Genome Inform* 8:110-24.
- Romero Z, Urbinati F, Geiger S, Cooper AR, Wherley J, Kaufman ML, Hollis RP, de Assin RR, Senadheera S, Sahagian A, et al. 2013. Beta-globin gene transfer to human bone marrow for sickle cell disease. *J Clin Invest* .
- Rupon JW, Wang SZ, Gnanapragasam M, Labropoulos S, Ginder GD. 2011a. MBD2 contributes to developmental silencing of the human epsilon-globin gene. *Blood Cells Mol Dis* 46(3):212-9.
- Rupon JW, Wang SZ, Gnanapragasam M, Labropoulos S, Ginder GD. 2011b. MBD2 contributes to developmental silencing of the human epsilon-globin gene. *Blood Cells Mol Dis* 46(3):212-9.
- Rupon JW, Wang SZ, Gaensler K, Lloyd J, Ginder GD. 2006a. Methyl binding domain protein 2 mediates gamma-globin gene silencing in adult human betaYAC transgenic mice. *Proc Natl Acad Sci U S A* 103(17):6617-22.
- Rupon JW, Wang SZ, Gaensler K, Lloyd J, Ginder GD. 2006b. Methyl binding domain protein 2 mediates gamma-globin gene silencing in adult human betaYAC transgenic mice. *Proc Natl Acad Sci U S A* 103(17):6617-22.
- Saito M and Ishikawa F. 2002. The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J Biol Chem* 277(38):35434-9.
- Sankaran VG, Xu J, Orkin SH. 2010a. Advances in the understanding of haemoglobin switching. *Br J Haematol* 149(2):181-94.
- Sankaran VG, Xu J, Orkin SH. 2010b. Advances in the understanding of haemoglobin switching. *Br J Haematol* 149(2):181-94.
- Sankaran VG, Menne TF, Scepanovic D, Vergilio JA, Ji P, Kim J, Thiru P, Orkin SH, Lander ES, Lodish HF. 2011. MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. *Proc Natl Acad Sci U S A* 108(4):1519-24.
- Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HK, Hirschhorn JN, Cantor AB, Orkin SH. 2008. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* 322(5909):1839-42.
- Sankaran VG, Xu J, Ragoczy T, Ippolito GC, Walkley CR, Maika SD, Fujiwara Y, Ito M, Groudine M, Bender MA, et al. 2009. Developmental and species-divergent globin switching are driven by BCL11A. *Nature* 460(7259):1093-7.
- Sansom OJ, Berger J, Bishop SM, Hendrich B, Bird A, Clarke AR. 2003. Deficiency of Mbd2 suppresses intestinal tumorigenesis. *Nat Genet* 34(2):145-7.
- Sarakul O, Vattanaviboon P, Tanaka Y, Fucharoen S, Abe Y, Svasti S, Umemura T. 2013. Enhanced erythroid cell differentiation in hypoxic condition is in part contributed by miR-210. *Blood Cells Mol Dis* 51(2):98-103.
- Sarraf SA and Stancheva I. 2004. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol Cell* 15(4):595-605.

- Satta S, Perseu L, Maccioni L, Giagu N, Galanello R. 2012. Delayed fetal hemoglobin switching in subjects with KLF1 gene mutation. *Blood Cells Mol Dis* 48(1):22-4.
- Sauntharajah Y, Hillery CA, Lavelle D, Molokie R, Dorn L, Bressler L, Gavazova S, Chen YH, Hoffman R, DeSimone J. 2003. Effects of 5-aza-2'-deoxycytidine on fetal hemoglobin levels, red cell adhesion, and hematopoietic differentiation in patients with sickle cell disease. *Blood* 102(12):3865-70.
- Saxonov S, Berg P, Brutlag DL. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* 103(5):1412-7.
- Scarsdale JN, Webb HD, Ginder GD, Williams DC, Jr. 2011a. Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res* 39(15):6741-52.
- Scarsdale JN, Webb HD, Ginder GD, Williams DC, Jr. 2011b. Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res* 39(15):6741-52.
- Schuck P. 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78(3):1606-19.
- Sebastiano V, Maeder ML, Angstman JF, Haddad B, Khayter C, Yeo DT, Goodwin MJ, Hawkins JS, Ramirez CL, Batista LF, et al. 2011. In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells* 29(11):1717-26.
- Shen Y and Bax A. 2013. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* 56(3):227-41.
- Shen Y, Delaglio F, Cornilescu G, Bax A. 2009. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44(4):213-23.
- Shenoy S. 2013. Hematopoietic stem-cell transplantation for sickle cell disease: Current evidence and opinions. *Ther Adv Hematol* 4(5):335-44.
- Shi L, Cui S, Engel JD, Tanabe O. 2013. Lysine-specific demethylase 1 is a therapeutic target for fetal hemoglobin induction. *Nat Med* 19(3):291-4.
- Singal R, Wang SZ, Sargent T, Zhu SZ, Ginder GD. 2002. Methylation of promoter proximal-transcribed sequences of an embryonic globin gene inhibits transcription in primary erythroid cells and promotes formation of a cell type-specific methyl cytosine binding complex. *J Biol Chem* 277(3):1897-905.
- Singal R, Ferris R, Little JA, Wang SZ, Ginder GD. 1997. Methylation of the minimal promoter of an embryonic globin gene silences transcription in primary erythroid cells. *Proc Natl Acad Sci U S A* 94(25):13724-9.
- Singleton BK, Frayne J, Anstee DJ. 2012. Blood group phenotypes resulting from mutations in erythroid transcription factors. *Curr Opin Hematol* 19(6):486-93.
- Smits AH, Jansen PW, Poser I, Hyman AA, Vermeulen M. 2013. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res* 41(1):e28.
- Sripichai O, Kiefer CM, Bhanu NV, Tanno T, Noh SJ, Goh SH, Russell JE, Rognerud CL, Ou CN, Oneal PA, et al. 2009. Cytokine-mediated increases in fetal hemoglobin are associated with globin gene histone modification and transcription factor reprogramming. *Blood* 114(11):2299-306.
- Stadhouders R, Aktuna S, Thongjuea S, Aghajani-farah A, Pourfarzad F, van Ijcken W, Lenhard B, Rooks H, Best S, Menzel S, et al. 2014. HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. *J Clin Invest* 124(4):1699-710.
- Stamatoyannopoulos G, Majerus PW, Perimutter RM, Varmus H. 2001. *The molecular basis of blood diseases*. 3rd ed. Philadelphia, PA: WB Saunders.

- Stauffer SR. 2007. Small molecule inhibition of the bcl-X(L)-BH3 protein-protein interaction: Proof-of-concept of an in vivo chemopotentiator ABT-737. *Curr Top Med Chem* 7(10):961-5.
- Stefanska B, Suderman M, Machnes Z, Bhattacharyya B, Hallett M, Szyf M. 2013. Transcription onset of genes critical in liver carcinogenesis is epigenetically regulated by methylated DNA-binding protein MBD2. *Carcinogenesis* 34(12):2738-49.
- Steinberg MH. 2008. Sickle cell anemia, the first molecular disease: Overview of molecular etiology, pathophysiology, and therapeutic approaches. *ScientificWorldJournal* 8:1295-324.
- Steinberg MH, Lu ZH, Barton FB, Terrin ML, Charache S, Dover GJ. 1997. Fetal hemoglobin in sickle cell anemia: Determinants of response to hydroxyurea. multicenter study of hydroxyurea. *Blood* 89(3):1078-88.
- Straussman R, Nejman D, Roberts D, Steinfeld I, Blum B, Benvenisty N, Simon I, Yakhini Z, Cedar H. 2009. Developmental programming of CpG island methylation profiles in the human genome. *Nat Struct Mol Biol* 16(5):564-71.
- Sun N and Zhao H. 2014. Seamless correction of the sickle cell disease mutation of the HBB gene in human induced pluripotent stem cells using TALENs. *Biotechnol Bioeng* 111(5):1048-53.
- Suzuki M, Yamazaki H, Mukai HY, Motohashi H, Shi L, Tanabe O, Engel JD, Yamamoto M. 2013. Disruption of the Hbs1l-myb locus causes hereditary persistence of fetal hemoglobin in a mouse model. *Mol Cell Biol* 33(8):1687-95.
- Swiech L, Heidenreich M, Banerjee A, Habib N, Li Y, Trombetta J, Sur M, Zhang F. 2014. In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat Biotechnol* .
- Takahashi S, Onodera K, Motohashi H, Suwabe N, Hayashi N, Yanai N, Nabesima Y, Yamamoto M. 1997. Arrest in primitive erythroid cell development caused by promoter-specific disruption of the GATA-1 gene. *J Biol Chem* 272(19):12611-5.
- Tanabe O, Katsuoka F, Campbell AD, Song W, Yamamoto M, Tanimoto K, Engel JD. 2002. An embryonic/fetal beta-type globin gene repressor contains a nuclear receptor TR2/TR4 heterodimer. *EMBO J* 21(13):3434-42.
- Tanabe O, McPhee D, Kobayashi S, Shen Y, Brandt W, Jiang X, Campbell AD, Chen YT, Chang C, Yamamoto M, et al. 2007. Embryonic and fetal beta-globin gene repression by the orphan nuclear receptors, TR2 and TR4. *EMBO J* 26(9):2295-306.
- Tanimoto K, Liu Q, Grosveld F, Bungert J, Engel JD. 2000. Context-dependent EKLF responsiveness defines the developmental specificity of the human epsilon-globin gene in erythroid cells of YAC transgenic mice. *Genes Dev* 14(21):2778-94.
- Thein SL, Menzel S, Lathrop M, Garner C. 2009. Control of fetal hemoglobin: New insights emerging from genomics and clinical implications. *Hum Mol Genet* 18(R2):R216-23.
- Thein SL, Menzel S, Peng X, Best S, Jiang J, Close J, Silver N, Gerovasilli A, Ping C, Yamaguchi M, et al. 2007. Intergenic variants of HBS1L-MYB are responsible for a major quantitative trait locus on chromosome 6q23 influencing fetal hemoglobin levels in adults. *Proc Natl Acad Sci U S A* 104(27):11346-51.
- Tomba P. 2012. Intrinsically disordered proteins: A 10-year recap. *Trends Biochem Sci* 37(12):509-16.
- Tomba P. 2002. Intrinsically unstructured proteins. *Trends Biochem Sci* 27(10):527-33.
- Toth G, Gardai SJ, Zago W, Bertocini CW, Cremades N, Roy SL, Tambe MA, Rochet JC, Galvagnion C, Skibinski G, et al. 2014. Targeting the intrinsically disordered structural ensemble of alpha-synuclein by small molecules as a potential therapeutic strategy for parkinson's disease. *PLoS One* 9(2):e87133.
- Tsai SF, Martin DI, Zon LI, D'Andrea AD, Wong GG, Orkin SH. 1989. Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature* 339(6224):446-51.

- Tsang AP, Fujiwara Y, Hom DB, Orkin SH. 1998. Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes Dev* 12(8):1176-88.
- Tsuzuki S and Enver T. 2002. Interactions of GATA-2 with the promyelocytic leukemia zinc finger (PLZF) protein, its homologue FAZF, and the t(11;17)-generated PLZF-retinoic acid receptor alpha oncoprotein. *Blood* 99(9):3404-10.
- Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G, et al. 2008. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci U S A* 105(5):1620-5.
- Uversky VN. 2011. Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol* 43(8):1090-103.
- Uversky VN, Li J, Fink AL. 2001. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J Biol Chem* 276(14):10737-44.
- Vacic V, Oldfield CJ, Mohan A, Radivojac P, Cortese MS, Uversky VN, Dunker AK. 2007. Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res* 6(6):2351-66.
- Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker J, Blobel GA. 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell* 17(3):453-62.
- van Dijk TB, Gillemans N, Pourfarzad F, van Lom K, von Lindern M, Grosveld F, Philipsen S. 2010. Fetal globin expression is regulated by friend of Prmt1. *Blood* 116(20):4349-52.
- Vegiopoulos A, Garcia P, Emambokus N, Frampton J. 2006. Coordination of erythropoiesis by the transcription factor c-myb. *Blood* 107(12):4703-10.
- Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED. 2005. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* 59(4):687-96.
- Wade PA. 2001. Methyl CpG binding proteins: Coupling chromatin architecture to gene regulation. *Oncogene* 20(24):3166-73.
- Wahlberg K, Jiang J, Rooks H, Jawaid K, Matsuda F, Yamaguchi M, Lathrop M, Thein SL, Best S. 2009. The HBS1L-MYB intergenic interval associated with elevated HbF levels shows characteristics of a distal regulatory region in erythroid cells. *Blood* 114(6):1254-62.
- Walavalkar NM, Gordon N, Williams DC, Jr. 2013. Unique features of the anti-parallel, heterodimeric coiled-coil interaction between methyl-cytosine binding domain 2 (MBD2) homologues and GATA zinc finger domain containing 2A (GATAD2A/p66alpha). *J Biol Chem* 288(5):3419-27.
- Wang CX, Sather BD, Wang X, Adair J, Khan I, Singh S, Lang S, Adams A, Curinga G, Kiem HP, et al. 2014. Rapamycin relieves lentiviral vector transduction resistance in human and mouse hematopoietic stem cells. *Blood* 124(6):913-23.
- WATSON J. 1948. A study of sickling of young erythrocytes in sickle cell anemia. *Blood* 3(4):465-9.
- Weatherall DJ and Clegg JB. 2001. Inherited haemoglobin disorders: An increasing global health problem. *Bull World Health Organ* 79(8):704-12.
- Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, Svergun DI, Blackledge M, Fersht AR. 2008. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc Natl Acad Sci U S A* 105(15):5762-7.
- Wijgerde M, Gribnau J, Trimborn T, Nuez B, Philipsen S, Grosveld F, Fraser P. 1996. The role of EKLF in human beta-globin gene competition. *Genes Dev* 10(22):2894-902.
- Wilusz JE, Sunwoo H, Spector DL. 2009. Long noncoding RNAs: Functional surprises from the RNA world. *Genes Dev* 23(13):1494-504.

- Xie F, Ye L, Chang JC, Beyer AI, Wang J, Muench MO, Kan YW. 2014. Seamless gene correction of beta-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Res* 24(9):1526-33.
- Xiong L, Wang F, Huang X, Liu ZH, Zhao T, Wu LY, Wu K, Ding X, Liu S, Wu Y, et al. 2012. DNA demethylation regulates the expression of miR-210 in neural progenitor cells subjected to hypoxia. *FEBS J* 279(23):4318-26.
- Xu J, Sankaran VG, Ni M, Menne TF, Puram RV, Kim W, Orkin SH. 2010. Transcriptional silencing of γ -globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev* 24(8):783-98.
- Xu J, Bauer DE, Kerenyi MA, Vo TD, Hou S, Hsu YJ, Yao H, Trowbridge JJ, Mandel G, Orkin SH. 2013. Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. *Proc Natl Acad Sci U S A* 110(16):6518-23.
- Yi Z, Cohen-Barak O, Hagiwara N, Kingsley PD, Fuchs DA, Erickson DT, Epner EM, Palis J, Brilliant MH. 2006. Sox6 directly silences epsilon globin expression in definitive erythropoiesis. *PLoS Genet* 2(2):e14.
- Yildirim O, Li R, Hung JH, Chen PB, Dong X, Ee LS, Weng Z, Rando OJ, Fazzio TG. 2011. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147(7):1498-510.
- Zhao Q, Zhou W, Rank G, Sutton R, Wang X, Cumming H, Cerruti L, Cunningham JM, Jane SM. 2006. Repression of human gamma-globin gene expression by a short isoform of the NF-E4 protein is associated with loss of NF-E2 and RNA polymerase II recruitment to the promoter. *Blood* 107(5):2138-45.
- Zhao Q, Rank G, Tan YT, Li H, Moritz RL, Simpson RJ, Cerruti L, Curtis DJ, Patel DJ, Allis CD, et al. 2009a. PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat Struct Mol Biol* 16(3):304-11.
- Zhao Q, Rank G, Tan YT, Li H, Moritz RL, Simpson RJ, Cerruti L, Curtis DJ, Patel DJ, Allis CD, et al. 2009b. PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat Struct Mol Biol* 16(3):304-11.
- Zhao X, Ueba T, Christie BR, Barkho B, McConnell MJ, Nakashima K, Lein ES, Eadie BD, Willhoite AR, Muotri AR, et al. 2003. Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. *Proc Natl Acad Sci U S A* 100(11):6777-82.
- Zhou D, Liu K, Sun CW, Pawlik KM, Townes TM. 2010. KLF1 regulates BCL11A expression and gamma-to beta-globin gene switching. *Nat Genet* 42(9):742-4.
- Zhou W, Zhao Q, Sutton R, Cumming H, Wang X, Cerruti L, Hall M, Wu R, Cunningham JM, Jane SM. 2004. The role of p22 NF-E4 in human globin gene switching. *J Biol Chem* 279(25):26227-32.
- Zhou X, Ruan J, Wang G, Zhang W. 2007. Characterization and identification of microRNA core promoters in four model species. *PLoS Comput Biol* 3(3):e37.

Vita

Megha Anupam Desai was born on April 11, 1987 in Vadodara, Gujarat, India, and is an Indian citizen. She graduated from Baroda High School, Vadodara, India in 2005. She received her Bachelor of Engineering in Biotechnology from NMAM Institute of Technology, Nitte, Karnataka, India in 2009.